Contrast invariant image analysis and PDE's

Frédéric Guichard   Jean-Michel Morel   Robert Ryan

Contents

Introduction

This book addresses the problem of low-level image analysis and, as such, is a contribution to image processing and imaging science. While the material touches on several aspects of image analysis—and peripherally on other parts of image processing—the main subject is image smoothing using partial differential equations (PDEs). The rational for a book devoted to smoothing is the assumption that a digital image must be smoothed before reliable features can be extracted.

The purpose of this introduction is to establish some of the language, conventions, and assumptions that are used throughout the book, to review part of the history of PDEs in image processing, and to introduce notation and background material.

I.1 Images

Since the objects of our study are ultimately digital images, we begin by defining what we mean by "digital image" and by describing some of the ways these images are obtained and some current assumptions about the "original images" from which the digital images are derived.

Most of the images dealt with will be natural images, that is, images from nature (people, landscapes, cityscapes, etc.). We include medical images and astronomical images, and we do not exclude drawings, paintings, and other man-made images. All of the images we consider will be grayscale images. Thus, mathematically, an image is a real-valued function u defined on some subset Ω of the plane $\mathbb{R}^2$. The value $u(\boldsymbol{x})$, $\boldsymbol{x} = (x, y) \in \Omega$, represents the gray level of the image at the point $\boldsymbol{x}$. If u is a digital image, then its domain of definition is a finite grid with evenly spaced points. It is often square with $2^n \times 2^n$ points. The gray levels $u(\boldsymbol{x})$ are typically coded with the integers 0–255, where 0 represents black and 255 represents white. If h is the distance between grid lines, then the squares with sides of length h centered at the points $u(\boldsymbol{x})$ are called *pixels*, where "pix" is slang for "picture" and "el" stands for "element."

The mathematical development in this book proceeds along two parallel lines. The first is theoretical and deals with images u that belong to function spaces, generally spaces of continuous functions that are defined on domains of $\mathbb{R}^2$. The second line concerns numerical algorithms, and for this the images are digital images. To understand the relations between the digital and continuous images, it is useful to consider some examples of how images are obtained and some of the assumptions we make about the processes and the images. Perhaps the simplest example is that of taking a picture of a natural scene with a digital camera. The scene—call it S—is focused at the focal plane of the camera forming a representation of S that we denote by u_f. When we take the picture, the image u_f is sampled, or captured, by an array of charged coupled devices (CCDs) producing the digital image u_d. This image, u_d, is the only representation of S that is directly available to us; the image u_f is not directly available to us. Even more elusive is the completely hypothetical image that we call u_S. This is the representation of S that would be formed at the focal plane of an ideal camera having perfect optics. A variation on this example is to capture u_f on film as the image u_p. Then u_p can be sampled (scanned) to produce a digital image u_d. For example, before the advent of CCDs, astronomical images were captured on Schmidt plates. Many of these plates have been scanned recently, and the digital images have been made available to astronomers via the Internet.

Aspects of the photographic example could be recast for medical imaging. Although photography plays an important role in medicine, images for diagnostic use are often obtained using other kinds of radiation. X-rays are perhaps closest to our photographic example. In this case, there is an image corresponding to u_p that can be scanned to produce a digital image u_d. Other medical imaging processes, such as scintigraphy and nuclear magnetic resonance, are more complicated, but these processes yield digital images. The images examined by the experts are often "negatives" produced from an original digital images. Irrespective of the process, digital images captured by some technology all have one characteristic in common: They are all noisy.

One way to relate the different representations of S, is to write

$$u_d = Tu_S + n,$$

where T is a hypothetical operator representing some technology and n is noise. In the case of photography, we might write this in two steps,

$$\begin{cases} u_f = P * u_S + n_1, \\ u_d = Ru_f + n_2, \end{cases}$$

where P represents the optics and R represents the sampling. This is a useful model in optical astronomy, since astronomers have considerable knowledge about the operators P and R and about the noises n_1 and n_2. Similarly, experts in other technologies know a great deal about the processes and noise sources. Noise and pixels are illustrated in Figure I.1

In the photographic example, the image u_f is a smoothed version of u_S. Furthermore, $Ru_f(\boldsymbol{x})$ is not exactly $u_f(\boldsymbol{x})$ but rather an average of values of u_f in a small neighborhood of $\boldsymbol{x}$, which is to say that the operator R does some smoothing. Thus, in this example, u_d is sampled from a smoothed version of S. We are going to assume that this is the case for the digital images considered in the book, except for digital images that are artificially generated. This is realistic, since all of the processes T that we can imagine for capturing images, smooth the original photon flux. In fact, this is more of an observation about technology than it is an assumption. We are also going to assume that, for any technology considered, the sampling rate used to produce u_d is high enough so that u_d is a "good" representation of the smoothed version of S, call it u_f, from which it was derived. Here, "good" means that the parallel development in the book mentioned above make sense; it means that, from a practical point of view, the theoretical development that uses smooth functions to model the images u_f is indeed related to the algorithmic development that uses the digital images u_d. We will say more about smoothing and sampling in section I.2.

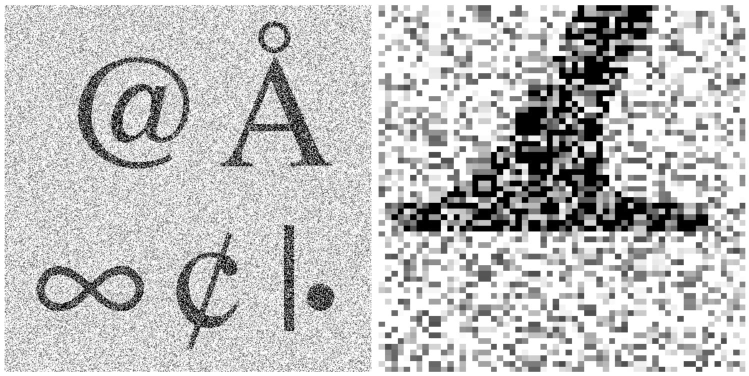

Figure I.1: A noisy image magnified to show the pixels.

It is widely assumed that the underlying "real image" u_S is either a measure or, for more optimistic authors, a function that has strong discontinuities. Rudin in 1987 [241] and De Giorgi and Ambrosio in 1988 [125] proposed independently the space $BV(\mathbb{R}^2)$ of functions with bounded variation as the correct function space for modeling the images u_S. A function f is in $BV(\mathbb{R}^2)$ if its partial derivatives $\partial f/\partial x$ and $\partial f/\partial y$, taken as distributions, are Radon measures with finite total mass. $BV(\mathbb{R}^2)$ looked at first well adapted to modeling digital images because it contains functions having step discontinuities. In fact, the characteristic functions of smooth domains in $\mathbb{R}^2$ belong to $BV(\mathbb{R}^2)$. However, in 1999, Alvarez, Gousseau, and Morel used a statistical device on digital images u_d to estimate how the corresponding images u_S oscillate [8]. They deduced by geometric-measure arguments, that the u_S have, in fact, unbounded variation. We may therefore accept the idea that these high-resolution images contain very strong oscillations. Although the images u_f are smoothed versions of the u_S, and hence the oscillations have been averaged, common sense tells us that they also have large derivatives at transitions between different observed objects, that is, on the apparent contours of physical objects. Furthermore, we expect that these large derivatives (along with noise) are passed to the digital images u_d.

I.2 Image processing

For the convenience of exposition, we divide image processing into separate disciplines. These are distinguished not so much by their techniques, which often overlap, as they are by their goals. We will briefly describe two of these areas: compression and restoration. The third area, image analysis, is the main subject of the book and will be discussed in more detail.

Image compression

Compression is based on the discrete nature of digital images, and it is motivated by economic necessity: Each form of storage and transmission has an associated cost, and hence one wishes to represent an image with the least number of bits that is compatible with end usage. There are two kinds of compression: lossless compression and lossy compression. Lossless compression algorithms are used to compress digital files where the decompressed file must agree bit-by-bit with the original file. Perhaps the best known example of lossless compression is the zip format. Lossless algorithms can be used on any digital file, including digital images. These algorithms take advantage of the structure of the file itself and have nothing to do with what the file represents. On the other hand, lossy compression algorithms take advantage of redundancies in natural images and subtleties of the human visual system. Done correctly, one can throw away information contained in an image without impairing its usefulness. The goal is to develop algorithms that provide high compression factors without objectionable visible alterations. Naturally, what is visually objectionable depends on how the decompressed image is used. This is nicely illustrated with our photographic example. Suppose that we capture the image u_f at our camera's highest resolution. If we are going to send u_d over the Internet to a publisher to be printed in a high-quality publication, then we want no loss of information and will probably send the entire file in the zip format. If, however, we just want the publisher to have a quick look at the image, then we would probably send u_d compressed as a .jpg file, using the Joint Photographic Expert Group (JPEG) standard for still image compression. This kind of compression is illustrated in Figure I.2.

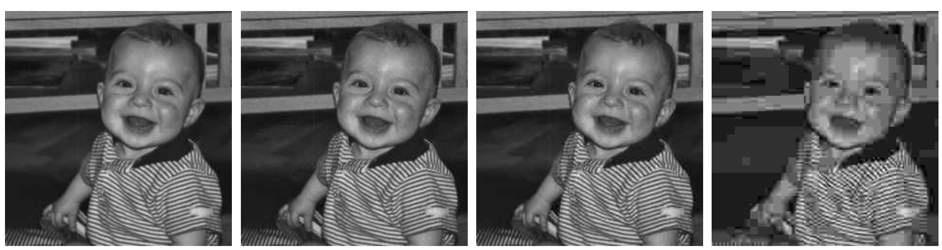

Figure I.2: Compression. Left to right: the original image and its increasingly compressed versions. The compression factors are roughly 7, 10, and 25. Up too a 10 factor, alterations are hardly visible.

Image restoration

A second area is restoration or denoising. Restoring digital images is much like restoring dirty or damaged paintings or photographs. Beginning with a digital image that contains blurs or other perturbations (all of which may be considered as noise), one wishes to produce a better version of the image; one wishes to enhance aspects of the image that have been attenuated or degraded. Image restoration plays an important role in law enforcement and legal proceedings. For example, surveillance cameras generally produce rather poor images that must often be denoised and enhanced as needed. Image restoration is also important in science. When the Hubble Space Telescope was first launched in 1990, and until it was repaired in 1993, the images it returned were all blurred due to a spherical aberration in the telescope's primary mirror. Elaborate (and costly) algorithms were developed to restore these poor images, and indeed useful images were obtained during this period. Restoration is illustrated in Figure I.3 with an artificial example. The image on the left has been ostensibly destroyed by introducing random-valued pixels amounting to 75% of the total pixel count. Nevertheless, the image can be significantly restored, and a restored version is shown on the right, by using a Vincent and Serra operator which we will study in Chapter **??**, the "area opening".

Figure I.3: Denoising. Left: an image with up to 75% of its pixels contaminated by simulated noise. Right: a denoised version by the Vincent-Serra algorithm (area opening).

Image analysis

A third area of image processing is low-level image analysis, and since this is the main topic of the book, it is important to explain what we mean by "low-level" and "analysis." "Analysis" is widely used in mathematics, with various shades of meaning. Our use of "analyze," and thus of "analysis," is very close to its common meaning, which is to decompose a whole into its constituent parts, to study the parts, and to study their relation to the whole. For our purposes, the constituent parts are, for the most part, the "edges" and "shapes" in an image. These objects, which are often called features, are things that we could, for a given image, point to and outline, although for a complex natural image this would be a tedious process. The goal of image analysis is to create algorithms that do this automatically.

The term "low-level" comes from the study of human vision and means extracting reliable, local geometric information from an image. At the same time, we would like the information to be minimal but rich enough to characterize the image. The goal here is not compression, although some of the techniques may provide a compressed representation of the image. Our goal is rather to answer questions like, Does a feature extracted from image A exist in image B? We are also interested in comparing features extracted from an image with features stored in a database. As an example, consider the level set at the left of Figure I.4. It consists of major features (roughly, the seven appendages) and noise. The noise, which is highly variable, prevents us from comparing the image directly with other images having similar shapes. Thus we ask for a sketchy version, where, however, all essential features are kept. The images on the right are such a sketchy versions, where most of the spurious details (or noise) have disappeared, but the main structures are maintained. These sketchy versions may lead to concise invariant encoding of the shape. Notice how the number of inflexion points of the shape has decreased in the simplification process. This is an example of what we mean by image analysis. The aim is not denoising or compression. The aim is to construct an invariant code that puts in evidence the "main parts" of an image (in this case, the appendages) and that facilitates fast recognition in a large database of shapes.

Figure I.4: Analysis of a shape. The original scanned shape is on the left. Simplified versions are to the right.

Edge detection and scale space

Since the earliest work in the 1960s, one of the goals of image analysis has been to locate the strong discontinuities in an image. This search is called *edge detection*, and it derives from early research that involved working with images of cubes. This seemingly simple goal turned out to be exceedingly difficult. Here is what David Marr wrote about the problem in the early 1980s ([198], p. 16):

> The first great revelation was that the problems are difficult. Of course, these days this fact is a commonplace. But in the 1960s almost no one realized that machine vision was difficult. The field had to go through the same experience as the machine translation field did in its fiascoes of the 1950s before it was at last realized that here were some problems that had to be taken seriously. The reason for this misconception is that we humans are ourselves so good at vision. The notion of a feature detector was well established by Barlow and by Hubel and Wiesel, and the idea that extracting edges and lines from images might be at all difficult simply did not occur to those who had not tried to do it. It turned out to be an elusive problem: Edges that are of critical importance from a three-dimensional point of view often cannot be found at all by looking at the intensity changes in an image. Any kind of textured image gives a multitude of noisy edge segments; variations in reflectance

> and illumination cause no end of trouble; and even if an edge has a clear existence at one point, it is as likely as not to fade out quite soon, appearing only in patches along its length in the image. The common and almost despairing feeling of the early investigators like B.K.P. Horn and T.O. Binford was that practically anything could happen in an image and furthermore that practically everything did.

The point we wish to emphasize is that textures and noise (which are often lumped together in image analysis) produce unwanted edges. The challenge was to separate the "true edges" from the noise. For example, one did not want to extract all of the small edges in a textured wall paper; one wanted the outline of the wall. The response was to blur out the textures and noise in a way that left the "true edges" intact, and then to extract these features. More formally, image analysis was reformulated as two processes: smoothing followed by edge detection. At the same time, a new doctrine, the *scale space*, was proposed. Scale space means that instead of speaking of features of an image at a given location, we speak of them at a given location and at a given scale, where the scale quantifies the amount of smoothing performed on the image before computing the features. We will see in experiments that "edges at scale 4" and "edges at scale 7" are different outputs of an edge detector.

Three requirements for image smoothing operators

We have advertised that this book is about image analysis, which we have just defined to be smoothing followed by edge detection, or feature extraction. In fact, the text focuses on smoothing and particularly on discussing and answering the question, What kind of smoothing should be used? To approach this problem, we need to introduce three concepts associated with image analysis operators. These concepts will be used to narrow the field of smoothing operators. We introduce them informally at first; more precise meanings will follow.

Localization

The first notion is *localization*. Roughly speaking, to say that an operator T is localized means it essentially uses information from a small neighborhood of $\boldsymbol{x}$ to compute the output $Tu(\boldsymbol{x})$. Recall that the sampling operator R in the photographic example was well localized. As another example, consider the classic Gaussian smoothing operators $\mathcal{G}_t$ defined by

$$\mathcal{G}_t u(\boldsymbol{x}) = G_t * u(\boldsymbol{x}) = \int_{\mathbb{R}^2} G_t(\boldsymbol{y}) u(\boldsymbol{x} - \boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

where $G_t(\boldsymbol{x}) = (1/4\pi t)\mathrm{e}^{-|\boldsymbol{x}|^2/4t}$. If $t > 0$ is small, then the Gaussian G_t is well localized around zero and $\mathcal{G}_t u(\boldsymbol{x})$ is essentially an average of the values of $u(\boldsymbol{x})$ in a small neighborhood of $\boldsymbol{x}$. The importance of localization is related to the occlusion problem: Most optical images consist of a superposition of different objects that partially obscure one another. It is clear that we must avoid confusing them in the analysis, as would, for example, G_t if t is large. It is for reasons like this that we want the analysis to be as local as possible.

We will prove in Chapter 1 under rather general conditions that $u(t, \boldsymbol{x}) = G_t * u_0(\boldsymbol{x})$ is the unique solution of the heat equation

$$\frac{\partial u}{\partial t} = \Delta u$$

with initial value u_0. Thus, we can say that smoothing u_0 with the Gaussian G_t is equivalent to applying the heat equation to u_0. We will see that the heat equation is possibly the worst candidate in our search for the ideal smoothing operator, since, except for small t, it is poorly localized and produces a very blurred image.

Iteration

One might conjecture that a way around this problem with the heat equation would be to replace G_t with a more suitable positive kernel. This is not the case, but it does serve to introduce the second concept, which is *iteration*. We will show in Chapter 2 that under reasonable assumptions and appropriate rescalings, iterating a convolution with a positive kernel leads to the Gaussian, and thus directly back to the heat equation. There is, however, a different point of view that leads to useful smoothing operators: Instead of looking for a different kernel, look for other PDEs that provide smoothing. This program leads to a class of nonlinear PDEs, where the Laplacian in the heat equation is replaced by various nonlinear operators. We will see that for these operators it is generally better, from the localization point of view, to iterate a well localized operator than to apply it directly at a large scale. This, of course, is just not true for the heat equation; if you iterate n times the convolution $G_t * u$ you get exactly $G_{nt} * u$. This is a good place to point out that if we are dealing with smoothing, localization, and iteration, then we are talking about parabolic PDEs. This announcement is heuristic, and the object of the book is to formalize and to make precise the necessity and the role of several PDEs in image analysis.

Invariance

Our last concept is *invariance*. Invariance requirements play a central role in image analysis because the objects to be recognized must be recognized under varying conditions of illumination (contrast invariance) and from different points of view (projective invariance). Contrast invariance is one of the central requirements of the theory of image analysis called mathematical morphology (see, for example, Matheron [202] or Serra [253]). This theory involves a number of contrast-invariant image analysis operators, including dilations, erosions, median filters, openings, and closings. We are going to use this theory by attempting to localize as much as possible these *morphomath* operators to exploit their behavior at small scales. We will then iterate these operators. This will lead to the proof that several geometric PDEs, namely, the curvature motions, are asymptotically related to certain morphomath operators in much the same way that linear smoothing is related to the heat equation. Thus, through these PDEs, one is able to combine the scale space doctrine and mathematical morphology. In particular, affine-invariant morphomath operators, which seemed at first to be computationally impractical, turn out to yield in their local iterated

Figure I.5: Shannon theory and sampling. Left to right: original image; smoothed image; sampled version of the original image; sampled version of the smoothed image. This illustrates the famous Shannon-Nyquist law that an image must be smoothed before sampling in order to avoid the aliasing artifacts.

version a very affordable PDE, the so called *affine morphological scale space* (AMSS) equation.

Shannon's sampling theory

We mentioned in section I.1 that most of the digital images u_d that come to us in practice have been sampled from a smoothed version, call it u_f, of the "real image" u_S. This was basically a comment about the technology. Another comment (or assumption) was that the sampling rate was high enough to capture all of the information in u_f that is needed in practice. What we mean by this is that the representations of u_f that we reconstruct from u_d show no signs that u_f was undersampled. This is an empirical statement; we will comment on the theory in a moment, but first we wish to illustrate in Figure I.5 what can happen if an image is undersampled.

We call the original image on the left Victor. Notice that Victor's sweater contains a striped pattern, which has a spatial frequency that is high relative to other aspects of the picture. If we attempt to reduce the size of Victor by simply sampling, for example, by taking one pixel in sixteen in a square pattern, we obtain a new image (the third panel) in which the sampling has created new and unstable patterns. Notice how new stripes have been created with a frequency and direction that has nothing to do with the original. This is called *aliasing*, and it is caused by high spatial frequencies being projected onto lower frequencies, which creates new patterns. If this had been a video instead of being a still photo, these newly created patterns would move and flicker in a totally uncontrolled way. This kind of moving pattern often appears in recent commercial DVDs. They have simply not been sampled at a high enough rate. The second panel in Figure I.5 is a version of Victor that has been smoothed enough so that we no longer see the stripes in the sweater. This image is sampled the same way—every fourth pixel horizontally and vertically—and appears in panel four. It is not a good image, but there are no longer the kinds of artifacts that appear in the third image. To compare the images we have magnified the sampled versions by a factor of four. This example also shows that simply subsampling an image is a poor way to compress it.

This pragmatic discussion and the experiment have their theoretical counterpart, namely, Shannon's theory of sampling. Briefly, Shannon's theorem, in the two-dimensional case, states that for an image to be accurately reconstructed from samples, the image must be bandlimited, which means that it contains no spatial frequencies greater than some bound λ, and the sampling rate must be higher than a factor of λ. Some implications of these statements are that the image u must be infinitely differentiable, that its domain of definition is all of $\mathbb{R}^2$, and that there must be an infinite number of samples to accurately reconstruct u. Furthermore, in Shannon's theory, the image u is reconstructed as an infinite series of trigonometric functions. Note that this is very different from what was done in Figure I.5. So what does this have to do with the problems addressed in this book? What does this have to do with, say, a hypothesized u_S in $BV(\mathbb{R}^2)$ that is definitely not bandlimited? Our answer, which may smack of smoke and mirrors, is that we always are working in two parallel worlds, the theoretical one and the practical one based on numerical computations, and that these two worlds live together in harmony at a certain scale. Here is an example of what we mean: Suppose that u is not a bandlimited image. To sample it properly we would first have to smooth it with a bandlimited kernel. Suppose that instead we smooth it with the Gaussian G_t, which is not bandlimited. Theoretically this is wrong, but practically, the spectrum of G_t, which is G_t itself, decays exponentially. If $|\boldsymbol{x}|^2/4t$ is sufficiently large, then $G_t(\boldsymbol{x})$ appears as zero in computations, and thus it is "essentially" bandlimited. Arguments like this could be made for other situations, but the important point for the reader to keep in mind is that the parallel developments, theory and practice, make sense in the limit.

In the next section, we present a survey of most of the PDEs that have been proposed for image analysis. This provides an informal account of the mathematics that will be developed in detail in the following chapters.

We wish to end this section with a mild disclaimer, and for this we take a page from *Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern where they comment on their theory of a zero-sum two-person game [277] p. 147:

> We are trying to find a satisfactory theory,—at this stage for the zero-sum two-person game. Consequently we are not arguing deductively from the firm basis of an existing theory—which has already stood all reasonable tests—but we are searching for such a theory.... This consists in imagining that we have a satisfactory theory of a certain desired type, trying to picture the consequences of this imaginary intellectual situation, and then drawing conclusions from this as to what the hypothetical theory must be like in detail. If this process is applied successfully, it may narrow the possibilities for the hypothetical theory of the type in question to such an extent that only one possibility is left,—i.e. that the theory is determined, discovered by this device. Of course, it can happen that the application is even more "successful," and that it narrows the possibilities down to nothing—i.e. that it demonstrates that a consistent theory of the kind desired is inconceivable.

We take much the same philosophical position, and here is our variation on the von Neumann–Morgenstern statement: We do not suggest that what will be developed here is a necessary future for image analysis. However, 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.

I.3 PDEs and image processing

We have argued that smoothing—suppressing high spatial frequencies—is a necessary part of image processing in at least two situations: An image needs to be smoothed before features can be extracted, and images must be smoothed before they are sampled. We have also mentioned that, while smoothing with the Gaussian is not a good candidate for the first situation (we will see that it is not contrast invariant, and it is not well localized except for small t), it is not unreasonable to use it numerically in the second situation, since it does a good job of suppressing high frequencies. These smoothing requirements and the fact that the Gaussian is the fundamental solution of the heat equation mean that the heat equation appears completely naturally in image processing, and indeed it is the first PDE to enter the picture in Chapters 1 and 2. Smoothing with the heat equation is illustrated in Figure I.6.

Figure I.6: Heat equation and smoothing. The original image is on the left; the heat equation has been applied at some scale, and the resulting blurred image is on the right.

There is another path hinted at in section I.1 that leads to the Gaussian and thus to the heat equation. Suppose that k is any positive kernel such that $k(\boldsymbol{x}) = k(|\boldsymbol{x}|)$ and such that k is localized in the sense that $k(\boldsymbol{x}) \to 0$ sufficiently rapidly as $|\boldsymbol{x}| \to \infty$. If k is normalized properly and if we write $k_h(\mathbf{x}) = (1/h)k(\mathbf{x}/h^{1/2})$, then

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

as $h \to 0$ whenever the image u_0 is sufficiently smooth. We write this as

$$k_h * u_0(\mathbf{x}) - u_0(\mathbf{x}) = h\Delta u_0(\mathbf{x}) + o(h). \tag{I.1}$$

Now let $u(t, \boldsymbol{x})$ denote the solution of the heat equation

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

If u_0 is sufficiently smooth, then we can write

$$u(t, \boldsymbol{x}) - u(0, \boldsymbol{x}) = t\Delta u_0(\mathbf{x}) + o(t). \tag{I.2}$$

The reverse heat equation

Equations (I.1) and (I.2) suggest that blurring u_0 with a kernel k_h for small h is equivalent to applying the heat equation to u_0 at some small scale t. This is true, and it will be made precise in Chapter 2. These equations also lead to another idea: We read in the paper [183] by Lindenbaum, Fischer, and Bruckstein that Kovasznay and Joseph [175] introduced in 1955 the notion that a slightly blurred image could be deblurred by subtracting a small amount of its Laplacian. Numerically, this amounts to subtracting a fraction λ of the Laplacian of the observed image from itself:

$$u_{\text{restored}} = u_{\text{observed}} - \lambda \Delta u_{\text{observed}}.$$

Dennis Gabor, who received the Nobel prize in 1971 for his invention of optical holography, studied this process and determined that the best value of λ was the one that doubled the steepest slope in the image [183]. Empirically, one can start with a small value of λ and repeat the process until a good image is obtained; with further repetitions the process blows up. Indeed, this process is just applying the reverse heat equation to the observed image, and the reverse heat equation is notoriously ill-posed. On the other hand, the Kovasznay–Joseph–Gabor method is efficient for sufficiently small λ and can be successfully applied to most images obtained from optical devices. This process is illustrated in Figure I.7. A few iterations can enhance the image (second panel), but the inverse heat equation finally blows up (third panel).

Figure I.7: Kovasznay–Joseph–Gabor deblurring. Left to right: original image; three iterations of the algorithm; ten iterations of the algorithm.

Figure I.8 shows that same experiment applied to an image of Victor that has been numerically blurred. Again, the process blows up, but it yields a significant improvement at some scales.

We have now seen the heat equation used in two senses, each with a different objective. In both cases, we have noted drawbacks. In the first instance, the heat equation (or Gaussian) was used to smooth an image, but as we have mentioned, this operator is not contrast invariant, and thus is not appropriate for any theory of image analysis that requires contrast-invariant operators. This does not mean that the Gaussian should be dismissed; it only means that it is not appropriate for our version of image analysis. To meet our objectives, we will replace the Laplacian, which is a linear isotropic operator, with nonlinear,

Figure I.8: Kovasznay–Joseph–Gabor deblurring. This is the same deblurring experiment as in Figure I.7, but it is applied to a much more blurred image.

nonisotropic smoothing operators. This will bring us to the central theme of the book: appropriate smoothing for a possible theory of image analysis.

In the second instance, the heat equation is run backward (the inverse heat equation) with the objective of restoring a blurred image. As we have seen, this is successful to some extent, but the drawback is that it is an unstable process. The practical problem is more complex than the fact that the inverse heat equation is not well posed. In the absence of noise, the best way to deblurr a slightly blurred image is to use the inverse heat equation. However, in the presence of noise, this isotropic operator acts equally in all direction, and while it enhances the definition of edges, the edges become jagged due to the noise. This observation led Gabor to try to improve matters by using more directional operators in place of the Laplacian. Gabor was concerned with image restoration, but his ideas will appear later in our story in connection with smoothing. (For an account of Gabor's work see [183].)

Shock filters

The objective for running the heat equation backward is image restoration, and although restoration is not the main subject of the book, we are going to pause here to describe two ways to improve the stability of the inverse heat equation. Image restoration is an extremely important area of image processing, and the techniques we describe illustrate another use of PDEs in image processing. There are indeed stable ways to "reverse" the heat equation. More precisely, there are "inverse diffusions" that deblurr an image and reach a steady state. The first example, due to Rudin in 1987 [241] and Osher and Rudin in 1990 [223] is a pseudoinverse for the heat equation, where the propagation term $|Du| = |(u_x, u_y)|$ is controlled by the sign of the Laplacian:

$$\frac{\partial u}{\partial t} = -\mathrm{sign}(\Delta u)|Du|. \tag{I.3}$$

This equation is called a *shock filter*. We will see later that this operator propagates the level lines of an image with a constant speed and in the same direction as the reverse heat equation would propagate these lines; hence it acts as a pseudoinverse for the heat equation. This motion enhances the definition of the contours and thus sharpens the image. Equation (I.3) is similar to a classic nonlinear filter introduced by Kramer in the seventies [176]. Kramer's filter can be interpreted in terms of a PDE using the same kinds of heuristic arguments that have been used to derive the heat equation. This equation is

$$\frac{\partial u}{\partial t} = -\text{sign}(D^2u(Du, Du))|Du|, \tag{I.4}$$

where the Laplacian has been replaced by

$$D^2u(Du, Du) = u_{xx}(u_x)^2 + 2u_{xy}u_xu_y + u_{yy}(u_y)^2. \tag{I.5}$$

We will see in Chapter 2 that $D^2u(Du, Du)/|Du|^2$ is the second derivative of u in the direction of its gradient Du, and we will interpret the differential operator (I.5) as Haralick's edge detector. Kramer's equation yields a slightly better version of a shock filter. The actions of these filters are illustrated in Figure I.9. The image on the left is a blurred image of Victor. The next image has been deblurred using the Rudin–Osher shock filter. This is a pseudoinverse of the heat equation that attains a steady state. The third image has been deblurred using Kramer's improved shock filter, which also attains steady state. The fourth image was deblurred using the Rudin–Osher–Fatemi restoration scheme, which is described below [242].

Figure I.9: Deblurring with shock filters and a variational method. Left to right: blurred image; Rudin–Osher shock filter; Kramer's improved shock filter; Rudin–Osher–Fatemi restoration method.

The deblurring algorithms (I.3) and (I.4) work to the extent that, experimentally, they attain steady states and do not blow up. However, a third deblurring method, the Rudin–Osher–Fatemi algorithm, is definitely better. It poses the deblurring problem as an inverse problem. It is very efficient when the observed image u_0 is of the form $k * u + n$, where k is known and where the statistics of the noise n are also known. Given the observed image u_0, one tries to find a restored version u such that $k * u$ is as close as possible to u_0 and such that the oscillation of u is nonetheless bounded. This is done by finding u that minimizes the functional

$$\int \big(|Du(\boldsymbol{x})| + \lambda(k * u(\boldsymbol{x}) - u_0(\boldsymbol{x}))^2\big)\, \mathrm{d}\mathbf{x}. \tag{I.6}$$

The parameter λ controls the oscillation in the restored version u. If λ is large, the restored version will closely satisfy the equation $k * u = u_0$, but it may be very oscillatory. If instead λ is small, the solution is smooth but inaccurate. This parameter can be computed in principle as a Lagrange multiplier. The obtained restoration can be remarkable. The best result we can obtain with the blurred Victor is shown in the fourth panel of Figure I.9. This scheme was selected by the French Space Agency (CNES) after a benchmark for satellite image deblurring, and it is currently being used by the CNES for satellite image restoration. This total variation restoration method also has fast wavelet packets versions.

From the heat equation to wavelets

The observation by Kovasznay, Joseph, and Gabor (and undoubtedly others) that the difference between a smoothed image and the original image is related to the Laplacian of the original image is also the departure of one of the paths that lead to wavelet theory. Here, very briefly, is the idea: If we convolve an image with an appropriate smoothing kernel and then take the difference, we obtain a new image related to the Laplacian of the original image (see equation (I.1)). This new "Laplacian image" turns out to be faded with respect to the original, and if one retains only the values greater than some threshold, the image is often sparse. This is illustrated in Figure I.10. The last panel on the right shows in black the values of this Laplacian image of Victor that differ significantly from zero. Here, and in most natural images, this representation is sparse and thus useful for compression. This experiment simulates the first step of a well-known algorithm due to Burt and Adelson.

In 1983, Burt and Adelson developed a compression algorithm called the Laplacian pyramid based on this idea [49]. Their algorithm consists of iterating two operations: a convolution followed by subsampling. After each convolution, one keeps only the difference $k_n * u_n - u_n$, where n is used here to indicate that each step takes place at a different scale due to the subsampling. The image is then coded by the (finite) sequence of these differences. These differences resemble the Laplacian of u_n, hence the name "Laplacian pyramid." An important aspect of this algorithm is that the discrete kernels k_n, which are low-pass filters, are all the same kernel k; the index n merely indicates that k is adjusted for the scale of the space where the subsampled image u_n lives. Ironically, the smoothing function cannot be the Gaussian, since the requirements for reconstructing the image from its coded version rule out the Gaussian. Burt and Adelson's algorithm turned out to be one of the key steps that led to multiresolution analyses and wavelets. Burt and Adelson were interested in compression, and, indeed, the differences $k_n * u_n - u_n$ tend to be sparse for natural images. On the other hand, we are interested in image analysis, and for us, the Burt and Adelson algorithm has the drawback that it is not translation invariant or isotropic because of the multiscale subsampling.

Back to edge detection

Early research in computer vision focused on *edge detection* as a main tool for image representation and analysis. It was assumed that the apparent contours of objects, and also the boundaries of the facets of objects, produce step discontinuities, while inside these boundaries, the image oscillates mildly. The apparent contour points, or *edges points*, were to be computed as points where the gradient is in some sense largest. Two ways were proposed to do this: Marr and Hildreth proposed computing the points where Δu crosses zero, the now-famous

Figure I.10: The Laplacian pyramid of Burt and Adelson. Left to right: the original image; the image blurred by Gaussian convolution; the difference between the original image and the blurred version, which approximates the Laplacian of the original image; the points where this Laplacian image is large.

zero-crossings [199]. A significant improvement was made by Harakick who defined the boundaries, or edges, of an image as those points where $|Du|$ attains a local maximum along the gradient lines [135]. Two years later, Canny implemented Haralick's detector in an algorithm that consists of Gaussian smoothing followed by computing the (edge) points where $D^2u(Du, Du) = 0$ and $|Du|$ is above some threshold [51]. We refer to this algorithm as the Haralick–Canny edge detector. The fourth panel in Figure I.11 displays what happens when we smooth the image with the Gaussian (the heat equation) and then compute the points where $D^2u(Du, Du) = 0$ and $|Du|$ is above some threshold. If this computation is done on the raw image (first panel), then "edges" show up everywhere (second panel) because the raw image is a highly oscillatory function and contains a very dense set of inflexion points. After applying the heat equation and letting it evolve to some scale (third panel), we see that the Haralick–Canny edge detector is able to extract some meaningful structure.

Figure I.11: Heat equation and Haralick's edge detector. Left to right: original image; edge points found in the original image using Haralick's detector; blurred image; edges found in the blurred image using the Haralick–Canny detector. The image "edges" are singled out after the image has been smoothed. This smoothing eliminates tiny oscillations and maintains the big ones.

The Perona-Malik equation

Given certain natural requirements such as isotropy, localization, and scale invariance, the heat equation is the only good linear smoothing operator. There are, however, many nonlinear ways to smooth an image. The first one was proposed by Perona and Malik in 1987 [231, 232]. Roughly, the idea is to smooth what needs to be smoothed, namely, the irrelevant homogeneous regions, and to enhance the boundaries. With this in mind, the diffusion should look like the heat equation when $|Du|$ is small, but it should act like the inverse heat equation when $|Du|$ is large. Here is an example of a Perona–Malik equation in divergence form:

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(|Du|)Du), \tag{I.7}$$

where $g(s) = 1/(1+\lambda^2 s^2)$. It is easily checked that we have a diffusion equation when $\lambda|Du| \leq 1$ and an inverse diffusion equation when $\lambda|Du| > 1$. To see this, consider the second derivative of u in the direction of Du,

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

and the second derivative of u in the orthogonal direction,

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

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 then becomes

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

The first term in this representation always appears as a one-dimensional diffusion in the direction orthogonal to the gradient, tuned by the size of the gradient. The nature of the second term depends on the value of the gradient; it can be either diffusion in the direction Du or inverse diffusion in the same direction. This model indeed mixes the heat equation and the reverse heat equation. Figure I.12 is used to compare the Perona–Malik equation with the classical heat equation (illustrated in Figure I.11) in terms of accuracy of the boundaries obtained by the Haralick–Canny edge detector (see Chapter 3). At a comparable scale of smoothing, we clearly gain some accuracy in the boundaries and remove more "spurious" boundaries using this Perona–Malik equation. The representation is both more sparse and more accurate.

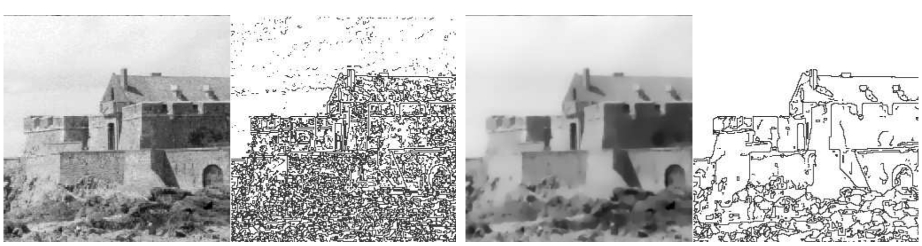

Figure I.12: A Perona–Malik equation and edge detection. This is the same experiment as in Figure I.11, but here the Perona–Malik equation is used in place of the heat equation. Notice that the edge map looks slightly better in this case.

The ambitious Perona–Malik model attempts to build into a single operator the ability to perform two very different tasks, namely, restoration and analysis. This has its cost: The model contains a "contrast threshold" λ^{-1} that must be set manually, and although experimental results have been impressive, the mathematical existence and uniqueness of solutions are not guaranteed, despite some partial results by Kichenassamy [165] and Weickert and Benhamouda [281]. There are three parameters involved in the overall smoothing and edge-detecting scheme: the gradient threshold λ^{-1} in the equation (3.2), the smoothing scale(s) t (or the time that equation (3.2) evolves), and the gradient threshold in the Haralick–Canny detector. We can use the same gradient threshold in both the Haralick–Canny detector and the Perona–Malik equation, but this still leaves us with a two-parameter algorithm. Can these parameters be dealt with automatically for an image analysis scheme? This question seems to have no general answer at present. An interesting attempt based on statistical arguments had been made, however, by Black et al. [40].

A proliferation of PDE's

If one believes that some nonlinear diffusion might be a good image analysis model, why not try them all? This is exactly what has happened during the last ten years. We can claim with some certainty that almost all possible nonlinear parabolic equations have been proposed. A few of the proposed models are even systems of PDEs. The common theme in this proliferation of models is this: Each attempt fixes one intrinsic diffusion direction and tunes the diffusion using the size of the gradient or the value of an estimate of the gradient. To keep the size of this introduction reasonable, we will focus on a few of the simplest models.

We begin with the Rudin–Osher–Fatemi model [242]. In this model the BV norm of u, $\int |Du(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}$, is one of the terms in the expression (I.6) that is minimized to obtain a restored image. It is this term that provides the smoothing. The gradient descent for $\int |Du(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}$ translates into the equation

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\frac{Du}{|Du|}\right) = \frac{1}{|Du|} u_{\eta\eta}.$$

Written this way, the method appears as a diffusion in the direction orthogonal to the gradient, tuned by the size of the gradient. Andreu et al. proved that this equation is well posed in the space BV of functions of bounded variation [18, 19]. A variant of this model was proposed independently by Alvarez, Lions, and Morel [12]. In this case, the relevant equation is

$$\frac{\partial u}{\partial t} = \frac{1}{|k * Du|}\,|Du| \operatorname{div}\left(\frac{Du}{|Du|}\right) = \frac{1}{|k * Du|} u_{\eta\eta},$$

and again the diffusion is in the direction $Du^{\perp}$ orthogonal to the gradient. Note that the rate of diffusion depends on the average value $k * Du$ of the gradient in a neighborhood of $\boldsymbol{x}$, whereas the direction of diffusion, $Du^{\perp}(\boldsymbol{x})/|Du(\boldsymbol{x})|$, depends on the value of $Du(\boldsymbol{x})$ at $\boldsymbol{x}$. The kernel k is usually the Gaussian. Kimia, Tannenbaum, and Zucker, working in a more general shape-analysis framework, proposed the simplest equation of our list [168]:

$$\frac{\partial u}{\partial t} = |Du| \operatorname{div}\left(\frac{Du}{|Du|}\right) = D^2 u\left(\frac{Du^{\perp}}{|Du|}, \frac{Du^{\perp}}{|Du|}\right) = u_{\eta\eta}. \qquad \text{(I.8)}$$

This equation had been proposed earlier in another context by Sethian as a tool for front-propagation algorithms [258]. This equation is a “pure” diffusion in the direction orthogonal to the gradient. We call this equation *the curvature equation*; this is to distinguish it from other equations that depend on the curvature of u in some other way. These latter will be called *curvature equations.* When we refer to the action of the equations, we often write *curvature motions* or *curvature-dependent motions.* (See Chapters 11 and 12.)

The Weickert equation can be viewed as a variant of the curvature equation [280]. It uses a nonlocal estimate of the direction orthogonal to the gradient for the diffusion direction. This direction is computed as the direction v of the eigenvector corresponding to the smallest eigenvalue of $k * (Du \otimes Du)$, where $(\boldsymbol{y} \otimes \boldsymbol{y})(\boldsymbol{x}) = (\boldsymbol{x} \cdot \boldsymbol{y})\boldsymbol{y}$. Note that if the convolution kernel is removed, then this eigenvector is simply $Du^{\perp}$. So the equation writes

$$\frac{\partial u}{\partial t} = u_{\eta\eta}, \tag{I.9}$$

where η denotes the coordinate in the direction v. The three models just described can be interpreted as diffusions in a direction orthogonal to the gradient (or an estimate of this direction), tuned by the size of the gradient. They are illustrated in Figure I.13. (The original image is in the first panel of Figure I.14.)

Carmona and Zhong proposed a diffusion in the direction of the eigenvector w corresponding to the smallest eigenvalue of D^2u [56]. So the equation is again 2.19, but this time η denotes the coordinate in the direction of w. This is illustrated in panel three of Figure I.14. Sochen, Kimmel, and Malladi propose instead a nondegenerate diffusion associated with a minimal surface variational formulation [260]. Their idea was to make a gradient descent for the area, $\int \sqrt{1+|Du(\boldsymbol{x})|^2}\,\mathrm{d}\boldsymbol{x}$, of the graph of u. This leads to the diffusion equation

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

At points where Du is large this equation behaves like $\frac{\partial u}{\partial t} = \operatorname{div}\left(\frac{Du}{|Du|}\right)$, where we retrieve the Rudin-Osher-Fatemi model of Section I.3. At points where Du is small we have $\frac{\partial u}{\partial t} = \operatorname{div}(Du)$ which is the heat equation. This equation is illustrated in panel four of Figure I.14. Other diffusions have also been considered. For purposes of interpolation, Caselles, Morel, and Sbert proposed a diffusion that may be interpreted as the strongest possible image smoothing [64],

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

This equation is not used for preprocessing the image as the others are; rather, it is a way to interpolate between the level lines of an image with sparse level lines (Figure I.15). Among the models mentioned, only the curvature motion proposed by Kimia, Tannenbaum, and Zucker was specifically introduced as a shape analysis tool. We are going to explain this, but to do so we must say more about image analysis.

Figure I.13: Diffusion models I. Left to right: Osher, Sethian 1988: the curvature equation; Rudin, Osher, Fatemi 1992: minimization of the image's total variation; Alvarez, Lions, Morel 1992: nonlocal variant of the preceding; Weickert 1994: nonlocal variant of the curvature equation. All of these models diffuse only in the direction orthogonal to the gradient, using a more or less local estimate of this direction. This explains why the results of the filters are so similar. However, the Weickert model captures better the texture direction.

Principles of image analysis

There are probably as many ways to approach image analysis as there are uses of digital images, and today the range of applications covers much of human activity. Most scientific and technical activities, including particularly medicine, and even sound analysis (visual sonograms), involve the perceptual analysis of images. Our goal is to look for fundamental principles that underlie most of these applications and to develop algorithms that are widely applicable. From a less lofty point of view, we wish to examine the collection of existing and potential image operators to determine which among them fit our vision of

Figure I.14: Diffusion models II. Left to right: original image; Perona–Malik equation 1987, creating blurry parts separated by sharp edges; Carmona, Zhong 1998 which actually blurs the whole image: diffusion along the least eigenvector of D^2u; Sochen, Kimmel, Malladi 1998: minimization of the image graph area. This last equation has effects similar to the Perona-Malik model.

Figure I.15: Diffusion models III. Left to right: original image; quantized image (only 10 levels are kept - 3.32 bits/pixel); the quantized image reinterpolated using the Caselles–Sbert algorithm 1998. They apply a diffusion on the quantized image with values on the remaining level lines taken as boundary conditions.

image analysis. Instead of examining an endless list of partial and specific requirements, we rely on a mathematical shortcut, well known in mechanics, that consists of stating a short list of invariance requirements. These invariance requirements will lead to a classification of models and point out the ones that are the most suitable as image analysis tools. The first invariance requirement is the Wertheimer principle according to which visual perception (and therefore, we add, image analysis) should be independent of the image contrast [287]. We formalize this as follows:

Contrast-invariant classes. Two images u and v are said to be (perceptually) equivalent if there is a continuous increasing function g such that $v = g(u)$. In this case, u and v are said to belong to the same contrast-invariant class. ("Increasing" always means "strictly increasing.")

Contrast invariance requirement. An image analysis operator T must act directly on the equivalence class. As a consequence, we ask that $T(g(u)) = g(Tu)$, which means that the image analysis operator commutes with contrast changes.

The contrast invariance requirement rules out the heat equation and all of the models described above except the curvature motion (I.8). Contrast invariance led Matheron in 1975 to formulate image analysis as set analysis, namely, the analysis of the level sets of an image. The *upper level set* of an image u at level λ is the set

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

We 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. if g is a continuous increasing contrast change, then

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

According to mathematical morphology, the image analysis doctrine founded by Matheron and Serra, the essential image shape information is contained in its level sets. It can be proved (Chapter 5) that an image can be reconstructed, up to a contrast change, from its set of level sets [202]. Figure I.16 shows an image and one of its level sets.

Figure I.16: An image and one of its level sets. On the right is 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 an image. Level sets are contrast invariant.

The contrast invariance requirement leads to powerful and simple denoising operators like the so-called *extrema killer, or area opening*, (Chapter 7) defined by Vincent in 1993 [276]. This image operator simply removes all connected components of upper and lower level sets with areas smaller than some fixed value. This operator is not a PDE; actually it's much simpler. Its effect is amazingly good for impulse noise, which includes the local destruction of the image and spots. The action of the extrema killer is illustrated in Figure I.17. The original image is in the first panel. In the third panel, the image has been degraded by adding "salt and pepper" noise to 75% of the pixels. The next panel shows its restoration using the extrema killer set to remove upper and lower level sets with areas smaller than 80 pixels. The second panel shows the result of the same operator applied to the original.

Level lines as a complete contrast invariant representation

In 1996, Caselles, Coll, and Morel further localized the contrast invariance requirement in image analysis. They proposed as the main objects of analysis the *level lines* of an image, that is, the boundaries of its level sets [61]. For this program—and the previous one involving level sets—to make sense, the levels sets and level lines must have certain topological and analytic properties. Level sets and isolevel sets $\{\mathbf{x} \mid u(\mathbf{x}) = \lambda\}$, which we would like to be the "level lines,"

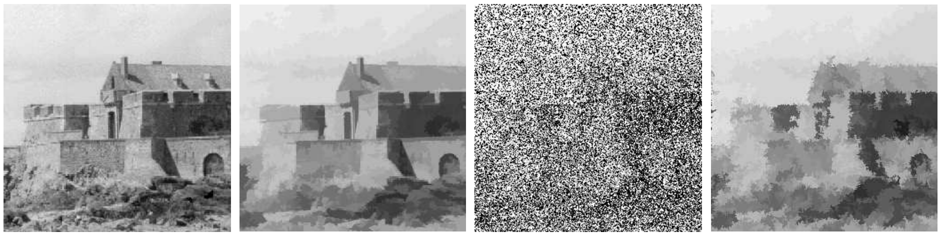

Figure I.17: The extrema killer filter. Left to right: original image; extrema killer applied with area threshold equal 80 pixels; 75% salt and pepper noise added to the original image; the same filter applied.

can be defined for any image (or function) u, but they will not necessarily be useful for image analysis. In particular, we cannot directly define useful level sets and level lines for a digital image u_d. What is needed is a representation of u_d for which these concepts make sense. But this is not a problem. By the assumptions of section I.1, a digital representation u_d of a natural image S has been obtained by suitably sampling a smooth version of S, call it u_f, and a smooth approximation of u_f is available to us by interpolation. There are, of course, different interpolation methods to produce smooth representations of u_d. One can also obtain a useful discontinuous representation by considering the extension of u_d that is constant on each pixel. For an interpolation method to be useful, the level lines should have certain minimal properties: They should be composed of a finite number of rectifiable Jordan curves, and they should be nested. This means that they do not cross, and thus that they form a tree by inclusion (Section 11.2.)

A study by Kronrod in 1950 shows that if the function u is continuous, then the isolevels sets $\{\mathbf{x} \mid u(\mathbf{x}) = \lambda\}$ are nested and thus form a tree when ordered by inclusion [178]. These isolevel sets are not necessarily curves; they are curves, however, if u has continuous first derivatives. Monasse proved Kronrod's result for lower semicontinuous and upper semicontinuous functions in 2000 [206] (see also [30]). His result implies that the extension of u_d that is constant on each pixel yields a nested set of Jordan curves bounding the pixels. Thus we have at least two ways to associate a set of nested Jordan curves with a digital image u_d, depending on how u_d is interpolated. Given an interpolation method, we call this set of nested curves a *topographic map* of the image.[1] By introducing the topographic map, the search for image smoothing, which had already been reduced to set smoothing, is further reduced to curve smoothing. Of course, we require that this smoothing preserves curve inclusion. Level lines of an image at a fixed level are shown in Figure I.18.

[1]The use of level lines is also consistent with the "BV assumption" mentioned in section I.1, according to which the correct function space for modeling images is the space BV of functions of bounded variation. In this case, the coarea formula can be used to associate a set of Jordan curves with an image (see [16]) It is, however, in general false for BV functions that the boundaries of lower and upper level sets form a nested set of curves; these curves may cross (see again [206].)

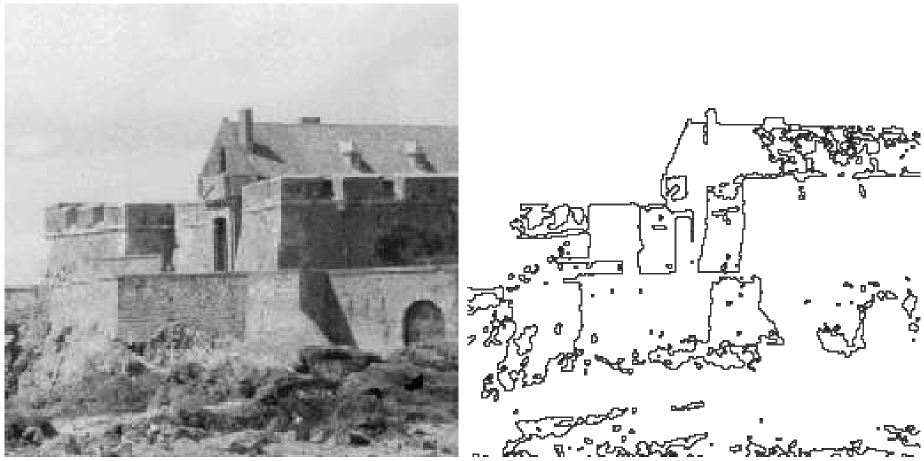

Figure I.18: Level lines of an image. Level lines, defined as the boundaries of level sets, can be defined to be a nested set of Jordan curves. They provide a contrast-invariant representation of the image. On the right are the level lines at level 183 of the left image.

Contrast invariant PDE's

Chen, Giga, and Goto [70, 71] and Alvarez et al. [11] proved that if one adds contrast invariance to the usual isotropic invariance requirement for image processing, then all multiscale image analyses should have a curvature-dependent motion of the form

$$\frac{\partial u}{\partial t} = F(\mathrm{curv}(u), t)|Du|, \tag{I.10}$$

where F is increasing with respect to its first argument (see chapters 21 and 22). This equation can be interpreted as follows: Consider a point $\mathbf{x}$ on a given level curve C of u at time t. Let $n(\mathbf{x})$ denote the unit vector normal to C at $\mathbf{x}$ and let $\mathrm{curv}(\mathbf{x})$ denote its curvature. Then the preceding equation is associated with the curve motion equation

$$\frac{\partial \mathbf{x}}{\partial t} = F(|\boldsymbol{\kappa}|(\mathbf{x}), t)n(\mathbf{x})$$

that describes how the point $\mathbf{x}$ moves in the direction of the normal. The formula defining $curv(u)$ at a point $\mathbf{x}$ is (Chapter 11)

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

The curvature vector at a point of a C^2 curve is the second derivative for a curve $\mathbf{x}(s)$ parameterized by length : $\boldsymbol{\kappa} = \mathrm{d}^2\mathbf{x}/\mathrm{d}s^2$. We refer to Chapter 11 for the detailed definitions and the links between the curvature vector of a level line of u and $curv(u)$. Not much more can be said at this level of generality about F. Two specific cases play prominent roles in this subject. The first case is $F(\mathrm{curv}(u), t) = \mathrm{curv}(u)$, the curvature equation (I.8). The second case is $F(\mathrm{curv}(u), t) = (\mathrm{curv}(u))^{1/3}$.

This particular one-third power form for the curvature dependence provides an important additional invariance, namely, affine invariance. We would like to have complete projective invariance, but a theorem proved by Alvarez et al. shows that this is impossible [11] (Chapter 22). The best we can have is invariance with respect to the so-called Chinese perspective, which preserves parallelism. Most of these equations, particularly when F is a power of the curvature, have a viscosity solution in the sense of Crandall and Lions [82]. This was shown in 1995 by Ishii and Souganidis [157]. We refer to Chapters 19 and 20 for all details.

As we have mentioned, contrast-invariant processing can be reduced to level set processing and, finally, to level curve processing. The equations mentioned above are indeed equivalent to curve evolution models if existence and regularity have been established. These results exist for the most important cases, namely, for $F(\mathrm{curv}(u),t) = \mathrm{curv}(u)$, called *curve shortening*, and for $F(\mathrm{curv}(u),t) = (\mathrm{curv}(u))^{1/3}$, known as *affine shortening*. Grayson proved existence, uniqueness, and analyticity for the curve shortening equation [128],

$$\frac{\partial \mathbf{x}}{\partial t} = \mathrm{curv}(\mathbf{x}) n(\mathbf{x}), \tag{I.11}$$

NE PAS LAISSER COMMME C'EST : *curv* n'est pas la meme notation qu'apres et n'est pas meme defini!

and Angenent, Sapiro, and Tannenbaum proved the same results for the affine shortening equation [21],

$$\frac{\partial \mathbf{x}}{\partial t} = (\mathrm{curv}(\mathbf{x}))^{\frac{1}{3}} n(\mathbf{x}). \tag{I.12}$$

These results are very important for image analysis because they ensure that the shortening processes do indeed reduce a curve to a more and more sketchy version of itself.

Affine invariance

An experimental verification of affine invariance for affine shortening is illustrated in Figure I.19. The numerical tests were made using a very fast numerical scheme for the affine shortening designed by Lionel Moisan [205]. The principle of this algorithm is explained in Chapter 16. Unlike many numerical schemes, this one is itself affine invariant. Each of the three panels in Figure I.19 contains three shapes. The first panel shows the action of an affine transformation A: Call the first shape in the first panel X; then the second shape is $A(X)$ and the third shape is $A^{-1}A(X) = X$. The second panel shows that affine shortening, S, commutes with A: The shapes are, from left to right, $S(X)$, $SA(X)$, and $A^{-1}SA(X)$. Since this third shape is the same as the first, we see that $A^{-1}SA(X) = S(X)$, or that $SA(X) = AS(X)$. The third panel shows the same experiment with affine shortening replaced with curve shortening. Since the first and third shapes are different, this illustrates that A does not commute with curve shortening, and hence that curve shortening is not affine invariant.

Evans and Spruck [99] (also [100, 101, 102]) and Chen, Giga, and Goto [70, 71] proved in 1991 that a continuous function moves by the curvature motion

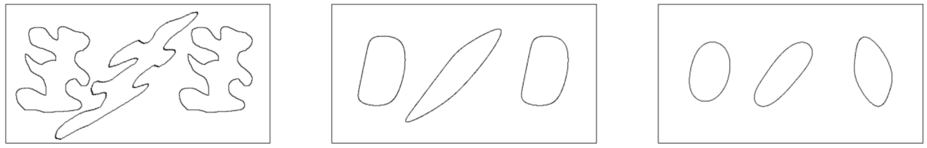

Figure I.19: Experimental verification of the affine invariance of the affine shortening (AMSS). The first panel contains three shapes, X, $A(X)$, and $A^{-1}A(X)$. The second panel contains $S(X)$, $SA(X)$, and $A^{-1}SA(X)$. The congruence of the first and third shapes implies that S and A commute. In the third panel, the same procedure has been applied using equation (I.11). Here the first and third shapes are not congruent, which shows that the curve shortening is not affine invariant, as expected.

(equation (I.10) with $F(\mathrm{curv}(u), t) = \mathrm{curv}(u)$) if and only if almost all of its level curves move by curve shortening (equation (I.11)). The same result is true for the affine invariant curve evolution (equation (I.10) with $F(\mathrm{curv}(u), t) = (\mathrm{curv}(u))^{1/3}$) and affine shortening (equation (I.12)).

In the case of the curvature motion, this result provides a mathematical justification for the now-classic Osher–Sethian numerical method for moving fronts [224]: They associate with some curve or surface C its signed distance function $u(\mathbf{x}) = \pm d(\mathbf{x}, C)$, and 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 classic numerical difference scheme. Thus, the evolution of the curve C is dealt with efficiently and accurately as a by-product of the evolution of u. The point of view that we adopt is slightly different from that of Osher and Sethian. We view the image as a generalized distance function to each of its level sets, since we are interested in all of them.

We show in Figure I.20 how the level lines are simplified by evolving the image numerically using affine invariant curvature motion. For clarity, we display only sixteen levels of level curves. Notice that the aim here is not subsampling; we keep the same resolution. Nor is the aim restoration; the processed image is clearly worse than the original. The aim is invariant simplification leading to shape recognition.

Figures I.21 and I.22 illustrate the effect of affine curvature motion on the values of the curvature of an image. In Figure I.21 the sea bird image has been smoothed by affine curvature motion at calibrated scale 1. In Figure I.22 the smoothing is stronger at calibrated scale 4. (A calibrated scale t means that at this scale a disk with radius t disappears.) The absolute values of the curvature of the smoothed images are shown in the upper-right panels of both figures, with the convention that the darkest points have the largest curvature. For clarity, the curvature is shown only at points where the gradient of the image was larger than 6 in a scale ranging from 0 to 255. Note how the density of points having large curvature is reduced in the second figure where the smoothing is stronger. On the other hand, the regions with large curvature are more concentrated with stronger smoothing. Each degree of smoothing produces a different curvature

Figure I.20: The affine and morphological scale space (AMSS model). Left to right: original image; level lines of this image (16 levels only); original image smoothed using the AMSS equation; level lines of the third image.

map of the original image, and thus curvature motions can be used as a nonlinear means to compute a "multiscale" curvature of the original image. The bottom two panels of the figures show, from left to right, the positive curvature and the negative curvature.

The snake method

Before proceeding to shape recognition, we mention that a variant of the curvature equation can be used for shape detection. This is a well-known method of contour detection, initially proposed by Kass, Witkin, and Terzopoulos [163]. Their method was very unstable. A better method is a variant of curvature motion proposed by Caselles, Catté, Coll, and Dibos [57] and improved simultaneously by Caselles, Kimmel, and Sapiro [62] and Malladi, Sethian, and Vemuri [191]. Here is how it works. The user draws roughly the desired contour in

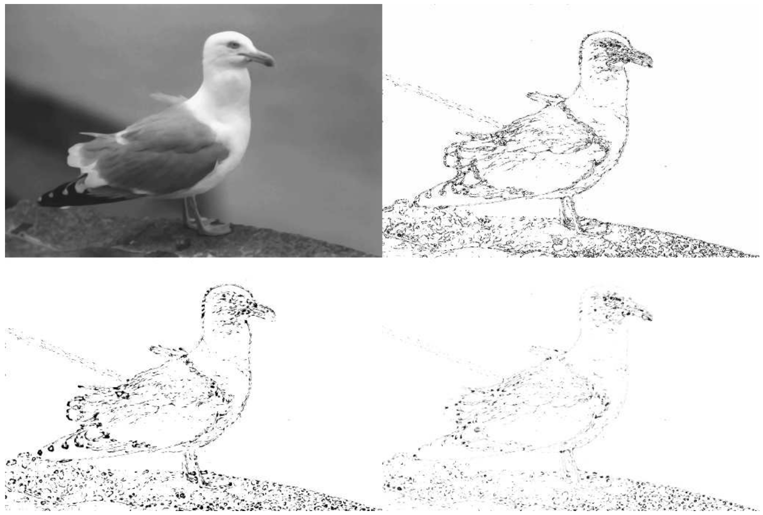

Figure I.21: Curvature scale space I. Top, left to right: original sea bird image smoothed by affine curvature motion at calibrated scale 1; the absolute value of the curvature. Bottom, left to right: the positive part of the curvature; the negative part. Compare with Figure I.22, where the calibrated smoothing scale is 4.

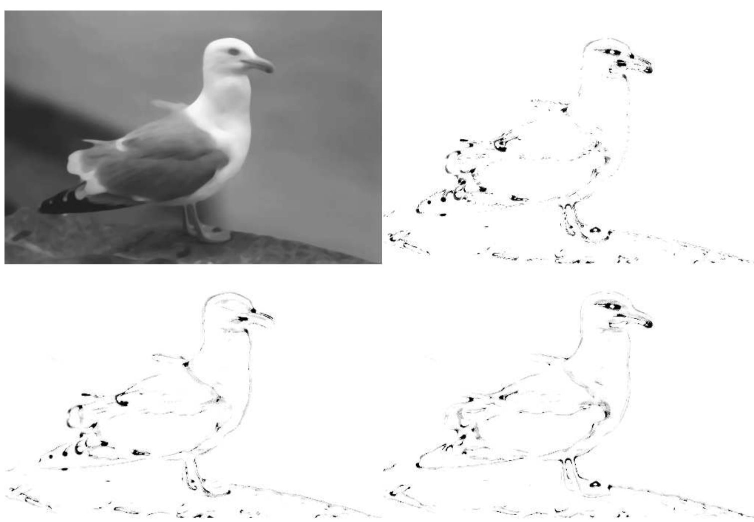

Figure I.22: Curvature scale space II. Top, left to right: original sea bird image smoothed by affine curvature motion at calibrated scale 4; the absolute value of the curvature. Bottom, left to right: the positive part of the curvature; the negative part. Compare with Figure I.21, where the calibrated smoothing scale is 1.

the image, and the algorithm then finds the best possible contour in terms of some variational criterion. This method is very useful in medical imaging. The motion of the contour is a tuned curvature motion that tends to minimize an energy function E. Given an original image u_0 containing some closed contour that we wish to approximate, we start with an edge map

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

that is, a function that vanishes on the edges of the image. The user then designates the contour of interest by drawing a polygon γ_0 roughly following 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

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

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

$$\frac{\partial v}{\partial t}(\boldsymbol{x}, t) = g(\boldsymbol{x})|Dv(\boldsymbol{x})|\mathrm{curv}(v)(\boldsymbol{x}) - Dv(\boldsymbol{x}) \cdot Dg(\boldsymbol{x}).$$

This algorithm is illustrated with a medical example in Figure I.23.

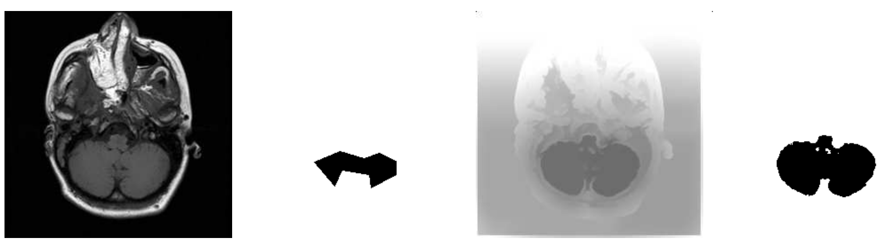

Figure I.23: Active contour, or "snake." Left to right: original image; initial contour; evolved distance function; final contour.

Shape retrieval

It seems to us that the most obvious application of invariant PDEs is shape retrieval in large databases. There are thousands of different definitions of shapes and a multitude of shape recognition algorithms. The real bottleneck has always been the ability to extract the relevant shapes. The discussion above points to a brute force strategy: All contrast-invariant local elements, or the level lines of the image, are candidates to be "shape elements." Of course, this notion 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 the level lines that are similar in both images? This would give a factual, a posteriori, definition of shapes. They would be defined as pieces of level lines common to two different images, irrespective of their relationships to real physical objects.

Of course, this brute force strategy would be impossible without the initial invariant filtering (AMSS). It is doable only if the level lines have been significantly simplified. This simplification entails the possibility of compressed invariant encoding. In Figure I.24, we present an experiment due to Lisani et al. [186]. Two images of a desk and the backs of chairs, viewed from different angles, are shown in the first two panels. All of the pieces of level lines in the two images that found a match in the other image are shown in the last two panels. Notice that several of these matches are doubled. Indeed, there are two similar chairs in each image. This brings to mind a Gestalt law that states that human perception tends to group similar shapes. We now see the numerical necessity of this perceptual grouping: A preliminary self-matching of each image, with grouping of similar shapes, must be performed before we can compare it with other images.

This concludes our overview of the use of PDEs in image analysis. The rest of the book is devoted to filling in the mathematical details that support most of the results mentioned in this introduction. We have tried to prove all of the mathematical statements, assuming only two or three years of mathematical training at the university level. Thus, for most of the PDEs addressed, and for all of the relevant ones, we prove the existence and uniqueness of solutions. We also develop invariant, monotone approximation schemes. This has been technically possible by combining tools from the recent, and remarkably simple, theory of viscosity solutions with the Matheron formalism for monotone set and function operators. Thus, the really necessary mathematical knowledge amounts to elementary differential calculus, linear algebra, and some results from the theory of Lebesgue integration, which are used in the chapters on the heat equation. Mathematical statements are not introduced as art for art's sake; all of the results are directed at proving the correctness of a model, of its properties, or of the associated numerical schemes. Numerical experiments, with detailed comments, are described throughout the text. They provide an independent development that is parallel to the central theoretical development. Most image processing algorithms mentioned in the text are accessible in the public software MegaWave. MegaWave was developed jointly by several university research groups in France, Spain and America, and it is available at `http://www.cmla.ens-cachan.fr`.

I.4 Notation and background material

$\mathbb{R}^N$ denotes the real N-dimensional Euclidian space. If $\mathbf{x} \in \mathbb{R}^N$ and $N > 2$, we write $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$; if $N = 2$, we usually write $\boldsymbol{x} = (x, y)$. For $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^N$, we denote their scalar product by $\boldsymbol{x} \cdot \boldsymbol{y} = x_1y_1 + x_2y_2 + \cdots + x_Ny_N$ and write

$$|\boldsymbol{x}| = (\boldsymbol{x} \cdot \boldsymbol{x})^{1/2} = (x_1^2 + x_2^2 + \cdots + x_N^2)^{1/2}.$$

Let Ω be an open set in $\mathbb{R}^N$, and let $n \in \mathbb{N}$ be a fixed integer. $C^n(\Omega)$ denotes the set of real-valued functions $f : \Omega \to \mathbb{R}$ that have continuous derivatives of all orders up to and including n. $f \in C^\infty(\Omega)$ means that f has continuous derivatives of all orders; $f \in C(\Omega) = C^0(\Omega)$ means that f is continuous on Ω. We will often write "f is C^n" as shorthand for $f \in C^n(\Omega)$, and we often omit the domain Ω if there is no chance of confusion.

We use multi-indices of the form $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N) \in \mathbb{N}^N$ as shorthand in several cases. For $\boldsymbol{x} \in \mathbb{R}^N$, we write $\boldsymbol{x}^\alpha$ and $|\boldsymbol{x}|^\alpha$ for $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_N^{\alpha_N}$ and $|x_1|^{\alpha_1} |x_2|^{\alpha_2} \cdots |x_N|^{\alpha_N}$, respectively. For $f \in C^n(\Omega)$, we abbreviate the partial

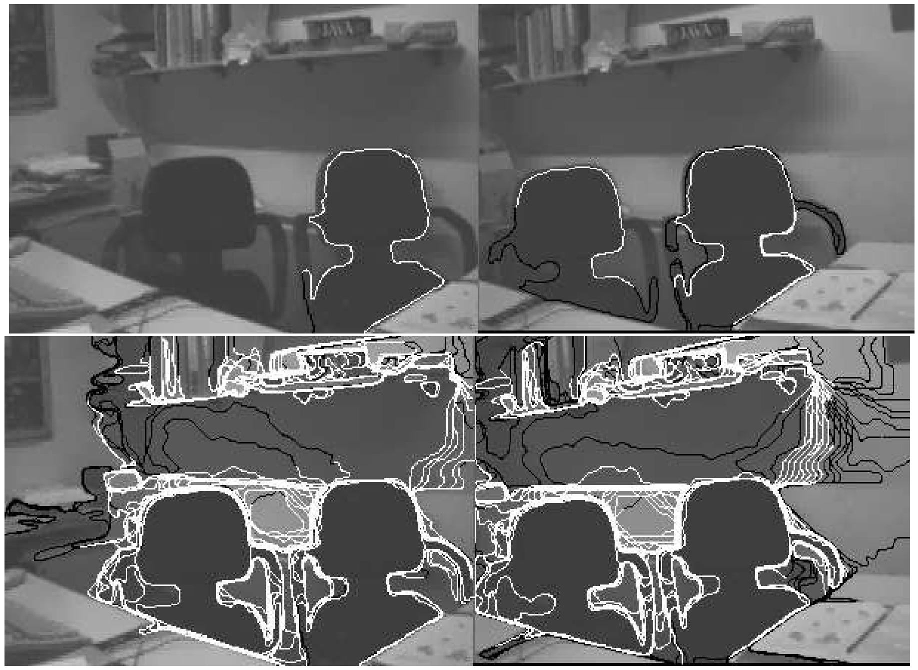

Figure I.24: A shape parser based on level lines. The two left images are of a desk and the backs of chairs viewed from different angles. In the far left panel, one level line has been selected (in white). In the second panel we show, also in white, all matching pieces of level lines. The match is ambiguous, as must be expected when the same object is repeated in the scene. In the two panels on the right, we display all the matching pairs of pieces of level lines (in white). The non matching parts of the same level lines are shown in black. Usually, recognized shape elements are pieces of level lines, seldom whole level lines. See []

derivatives of f by writing

$$\partial^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_N^{\alpha_N}},$$

where $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_N$ and $|\alpha| \le n$.

We also write the partial derivatives of $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_N)$ as $f_i = \partial f/\partial x_i$, $f_{ij} = \partial^2 f/\partial x_i \partial x_j$, and so on. In the two-dimensional case $f(\boldsymbol{x}) = f(x, y)$, we usually write $\partial f/\partial x = f_x$, $\partial f/\partial y = f_y$, $\partial^2 f/\partial x \partial y = f_{xy}$, and so on.

The gradient of f is denoted by Df. Thus, if $f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_N)$,

$$Df = (f_1, f_2, \ldots, f_N),$$

and

$$Df = (f_x, f_y)$$

in case $N = 2$. The Laplacian of f is denoted by Δf. Thus $\Delta f = f_{11} + f_{22} + \cdots + f_{NN}$ in general, and $\Delta f = f_{xx} + f_{yy}$ if $N = 2$.

We will often use the symbols O, o, and ε. They are defined as follows. We assume that h is a real variable that tends to a limit h_0 that can be finite or infinite. We assume that g is a positive function of h and that f is any other function of h. Then $f = O(g)$ means that there is a constant $C > 0$ such that $|f(h)| < Cg(h)$ for all values of h. The expression $f = o(g)$ means that $f(h)/g(h) \to 0$ as $h \to h_0$. We occasionally will use ε to denote a function of h that tends to zero as $h \to 0$. Thus, $f(h) = o(h)$ can be written equivalently as $f(h) = h\varepsilon(h)$.

Taylor's formula

An N-dimensional form of Taylor's formula is used several times in the book. We will first state it and then explain the notation. Assume that $f \in C^n(\Omega)$ for some open set $\Omega \in \mathbb{R}^N$, that $\boldsymbol{x}, \boldsymbol{y} \in \Omega$, and that the segment joining $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{y}$ is also in Ω. Then

$$f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + \frac{1}{1!}Df(\boldsymbol{x})\boldsymbol{y}^{(1)} + \frac{1}{2!}D^2f(\boldsymbol{x})\boldsymbol{y}^{(2)} + \cdots + \frac{1}{n!}D^nf(\boldsymbol{x})\boldsymbol{y}^{(n)} + o(|\boldsymbol{y}|^n).$$

This has been written compactly to resemble the one-dimensional case, but the price to be paid is to explain the meaning of $D^pf(\boldsymbol{x})\boldsymbol{y}^{(p)}$. We have already seen special cases of this expression in section I.3, for example, $D^2u(Du, Du)$ in equation (I.4). The expression $D^pf(\boldsymbol{x})\boldsymbol{y}^{(p)}$ is

$$D^pf(\boldsymbol{x})\boldsymbol{y}^{(p)} = D^pf(\boldsymbol{x})(\underbrace{\boldsymbol{y}, \boldsymbol{y}, \ldots, \boldsymbol{y}}_{p \text{ terms}}) = \sum_{(i_1, i_2, \ldots, i_p)} \frac{\partial^p f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_p}}(\mathbf{x}) y_{i_1} y_{i_2} \cdots y_{i_p},$$

where the sum is taken over all N^p different vectors $(i_1, i_2, \ldots, i_p)$, $i_j = 1, 2, \ldots, N$. Notice that $Df(\boldsymbol{x})\boldsymbol{y}^{(1)}$ is just $\sum_{j=1}^N f_j y_j = Df(\boldsymbol{x}) \cdot \boldsymbol{y}$, which is how we usually write it.

The implicit function theorem

Consider a real-valued C^1 function f defined on an open set Ω in $\mathbb{R}^N$. For ease of notation we write $\mathbf{z} = (\mathbf{x}, y)$, where $\mathbf{x} = (x_1, \ldots, x_{N-1})$ and $y = x_N$. Assume that $f(\mathbf{z}_0) = 0$ for a point $\mathbf{z}_0 \in \Omega$ and that $f_y(\mathbf{x}_0) \neq 0$. Then there is a neighborhood $M = M(\mathbf{x}_0)$ and a neighborhood $N = N(y_0)$ such that for every $\mathbf{x} \in M$ there is exactly one $y \in N$ such that $f(\mathbf{x}, y) = 0$. The function $y = \varphi(\mathbf{x})$ is C^1 on M and $y_0 = \varphi(\mathbf{x}_0)$. Furthermore, if $f \in C^n(\Omega)$, then $\varphi \in C^n(M)$.

Lebesgue integration

The Lebesgue integral, which first appeared in 1901 and is thus over a hundred years old, has become the workhorse of analysis. It plays a role in chapters 1 and 2 and appears briefly in other parts of the book. One does not need a profound understanding of abstract measure theory and integration to follow the arguments. One should, however, be familiar with a few key results and be comfortable with the basic manipulations of the integral. With this in mind, we restate some of these fundamentals.

The functions and sets in this book are always measurable. Thus we dispense in general with phrases like "let f be a measurable function." We denote by

$\mathcal{M}$ the set Lebesgue measurable subsets of $\mathbb{R}^N$. Since we shall sometimes need to complete $\mathbb{R}^N$ by a point at infinity, ∞, we still denote by $\mathcal{M}$ the measurable sets of $S_N = \mathbb{R}^N \cap \{\infty\}$ and take measure$(\{\infty\}) = 0$. A function f defined on a subset A of $\mathbb{R}^N$ is integrable, if

$$\int_A |f(\boldsymbol{x})|\,\mathrm{dx} < +\infty.$$

The Banach space of all integrable function defined on A is denoted as usual by $L^1(A)$; we write $\|f\|_{L^1(A)} = \int_A |f(\boldsymbol{x})|\,\mathrm{dx}$ to denote the norm of f in $L^1(A)$. The most important applications in the book are the two cases $A = \mathbb{R}^N$ and $A = [-1,1]^N$. Here are two results that we use in chapters 1 and 2. We state them not in the most general form, but rather in the simplest form suitable for our work.

A density theorem for $L^1(\mathbb{R}^N)$

If f is in $L^1(\mathbb{R}^N)$, then there exists a sequence of continuous functions $\{g_n\}$, each of which has compact support, such that $g_n \to f$ in $L^1(\mathbb{R}^N)$, that is, $\|g_n - f\| \to 0$ as $n \to +\infty$. This result is true for $L^1([-1,1]^N)$, in which case the g_n are continuous on $[-1,1]^N$.

Fubini's theorem

Suppose that f is a measurable function defined on $A \times B \in \mathbb{R}^N \times \mathbb{R}^N$. Fubini's theorem states that

$$\int_{A\times B} |f(\mathbf{z})|\,\mathrm{d}\boldsymbol{z} = \int_A \int_B |f(\mathbf{x},\mathbf{y})|\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} = \int_B \int_A |f(\mathbf{x},\mathbf{y})|\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x},$$

where we have written $\mathbf{z} = (\mathbf{x},\mathbf{y})$. It further states, that if any one of the integrals is finite, then

$$\int_{A\times B} f(\mathbf{z})\,\mathrm{d}\boldsymbol{z} = \int_B \int_A f(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} = \int_A \int_B f(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y}\,\mathrm{d}\mathbf{x}.$$

Lebesgue's dominated convergence theorem

If a sequence of functions $\{f_n\}$ is such that $f_n(\mathbf{x}) \to f(\mathbf{x})$ for almost every $\mathbf{x} \in \mathbb{R}^N$ as $n \to +\infty$, and if there is an integrable function g such that $|f_n(\mathbf{x})| \leq g(\mathbf{x})$ almost everywhere, then

$$\int_{\mathbb{R}^N} f_n(\mathbf{x})\,\mathrm{d}\mathbf{x} \to \int_{\mathbb{R}^N} f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

We often use the following direct consequence: if A_n is a decreasing sequence of measurable sets with bounded measure then $\mathrm{measure(A_n)} \mapsto \mathrm{measure(A)}$. To prove this, apply Lebesgue's theorem to the characteristic functions of A_n and A, $\mathbf{1}_{A_n}$ and $\mathbf{1}_A$.

We also use the following result, which is a direct consequence of the dominated convergence theorem.

Interchanging differentiation and integration

Suppose that a function f defined on $(t_0, t_1) \times \mathbb{R}^N$, where (t_0, t_1) is any interval of $\mathbb{R}$, is such that $t \mapsto f(t, \mathbf{x})$ is continuously differentiable (for almost every $\mathbf{x} \in \mathbb{R}^N$) on some interval $[a, b] \subset (t_0, t_1)$. If there exists an integrable function g such that for all $t \in [a, b]$

$$\left| \frac{\partial f}{\partial t}(t, \mathbf{x}) \right| \leq g(\mathbf{x}) \quad \text{almost everywhere,}$$

then the integral $I(t) = \int_{\mathbb{R}^N} f(t, \mathbf{x})\, \mathrm{d}\mathbf{x}$ is differentiable for $t \in (a, b)$ and

$$\frac{dI}{dt}(t) = \int_{\mathbb{R}^N} \frac{\partial f}{\partial t}(t, \mathbf{x})\, \mathrm{d}\mathbf{x}.$$

A brief but comprehensive discussion of the Lebesgue integral can be found in the classic textbook by Walter Rudin [243].

I.4.1 A framework for sets and images

We start by fixing a simple and handy functional framework for images and sets, which will be maintained throughout the book. Until now, we have been vague about the domain of definition of an image. On one hand, a real digital image is defined on a finite grid. On the other hand, standard interpolation methods give a continuous representation defined on a finite domain of $\mathbb{R}^N$, usually a rectangle. Now, it is convenient to have images defined on all of $\mathbb{R}^N$, but it is not convenient to extend them by making them zero outside their original domains of definition because that would make them discontinuous. So an usual way is to extend them into a continuous function tending to a constant at infinity. One way to do that is illustrated in Figure I.25. First, an extension to a wider domain is performed by reflection across the domain's boundary and periodization. Then, it is easy to let the function fade at infinity or to make it compactly supported. This also means that we fix a value at infinity for u, which we denote by $u(\infty)$. We denote the topological completion of $\mathbb{R}^N$ by this infinity point by $S_N = \mathbb{R}^N \cup \{\infty\}$, which can also be denoted $\overline{\mathbb{R}^N}$. Let us justify the notation.

Proposition 0.1. *Consider the sphere $S_N = \{\mathbf{z} \in \mathbb{R}^{N+1},\ ||\mathbf{z}|| = 1\}$. Then the mapping $T:\ \mathbb{R}^N \cup \{\infty\} \to S_N$ defined by*

$$T(\mathbf{x}) = \left(\frac{2\mathbf{x}}{1 + \mathbf{x}^2}, \frac{\mathbf{x}^2 - 1}{\mathbf{x}^2 + 1} \right)$$

is a homeomorphism (that is, a continuous bijection with continuous inverse.)

This is easily checked (Exercise 1.4).

Definition 0.2. *We denote by $\mathcal{F}$ the set of continuous functions on S_N, which can be identified with the set of continuous functions on $\mathbb{R}^N$ tending to some constant at infinity. The natural norm of $\mathcal{F}$ is*

$$\|u\|_{\mathcal{F}} = \sup_{\mathbf{x} \in \mathbb{R}^N} |u(\mathbf{x})|. \tag{I.13}$$

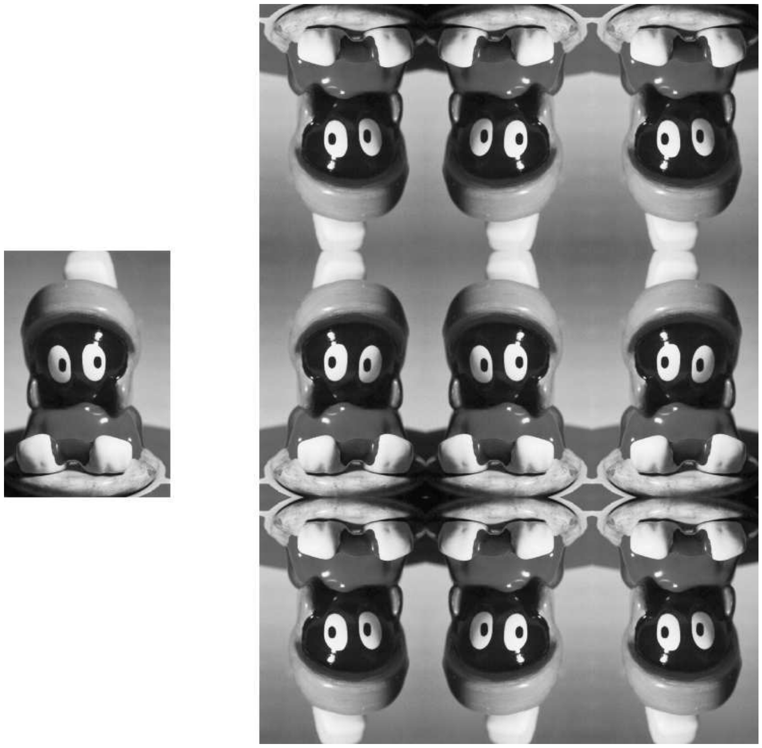

Figure I.25: Image extension by symmetry, followed by periodization. Then the image can be extended continuously to the rest of the plane into a function which is constant for $\mathbf{x}$ large. The purpose of these successive extensions of u to all of $\mathbb{R}^N$ is to facilitate the definition of certain operations on u, such as convolution with smoothing kernels, and, at the same time, to preserve the continuity of u. This method of extending a function is widely used in image processing; in particular, it is used in most compression and transmission standards. For instance, the discrete cosine transform (DCT) applied to the initial data u, restricted to $[0,1]^N$, is easily interpreted as an application of the FFT to the symmetric extension of u.

We say that an image u in $\mathcal{F}$ is C^1, if the function u is C^1 at each point $\mathbf{x} \in \mathbb{R}^N$. We define in the same way the C^2,... C^∞ functions of $\mathcal{F}$.

Definition 0.3. *We say that a function u defined on $\mathbb{R}^N$ is uniformly continuous if for every $\mathbf{x}$, $\mathbf{y}$,*

$$|u(\mathbf{x}+\mathbf{y}) - u(\mathbf{x})| \le \varepsilon(|\mathbf{y}|),$$

for some function ε called modulus of continuity of u, satisfying $\lim_{s\to 0} \varepsilon(s) = 0$.

Continuous functions on a compact set are uniformly continuous, so functions of $\mathcal{F}$ are uniformly continuous. We shall often consider the level sets of functions in $\mathcal{F}$, which simply are compact sets of S_N.

Definition 0.4. *We denote by $\mathcal{L}$ the set of all compact sets of S_N.*

These sets are easy to characterize:

Proposition 0.5. *The elements of $\mathcal{L}$ are of three kinds:*

- *compact subsets of* $\mathbb{R}^N$
- $F \cup \{\infty\}$, *where F is a compact set of* $\mathbb{R}^N$.
- $F \cup \{\infty\}$, *where F is an unbounded closed subset of* $\mathbb{R}^N$

Proof. Indeed, $B \cap \mathbb{R}^N$ is a closed set of $\mathbb{R}^N$ and is therefore either a bounded compact set or an unbounded closed set of $\mathbb{R}^N$. In the latter case, B must contain ∞. $\square$

Part I

Linear Image Analysis

Chapter 1

The Heat Equation

The heat equation is the prototype of all the PDEs used in image analysis. There are strong reasons for that and it is the aim of this chapter to explain some of them. Some more will be given in Chapter 21. Our first section is dedicated to a simple example of linear smoothing illustrating the relation between linear smoothing and the Laplacian. In the next section, we prove the existence and uniqueness of its solutions, which incidentally establishes the equivalence between the convolution with a Gaussian and the heat equation.

1.1 Linear smoothing and the Laplacian

Consider a continuous and bounded function u_0 defined on $\mathbb{R}^2$. If we wish to smooth u_0, then the simplest way to do so without favoring a particular direction is to replace $u_0(\mathbf{x})$ with the average of the values of u_0 in a disk $D(\mathbf{x}, h)$ of radius h centered at $\mathbf{x}$. This means that we replace $u_0(\mathbf{x})$ with

$$M_h u_0(\mathbf{x}) = \frac{1}{\pi h^2} \int_{D(\mathbf{x},h)} u_0(\mathbf{y})\,\mathrm{d}\mathbf{y} = \frac{1}{\pi h^2} \int_{D(0,h)} u_0(\mathbf{x}+\mathbf{y})\,\mathrm{d}\mathbf{y}. \tag{1.1}$$

Although the operator M_h is quite simple, it exhibits important characteristics of a general linear isotropic smoothing operator. For example, it is localizable: As h becomes small, M_h becomes more localized, that is, $M_h u_0(\mathbf{x})$ depends only on the values of $u_0(\mathbf{x})$ in a small neighborhood of $\mathbf{x}$. Smoothing an image by averaging over a small symmetric area is illustrated in Figure 1.1.

Our objective is to point out the relation between the action of M_h and the action of the Laplacian, or the heat equation. To do so, we assume enough regularity for u_0, namely that it is C^2. We shall actually prove in Theorem 2.2 that under that condition

$$M_h u_0(\mathbf{x}) = u_0(\mathbf{x}) + \frac{h^2}{8}\Delta u_0(\mathbf{x}) + h^2 \varepsilon(\mathbf{x}, h), \tag{1.2}$$

where $\varepsilon(\mathbf{x}, h)$ tends to 0 when $h \to 0$. As we have seen in the introduction, (1.2) provides the theoretical basis for deblurring an image by subtracting a small amount of its Laplacian. It also suggests that M_h acts as one step forward in the heat equation starting with initial condition u_0,

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \frac{1}{8}\Delta u(t, \mathbf{x}), \quad u(0, \mathbf{x}) = u_0(\mathbf{x}). \tag{1.3}$$

Figure 1.1: Local averaging algorithm. Left to right: original image; result of replacing the grey level at each pixel by the average of the grey levels over the neighboring pixels. The shape of the neighborhood is shown by the black spot displayed in the upper right-hand corner.

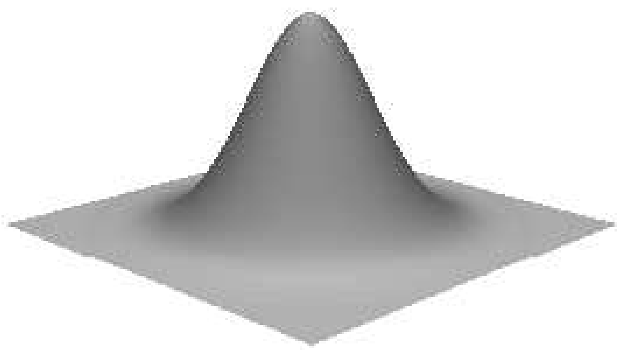

Figure 1.2: The Gaussian in two dimensions.

This statement is made more precise in Exercise 1.3. Equation (1.2) actually suggests that if we let $n \to +\infty$ and at the same time require that $nh^2 \to t$, then

$$(M_h^n u_0)(\mathbf{x}) \to u(t, \mathbf{x}) \tag{1.4}$$

where $u(t, x)$ is a solution of (1.3).

This heuristics justifies the need for a thorough analysis of the heat equation. The next chapter will prove that (1.4) is true under fairly general conditions. In the next section, we shall prove that the heat equation has a unique solution for a given continuous initial condition u_0, and that this solution at time t is equal to the convolution $G_t * u_0$, where G_t is the Gaussian (Figure 1.2). The effect on level lines of smoothing with the Gaussian is shown in Figure 1.4.

1.2 Existence and uniqueness of solutions of the heat equation

Definition 1.1. *We say that a function g defined on $\mathbb{R}^N$ belongs to the Schwartz class $\mathcal{S}$ if $g \in C^\infty(\mathbb{R}^N)$ and if 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 1.2. *If $g \in \mathcal{S}$, then $g \in L^1(\mathbb{R}^N)$, that is, $\int_{\mathbb{R}^N} |g(\mathbf{x})| d\mathbf{x} < +\infty$. For each pair of multi-indices α, β, the function $\mathbf{x}^\beta \partial^\alpha g$ also belongs to $\mathcal{S}$, and $\partial^\alpha g$ is uniformly continuous on $\mathbb{R}^N$.*

Proof. The second statement follows from the Leibnitz rule for differentiating a product. (A complete proof by induction is tedious but not profound.) By the definition of $\mathcal{S}$, there is a constant C such that $|\mathbf{x}|^{N+2}|g(\mathbf{x})| \leq C$. Thus there is another C such that $|g(\mathbf{x})| \leq C/(1 + |\mathbf{x}|^{N+2})$; since $C/(1 + |\mathbf{x}|^{N+2}) \in L^1(\mathbb{R}^N)$, $g \in L^1(\mathbb{R}^N)$. Finally, note that $|\partial^\alpha g(\mathbf{x})| \to 0$ as $|\mathbf{x}| \to \infty$. But any continuous function on $\mathbb{R}^N$ that tends to zero at infinity is uniformly continuous. □

Proposition 1.3 (The Gaussian and the heat equation). *For all $t > 0$, the function $\mathbf{x} \mapsto G_t(\mathbf{x}) = (1/(4\pi t)^{N/2})\mathrm{e}^{-|\mathbf{x}|^2/4t}$ belongs to $\mathcal{S}$ and satisfies the heat equation*

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

Proof. It is sufficient to prove the first statement for the function $g(\mathbf{x}) = \mathrm{e}^{-|\mathbf{x}|^2}$. An induction argument shows that $\partial^\alpha g(\mathbf{x}) = P_\alpha(\mathbf{x})\mathrm{e}^{-|\mathbf{x}|^2}$, where $P_\alpha(\mathbf{x})$ is a polynomial of degree $|\alpha|$ in the variables $x_1, x_2, \dots, x_N$. The fact that, for every $k \in \mathbb{N}$, $x^k \mathrm{e}^{-x^2} \to 0$ as $|x| \to +\infty$ finishes the proof. Differentiation shows that G_t satisfies the heat equation. □

Exercise 1.1. Check that G_t is solution of the heat equation. ∎

Linear image filtering is mainly done by convolving an image u with a positive integrable kernel g. This means that the smoothed image is given by the function $g * u$ defined as

$$g * u(\mathbf{x}) = \int_{\mathbb{R}^N} g(\mathbf{x} - \mathbf{y})u(\mathbf{y})\, \mathrm{d}\mathbf{y} = \int_{\mathbb{R}^N} g(\mathbf{y})u(\mathbf{x} - \mathbf{y})\, \mathrm{d}\mathbf{y}.$$

Note that the convolution, when it makes sense, is translation invariant. This means that $g * u(\mathbf{x} - \mathbf{z}) = g_\mathbf{z} * u(\mathbf{x})$, where $g_\mathbf{z}(\mathbf{x}) = g(\mathbf{x} - \mathbf{z})$. (Linear filtering with the Gaussian at several scales is illustrated in Figure 1.3.) The next result establishes properties of the convolution that we need for our treatment of the heat equation.

Proposition 1.4. *Assume that $u \in \mathcal{F}$ and that $g \in L^1(\mathbb{R}^N)$. Then the function $g * u$ belongs to $\mathcal{F}$ and satisfies the inequality*

$$\|g * u\|_\mathcal{F} \leq \|g\|_{L^1(\mathbb{R}^N)} \|u\|_\mathcal{F}. \tag{1.5}$$

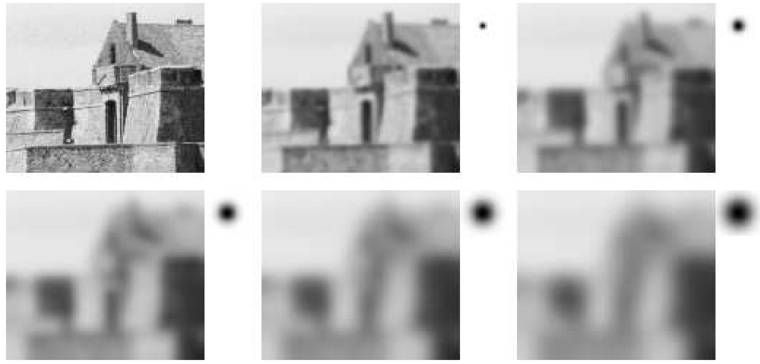

Figure 1.3: Convolution with Gaussian kernels (heat equation). Displayed from top-left to bottom-right are the original image and the results of convolutions with Gaussians of increasing variance. A grey level representation of the convolution kernel is put on the right of each convolved image to give an idea of the size of the involved neighborhood.

Proof.

$$|g * u(\mathbf{x})| \leq \int_{\mathbb{R}^N} |g(\mathbf{x}-\mathbf{y})||u(\mathbf{y})|\,\mathrm{d}\mathbf{y} \leq \|u\|_{\mathcal{F}} \int_{\mathbb{R}^N} |g(\mathbf{x}-\mathbf{y})|\,\mathrm{d}\mathbf{y} = \|u\|_{\mathcal{F}} \|g\|_{L^1_{\mathbb{R}^N}}.$$

□

Exercise 1.2. Verify that $g * u$ indeed is continuous and tends to $u(\infty)$ at infinity : this a direct application of Lebesgue Theorem. ■

We are now going to focus on kernels that, like the Gaussian, belong to $\mathcal{S}$.

Proposition 1.5. *If $u \in \mathcal{F}$ and $g \in \mathcal{S}$, then $g * u \in C^\infty(\mathbb{R}^N) \cap \mathcal{F}$ and*

$$\partial^\alpha (g * u) = (\partial^\alpha g) * u \tag{1.6}$$

for every multi-index α.

Proof. Since $g \in \mathcal{S}$, $g \in L^1(\mathbb{R}^N)$, as is $\partial^\alpha g$ for any multi-index α (Proposition 1.2). Thus by Proposition 1.4, $g * u$ belongs to $\mathcal{F}$. To prove (1.6), it is sufficient to prove it for $\alpha = (1, 0, \dots, 0)$. Indeed, we know that $\partial^\alpha g$ is in $\mathcal{S}$ if g is in S, so the general case follows from the case $\alpha = (1, 0, \dots, 0)$ by induction. Letting $\boldsymbol{e}_1 = (1, 0, \dots, 0)$ and using Taylor's formula with Lagrange's form for the remainder, we can write

$$\begin{aligned} g * u(\mathbf{x} + h\boldsymbol{e}_1) - g * u(\mathbf{x}) &= \int_{\mathbb{R}^N} (g(\mathbf{x} + h\boldsymbol{e}_1 - \mathbf{y}) - g(\mathbf{x}-\mathbf{y}))u(\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &= \int_{\mathbb{R}^N} (g(\mathbf{y} + h\boldsymbol{e}_1) - g(\mathbf{y}))u(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &= h\int_{\mathbb{R}^N} \frac{\partial g}{\partial x_1}(\mathbf{y})u(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y} \\ &\quad + \frac{h^2}{2}\int_{\mathbb{R}^N} \frac{\partial^2 g}{\partial x_1^2}(\mathbf{y} + \theta(\mathbf{y})h\boldsymbol{e}_1)u(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y}, \end{aligned} \tag{1.7}$$

where $0 \leq \theta(\mathbf{y}) \leq 1$. To complete the proof, we wish to have a bound on the last integral that is independent of $\mathbf{x} \in C$. This last integral is of the form $f * u$, where f is defined by $f(\mathbf{y}) = (\partial^2 g/\partial x_1^2)(\mathbf{y} + \theta(\mathbf{y})h\boldsymbol{e}_1)$. Since $g \in \mathcal{S}$, $\partial g/\partial x_1 \in \mathcal{S}$, and from this it is a simple computation to show that f decays rapidly at infinity. Having done this, Proposition 1.4 applies, and we deduce that $g * u$ is differentiable in x_1 and that $\partial(g * u)/\partial x_1 = (\partial g/\partial x_1) * u$. □

Proposition 1.6. *Assume that g decreases rapidly at infinity, that $g(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^N$, and that $\int_{\mathbb{R}^N} g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ and set, for $t > 0$, $g_t(\mathbf{x}) = (1/t^N)g(\mathbf{x}/t)$. Then: If $u_0 \in \mathcal{F}$, $g_t * u_0$ converges to u_0 uniformly as $t \to 0$. In addition, we have a maximum principle :*

$$\inf_{\mathbf{x}\in C} u_0(\mathbf{x}) \leq g_t * u_0(\mathbf{x}) \leq \sup_{\mathbf{x}\in C} u_0(\mathbf{x}). \tag{1.8}$$

Proof. Note first that g_t is normalized so that

$$\int_{\mathbb{R}^N} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y} = 1. \tag{1.9}$$

Next, since g decreases rapidly at infinity, a quick computation shows that, for any $\eta > 0$,

$$\int_{|\mathbf{y}|\geq\eta} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y} \to 0 \ \text{ as } \ t \to 0. \tag{1.10}$$

Using (1.9), we have

$$g_t * u_0(\mathbf{x}) - u_0(\mathbf{x}) = \int_{\mathbb{R}^N} g_t(\mathbf{y})(u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{x}))\,\mathrm{d}\mathbf{y}. \tag{1.11}$$

As already mentioned, $u_0 \in \mathcal{F}$ is uniformly continuous. Thus, for any $\varepsilon > 0$, there is an $\eta = \eta(\varepsilon) > 0$ such that $|u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{x})| \leq \varepsilon$ when $|\mathbf{y}| \leq \eta$. Using this inequality, we have

$$\begin{aligned}
|g_t * u_0(\mathbf{x}) - u_0(\mathbf{x})| \leq &\int_{|\mathbf{y}|<\eta} g_t(\mathbf{y})|u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{x})|\,\mathrm{d}\mathbf{y} \\
&+ \int_{|\mathbf{y}|\geq\eta} g_t(\mathbf{y})|u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{x})|\,\mathrm{d}\mathbf{y} \\
\leq &\varepsilon \int_{|\mathbf{y}|<\eta} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y} + 2\|u\|_{L^\infty(C)} \int_{|\mathbf{y}|\geq\eta} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y}.
\end{aligned}$$

Since $\int_{|\mathbf{y}|<\eta} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y} \leq 1$ and $\int_{|\mathbf{y}|\geq\eta} g_t(\mathbf{y})\,\mathrm{d}\mathbf{y} \to 0$ as $t \to 0$, we conclude that $g_t * u$ tends to u_0 uniformly in $\mathbf{x}$ as $t \to 0$. Relation (1.8) is an immediate consequence of the assumption that $g_t(\mathbf{x}) \geq 0$ and equation (1.9). □

Lemma 1.7. *Let $u_0 \in \mathcal{F}$ and $u(t,\mathbf{x}) = (G_t * u_0)(\mathbf{x})$. Then for every $t_0 > 0$, $u(t,\mathbf{x}) \to u_0(\infty)$ uniformly for $t \leq t_0$ as $\mathbf{x} \to \infty$.*

Proof. By assumption,

$$\forall \varepsilon > 0,\ \exists R,\ |\mathbf{x}| \geq R \Rightarrow |u_0(\mathbf{x}) - u_0(\infty)| < \varepsilon. \tag{1.12}$$

As a direct consequence of Lebesgue's theorem,

$$\forall \varepsilon > 0,\ \exists r(\varepsilon),\ r \geq r(\varepsilon) \Rightarrow \int_{|\mathbf{y}|\geq r} G_{t_0}(\mathbf{y})d\mathbf{y} < \varepsilon. \tag{1.13}$$

By using $\int G_t(\mathbf{y})d\mathbf{y} = 1$, we have

$$|u(t,\mathbf{x}) - u(\infty)| \leq \int_{|\mathbf{y}|\leq r} G_t(\mathbf{y})|u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{y})|d\mathbf{y} + \int_{|\mathbf{y}|\geq r} G_t(\mathbf{y})|u_0(\mathbf{x}-\mathbf{y}) - u_0(\mathbf{y})|d\mathbf{y}. \tag{1.14}$$

Using (1.13), the second term in (1.14) is bound from above for $r \geq r(\varepsilon)$ and $t \leq t_0$ by

$$(2\sup|u_0|)\int_{|\mathbf{y}|\geq r} G_{t_0}(\mathbf{y}) \leq (2\sup|u_0|)\varepsilon.$$

Fix therefore $r \geq r(\varepsilon)$. Then using $\int G_t = 1$, the first term in (1.14) is bound by ε by (1.12) for $|\mathbf{x}| \geq R + r$.

□

Lemma 1.8. *Let $u_0 \in \mathcal{F}$ and G_t the gaussian. Then*

$$(\partial G_t/\partial t) * u_0 = \partial(G_t * u_0)/\partial t.$$

Proof. Proposition 1.5 does not apply directly, since it applies to the spatial partial derivatives of G_t but not to the derivative with respect to t. Observe, however, that a slight modification of the proof of this proposition does the job: Replace g with G_t and $\mathbf{x}_1$ with t. Then the crux of the matter is to notice that, given an interval $0 < t_0 < t_1$, there is a rapidly decreasing function f such that $|(\partial^2 G_t/\partial t^2)(t + \theta(t)h, \mathbf{y})| \leq f(\mathbf{y})$ uniformly for $t \in [t_0, t_1]$, where f depends on t_0 and t_1 but not on h. Then Proposition 1.4 applies, and the last integral in equation (1.7) is uniformly bounded. □

All of the tools are in place to state and prove the main theorem of this chapter.

Theorem 1.9 (Existence and uniqueness of solutions of the heat equation). *Assume that $u_0 \in \mathcal{F}$ and define for $t > 0$ and $\mathbf{x} \in \mathbb{R}^N$, $u(t,\mathbf{x}) = (G_t * u_0)(\mathbf{x})$, $u(t,\infty) = u_0(\infty)$ and $u(0,\mathbf{x}) = u_0(\mathbf{x})$. Then*

(i) *u is C^∞ and bounded on $(0, +\infty) \times \mathbb{R}^N$;*

(ii) *$\mathbf{x} \to u(t,\mathbf{x})$ belongs to $\mathcal{F}$ for every $t \geq 0$;*

(iii) *for any $t_0 \geq 0$, $u(t,\mathbf{x})$ tends uniformly for $t \leq t_0$ to $u(\infty)$ as $\mathbf{x} \to \infty$;*

(iv) *$u(t,\mathbf{x})$ tends uniformly to $u(0,\mathbf{x})$ as $t \to 0$;*

(v) $u(t,\mathbf{x})$ *satisfies the heat equation with initial value* u_0*;*

$$\frac{\partial u}{\partial t} = \Delta u \quad \textit{and} \quad u(0,\mathbf{x}) = u_0(\mathbf{x}); \tag{1.15}$$

(vi) *More specifically,*

$$\sup_{\mathbf{x}\in\mathbb{R}^N,\ t\geq 0} |u(t,\mathbf{x})| \leq \|u_0\|_{\mathcal{F}}. \tag{1.16}$$

Conversely, given $u_0 \in \mathcal{F}$, $u(t,\mathbf{x}) = (G_t * u_0)(\mathbf{x})$ *is the only* C^2 *bounded solution* u *of* (1.15) *that satisfies properties (ii)-(v).*

Proof. Let us prove properties (i)-(vi). For each $t > 0$, $G_t \in \mathcal{S}$, so by Proposition 1.5 and Lemma 1.8,

$$\frac{\partial u}{\partial t} - \Delta u = u * \left(\frac{\partial G_t}{\partial t} - \Delta G_t\right). \tag{1.17}$$

Proposition 1.5 also tells us that $u(t,\cdot) \in C^\infty(\mathbb{R}^N) \cap \mathcal{F}$ for each $t > 0$. The right-hand side of (1.17) is zero by Proposition 1.3, and the fact that $|u(t,\mathbf{x}) - u_0(\mathbf{x})| \to 0$ uniformly as $t \to 0$ follows from Proposition 1.6. The inequality (1.16) is a direct application of Proposition 1.4. Relation (iii) comes from Lemma 1.7.

Uniqueness proof. If both v and w are solutions of the heat equation with the same initial condition $u_0 \in \mathcal{F}$, then $u = v - w$ is in $\mathcal{F}$ and satisfies (1.15) with the initial condition $u_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^N$. Also, by the assumptions of (ii), u is bounded on $[0,+\infty) \times \mathbb{R}^N$ and is C^2 on $(0,+\infty) \times \mathbb{R}^N$. We wish to show that $u(t,\mathbf{x}) = 0$ for all $t > 0$ and all $\mathbf{x} \in \mathbb{R}^N$. Assume that this is not the case. Then there is some point $(t,\mathbf{x})$ where $u(t,\mathbf{x}) \neq 0$. Assume that $u(t,\mathbf{x}) > 0$, by changing u to $-u$ if necessary.

We now consider the function u^ε defined by $u^\varepsilon(t,\mathbf{x}) = \mathrm{e}^{-\varepsilon t}u(t,\mathbf{x})$. This function tends to zero uniformly in $\mathbf{x}$ as $t \to 0$ and as $t \to +\infty$. It also tends uniformly to zero for each $t \leq t_0$ when $\mathbf{x} \to \infty$. These conditions imply that u^ε attains its supremum at some point $(t_0,\mathbf{x}_0) \in (0,+\infty) \times \mathbb{R}^N$, and this means that $\Delta u^\varepsilon(t_0,\mathbf{x}_0) = \mathrm{e}^{-\varepsilon t}\Delta u(t_0,\mathbf{x}_0) \leq 0$ and $(\partial u^\varepsilon/\partial t)(t_0,\mathbf{x}_0) = 0$. Here is the payoff: Using the fact that u is a solution of the heat equation, we have the following relations:

$$\begin{aligned} 0 = \frac{\partial u^\varepsilon}{\partial t}(t_0,\mathbf{x}_0) &= -\varepsilon u^\varepsilon(t_0,\mathbf{x}_0) + \mathrm{e}^{-\varepsilon t}\frac{\partial u}{\partial t}(t_0,\mathbf{x}_0) \\ &= -\varepsilon u^\varepsilon(t_0,\mathbf{x}_0) + \mathrm{e}^{-\varepsilon t}\Delta u(t_0,\mathbf{x}_0) \leq -\varepsilon u^\varepsilon(t_0,\mathbf{x}_0) < 0. \end{aligned}$$

This contradiction completes the uniqueness proof. □

1.3 Exercises

Exercise 1.3. The aim of this exercise is to prove relation (1.2) and its consequence: A local average is equivalent to one step forward of the heat equation. Theorem 2.2 yields actually a more general statement.

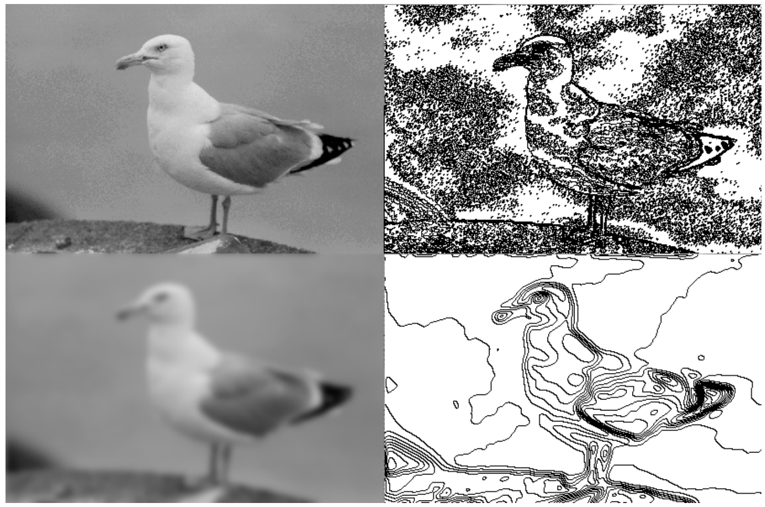

Figure 1.4: Level lines and the heat equation. Top, left to right: original 410×270 grey level image; level lines of original image for levels at multiples of 12. Bottom, left to right: original image smoothed by the heat equation (convolution with the Gaussian). The standard deviation of the Gaussian is 4, which means that its spatial range is comparable to a disk of radius 4. The image gets blurred by the convolution, which averages grey level values and removes all sharp edges. This can be appreciated on the right, where we have displayed all level lines for levels at multiples of 12. Note how some level lines on the boundaries of the image have split into parallel level lines that have drifted away from each other. The image has become smooth, but it is losing its structure.

1) Expanding u_0 around the point $\mathbf{x}$ using Taylor's formula, write

$$u_0(\mathbf{x}+\mathbf{y}) = u_0(\mathbf{x}) + Du_0(\mathbf{x})\cdot\mathbf{y} + \frac{1}{2}D^2u_0(\mathbf{x})(\mathbf{y},\mathbf{y}) + o(|\mathbf{y}|^2). \tag{1.18}$$

Expand the various terms using the coordinates (x, y) of $\mathbf{x}$.

2) Apply M_h to both sides of this expansion and deduce relation (1.2).

3) Assume $u_0 \in \mathcal{F}$ and consider the solution $u(t,\mathbf{x})$ of the heat equation (1.3) Then, for fixed $t_0 > 0$ and $\mathbf{x}$, apply M_h to the function $u^{t_0} : \mathbf{x} \to u(t_0,\mathbf{x})$ and write equation (1.2) for u^{t_0}. Using that $u(t,\mathbf{x})$ is a solution of the heat equation and its Taylor expansion between t_0 and $t_0 + h$, deduce that

$$M_h u(t_0,\mathbf{x}) = u(t_0 + h^2,\mathbf{x}) + h^2\varepsilon(t_0,\mathbf{x},h). \tag{1.19}$$

■

Exercise 1.4. Consider the sphere $S_N = \{\mathbf{z} \in \mathbb{R}^{N+1},\ ||\mathbf{z}|| = 1\}$. Prove that the mapping $T : \mathbb{R}^N \cup \{\infty\} \to S_N$ defined in Proposition 0.1 by

$$T(\mathbf{x}) = \left(\frac{2\mathbf{x}}{1+\mathbf{x}^2}, \frac{\mathbf{x}^2-1}{\mathbf{x}^2+1}\right),\ T(\infty) = (0,1).$$

is a homeomorphism.

Exercise 1.5. A natural norm for $\mathcal{F} \cap C^1$ is

$$\|u\|_{\mathcal{F} \cap C^1} = \sup_{\mathbf{x} \in \mathbb{R}^N} |u(\mathbf{x})| + |Du(\mathbf{x})|. \tag{1.20}$$

Prove that $\mathcal{F} \cap C^1$ is complete, namely that if $u_n \to u$ for the preceding norm, then $u(\mathbf{x})$ tends to a constant and $Du(\mathbf{x})$ tends to zero as $|\mathbf{x}|$ tends to infinity.

Exercise 1.6. Let u_0 be a continuous function defined on $\mathbb{R}^N$ having the property that there exist a constant $C > 0$ and an integer k such that

$$|u_0(\mathbf{x})| \leq C(1 + |\mathbf{x}|^k)$$

for all $\mathbf{x} \in \mathbb{R}^N$. Show that the function u defined by $u(t, \mathbf{x}) = G_t * u_0(\mathbf{x})$ is well defined and C^∞ on $(0, \infty) \times \mathbb{R}^N$ and that it is a classical solution of the heat equation. Hints: Everything follows from the fact that the Gaussian and all of its derivatives decay exponentially at infinity. ∎

Exercise 1.7. We want to prove the general principle that any linear, translation invariant and continuous operator T is a convolution, that is $Tu = g * u$ for some kernel g. This is one of the fundamental principles of both mechanics and signal processing, and it has many generalizations that depend on the domain, range, and continuity properties of T. For instance, assume that T is translation invariant (commutes with translations) and is continuous from $L^2(\mathbb{R}^N)$ into $L^\infty(\mathbb{R}^N) \cap C^0(\mathbb{R}^N)$. Show that $Tu = g * u$, where the convolution kernel g is in $L^2(\mathbb{R}^N)$. This is a direct consequence of Riesz theorem, which states that every bounded linear functional on $L^2(\mathbb{R}^N)$ has the form $f \mapsto \int_{\mathbb{R}^N} f(\mathbf{x})g(\mathbf{x})\,\mathrm{d}\mathbf{x}$ for some $g \in L^2(\mathbb{R}^N)$. Show that if $u \geq 0$ ($u(\mathbf{x}) \geq 0$ for all $\mathbf{x}$) implies $Tu \geq 0$, then $g \geq 0$. ∎

1.4 Comments and references

The heat equation. One should not conclude from Theorem 1.9 that the solutions of the heat equation are always unique. The assumption in (ii) that the solution was bounded is crucial. In fact, without this assumption, there are solutions u that grow so fast that gu is not in $L^1(\mathbb{R}^N)$ for $g \in \mathcal{S}$ (see, for example, [266, page 217]). The existence and uniqueness proof of Theorem 1.9 is classic and can be found in most textbooks on partial differential equations, such as Evans [98], Taylor [266], or Brezis [46].

Convolution. The heat equation—its solutions and their uniqueness—has been the main topic in this chapter, but to approach this, we have studied several aspects of the convolution, such as the continuity property (1.5). We also noted that the convolution commutes with translation. Conversely, as a general principle, any linear, translation invariant and continuous operator T is a convolution, that is, $Tu = g * u$ for some kernel g. This is a direct consequence of a result discovered independently by F. Riesz and M. Fréchet in 1907 (see [238, page 61] and exercise 1.7). Since we want smoothing to be translation invariant and continuous in some topology, this means that linear smoothing operators—which are called filters in the context of signal and image processing—are described by their convolution kernels. The Gaussian serves as a model for linear filters because it is the only one whose shape is stable under iteration. Other positive filters change their shape when iterated. This fact will be made precise in the next chapter where we show that a large class of iterated linear filters behaves asymptotically as a convolution with the Gaussian.

Smoothing and the Laplacian. One of the first tools proposed in the early days of image processing in the 1960s came, not surprisingly, directly from signal processing. The idea was to restore an image by averaging the gray levels locally (see, for example, [126] and [140]). The observation that the difference between an image and its local average is proportional to the Laplacian of the image has proved to be one of the most fruitful contributions to image processing. As noted in the Introduction, this method for deblurring an image was introduced by Kovasznay and Joseph in 1955 [175], and it was studied and optimized by Gabor in 1965 [117] (information taken from [183]). (See also [151] and [152].) Burt and Adelson based their Laplacian pyramid algorithm on this idea, and this was one of the results that led to multiresolution analysis and wavelets [49].

Chapter 2

Iterated Linear Filters and the Heat Equation

The title of this chapter is self-explanatory. The next section fixes fairly general conditions so that the difference of a smoothed image and the original be proportional to the Laplacian. The second section proves the main result, namely the convergence of iterated linear filters to the heat equation. So the choice of a smoothing convolution kernel is somewhat forced : Iterating the convolution with a smoothing kernel is asymptotically equivalent to the convolution with a Gauss function. This result is known in Probability as the central limit theorem, where it has a quite different interpretation. In image processing, it justifies the prominent role of Gaussian filtering. A last section is devoted to linear directional filters and their associated differential operators.

2.1 Smoothing and the Laplacian

There are minimal requirements on the smoothing kernels g which we state in the next definition.

Definition 2.1. *We say that a real-valued kernel $g \in L^1(\mathbb{R}^N)$ is Laplacian consistent if it satisfies the following moment conditions:*

(i) $\int_{\mathbb{R}^N} g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$.

(ii) *For $i = 1, 2, \dots, N$, $\int_{\mathbb{R}^N} x_i g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$.*

(iii) *For each pair $i, j = 1, 2, \dots, N$, $i \neq j$, $\int_{\mathbb{R}^N} x_i x_j g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$.*

(iv) *For $i = 1, 2, \dots, N$, $\int_{\mathbb{R}^N} x_i^2 g(\mathbf{x})\,\mathrm{d}\mathbf{x} = \sigma$, where $\sigma > 0$.*

(v) $\int_{\mathbb{R}^N} |\mathbf{x}|^3 |g(\mathbf{x})|\,\mathrm{d}\mathbf{x} < +\infty$.

Note that we do not assume that $g \geq 0$; in fact, many important filters used in signal and image processing are not positive. However, condition (i) implies that g is "on average" positive. A discussion of the necessity of the requirements $(i) - (v)$ is performed in Exercise 2.4.

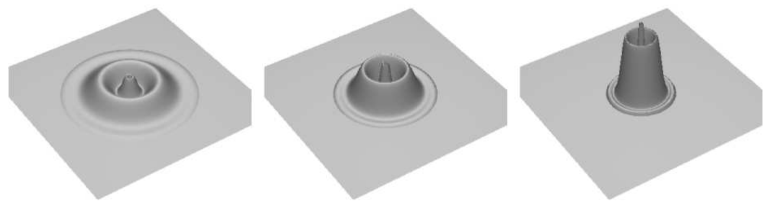

Figure 2.1: The rescalings $g_t(\mathbf{x}) = (1/t^2)g(\mathbf{x}/t)$ of a kernel for t=4, 3, and 2.

We say that a function g is *radial* if $g(\mathbf{x}) = g(|\mathbf{x}|)$, $\mathbf{x} \in \mathbb{R}^N$. This is equivalent to saying that g is invariant under all rotations around the origin in $\mathbb{R}^N$. As pointed out in Exercise 2.3, any radial function $g \in L^1(\mathbb{R}^N)$ can be rescaled to be Laplacian consistent if it decays fast enough at infinity and if $\int_{\mathbb{R}^N} \mathbf{x}_i^2 g(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and $\int_{\mathbb{R}^N} g(\mathbf{x})\,\mathrm{d}\mathbf{x}$ have the same sign.

We consider rescalings of a kernel g defined by

$$g_h(\mathbf{x}) = \frac{1}{h^{N/2}}\, g\Big(\frac{\mathbf{x}}{h^{1/2}}\Big) \tag{2.1}$$

for $h > 0$ (see Figure 2.1). Notice that this rescaling differs slightly from the one used in Section 1.2. We have used the factor $h^{1/2}$ here because it agrees with the factor $t^{1/2}$ in the Gaussian. We denote the convolution of g with itself n times by g^{n*}. The main result of this section concerns the behavior of g_h^{n*} as $n \to +\infty$ and $h \to 0$.

Exercise 2.1. Prove the following two statements:

(i) g_h is Laplacian consistent if and only it g is Laplacian consistent.

(ii) If $g \in L^1(\mathbb{R}^N)$, then $(g_h)^{n*} = (g^{n*})_h$. ∎

Our first result concerns the behavior of g_h as $h \to 0$. This will establish a more general and precise form of equation (1.2).

Theorem 2.2. *If g is Laplacian consistent, then for every $u \in \mathcal{F} \cap C^3$,*

$$g_h * u(\mathbf{x}) - u(\mathbf{x}) = h\frac{\sigma}{2}\Delta u(\mathbf{x}) + \varepsilon(h, \mathbf{x}) \tag{2.2}$$

where $|\varepsilon(h,\mathbf{x})| \le Ch^{3/2}$.

Proof. We use condition (i), the definition of g_h, and rescaling inside the integral to see that

$$\begin{aligned} g_h * u(\mathbf{x}) - u(\mathbf{x}) &= \int_{\mathbb{R}^N} \frac{1}{h^{N/2}}\, g\Big(\frac{\mathbf{x}}{h^{1/2}}\Big)\big(u(\mathbf{x}-\mathbf{y}) - u(\mathbf{x})\big)\,\mathrm{d}\mathbf{y} \\ &= \int_{\mathbb{R}^N} g(\mathbf{z})\big(u(\mathbf{x} - h^{1/2}\mathbf{z}) - u(\mathbf{x})\big)\,\mathrm{d}\mathbf{z}. \end{aligned}$$

Using Taylor's formula with the Lagrange remainder, we have

$$\begin{aligned} u(\mathbf{x} - h^{1/2}\mathbf{z}) - u(\mathbf{x}) \quad &= \quad -h^{1/2}Du(\mathbf{x})\cdot\mathbf{z} + \frac{h}{2}D^2u(\mathbf{x})(\mathbf{z},\mathbf{z}) \\ &\quad -\frac{1}{6}h^{3/2}D^3u(\mathbf{x} - h^{1/2}\theta\mathbf{z})(\mathbf{z},\mathbf{z},\mathbf{z}), \end{aligned}$$

where $\theta = \theta(\mathbf{x}, \mathbf{z}, h) \in [0,1]$. By condition (ii), $\int_{\mathbb{R}^N} g(\mathbf{z}) Du(\mathbf{x}) \cdot \mathbf{z} \, d\mathbf{z} = 0$; by conditions (iii) and (iv), $\int_{\mathbb{R}^N} g(\mathbf{z}) D^2u(\mathbf{x})(\mathbf{z}, \mathbf{z}) \, d\mathbf{z} = \sigma \Delta u(\mathbf{x})$. Thus,

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

We denote the error term by $\varepsilon(h, \mathbf{x})$. Then we have the following estimate:

$$\begin{aligned} |\varepsilon(h, \mathbf{x})| \leq & \frac{1}{6} h^{3/2} \int_{\mathbb{R}^N} |g(\mathbf{z}) D^3 u(\mathbf{x} - h^{1/2} \theta \mathbf{z})(\mathbf{z}, \mathbf{z}, \mathbf{z})| \, d\mathbf{z} \\ \leq & \frac{1}{6} h^{3/2} N^{3/2} \sup_{\alpha, \mathbf{x}} |\partial^\alpha u(\mathbf{x})| \int_{\mathbb{R}^N} |\mathbf{z}|^3 |g(\mathbf{z})| \, d\mathbf{z}, \end{aligned}$$

where the supremum is taken over all vectors $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_N)$, $\alpha_j \in \{1, 2, 3\}$, such that $|\alpha| = 3$ and over all $\mathbf{x} \in \mathbb{R}^N$. $\square$

The preceding theorem shows a direct relation between smoothing with a Laplacian-consistent kernel and the heat equation. It also shows why we require σ to be positive: If it is not positive, the kernel is associated with the inverse heat equation (see Exercise 2.4.)

2.2 The convergence theorem

The result of the next theorem is illustrated in Figure 2.2.

Theorem 2.3. *Let g be a nonnegative Laplacian-consistent kernel with $\sigma = 2$ and define g_h by* (2.1). *Write $T_h u_0 = g_h * u_0$ for $u_0 \in \mathcal{F}$, and let $u(t, \cdot) = G_t * u_0$ be the solution of the heat equation* (1.15). *Then, for each $t > 0$,*

$$(T_h^n u_0)(\mathbf{x}) \to u(t, \mathbf{x}) \text{ uniformly in } \mathbf{x} \text{ as } n \to +\infty \text{ and } nh \to t. \tag{2.3}$$

Proof. Let us start with some preliminaries. We have $(g_h * u_0)(\infty) = u_0(\infty)$ and therefore $T_h^n u_0(\infty) = u_0(\infty)$. The norm in $\mathcal{F}$ is $||u||_{\mathcal{F}} = \sup_{\mathbf{x} \in S_N} |u(\mathbf{x})| = \sup_{\mathbf{x} \in \mathbb{R}^N} |u(\mathbf{x})|$. The first order of business is to say precisely what is meant by the asymptotic limit (2.3): Given $t > 0$ and given $\varepsilon > 0$, there exists an $n_0 = n_0(t, \varepsilon)$ and a $\delta = \delta(t, \varepsilon)$ such that $\|T_h^n u_0 - u(t, \cdot)\|_{\mathcal{F}} \leq \varepsilon$ if $n > n_0$ and $|nh - t| \leq \delta$. This is what we must prove. We will first prove the result when $h = t/n$. We will then show that the result is true when h is suitably close to t/n.

We begin with comments about the notation. By Exercise 2.1, $(T_h)^n = (T^n)_h$, so there is no ambiguity in writing T_h^n. We will be applying T_h^n to the solution u of the heat equation, which is C^∞ on $(0, +\infty) \times \mathbb{R}^N$. In this situation, t is considered to be a parameter, and we write $T_h^n u(t, \mathbf{x})$ as shorthand for $T_h^n u(t, \cdot)(\mathbf{x})$. Throughout the proof, we will be dealing with error terms that we write as $O(h^r)$. These terms invariably depend on h, t, and $\mathbf{x}$. However, in all cases, given a closed interval $[t_1, t_2] \subset (0, +\infty)$, there will be a constant C such that $|O(h^r)| \leq Ch^r$ uniformly for $t \in [t_1, t_2]$ and $\mathbf{x} \in \mathbb{R}^N$. Finally, keep in mind that all functions of $\mathbf{x}$ tend to $u_0(\infty)$ as $\mathbf{x} \to \infty$.

We wish to fix an interval $[t_1, t_2]$, but since this depends on the point t in (2.3) and on ε, we must first choose these numbers. Thus, choose $\tau > 0$ and keep it fixed. This will be the "t" in (2.3). Next, choose $\varepsilon > 0$. Here are the conditions we wish t_1 and t_2 to satisfy:

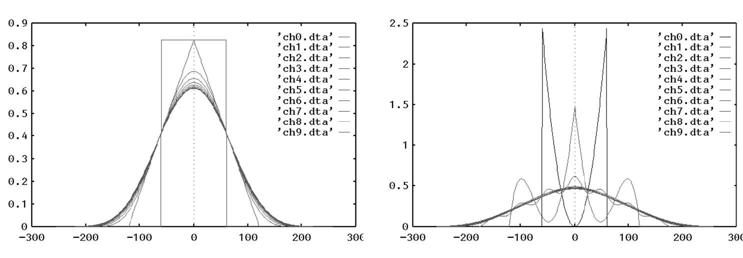

Figure 2.2: Iterated linear smoothing converges to the heat equation. In this experiment with one-dimensional functions, it can be appreciated how fast an iterated convolution of a positive kernel converges to a Gaussian. On the left are displayed nine iterations of the convolution of the characteristic function of an interval with itself, with appropriate rescalings. On the right, the same experiment is repeated with a much more irregular kernel. The convergence is almost as fast as the first case.

(1) t_1 is small enough so $\|u(t_1,\cdot) - u_0\|_{\mathcal{F}} < \varepsilon$. (This is possible by Theorem 1.9.)

(2) t_1 is small enough so $\|u(t_1+\tau,\cdot) - u(\tau,\cdot)\|_{\mathcal{F}} < \varepsilon$. (Again, by Theorem 1.9.)

(3) t_2 is large enough so $t_1 + \tau < t_2$.

There is no problem meeting these conditions, so we fix the interval $[t_1, t_2] \subset (0, +\infty)$.

Step 1, main argument : proof that

$$\lim_{\substack{n\to+\infty \\ nh=\tau}} T_h^n u(t_1,\mathbf{x}) = u(t_1+\tau,\mathbf{x}), \tag{2.4}$$

where the convergence is uniform for $\mathbf{x} \in \mathbb{R}^N$.
We can use Theorem 2.2 to write

$$T_h u(t,\mathbf{x}) - u(t,\mathbf{x}) = h\Delta u(t,\mathbf{x}) + O(h^{3/2}), \tag{2.5}$$

where $t \in [t_1, t_2]$. That the error function is bounded uniformly by $Ch^{3/2}$ on $[t_1,t_2]\times\mathbb{R}^N$ follows from the fact that $\sup_{\alpha,t,\mathbf{x}} |\partial^\alpha u(t,\mathbf{x})|$ is finite for $(t,\mathbf{x}) \in [t_1,t_2]\times\mathbb{R}^N$ (see the proof of Theorem 2.2). Since u is a solution of the heat equation, we also have

$$u(t+h,\mathbf{x}) - u(t,\mathbf{x}) = h\Delta u(t,\mathbf{x}) + O(h^2). \tag{2.6}$$

This time the error term is bounded uniformly by Ch^2 on $[t_1,t_2]\times\mathbb{R}^N$ because u is C^∞ on $(0,+\infty)\times\mathbb{R}^N$. By subtracting (2.6) from (2.5) we see that

$$T_h u(t,\mathbf{x}) = u(t+h,\mathbf{x}) + O(h^{3/2}). \tag{2.7}$$

This shows that applying T_h to a solution of the heat equation at time t advances the solution to time $t+h$, plus an error term.

So far we have not used the assumption that g is nonnegative. Thus, (2.7) is true for any Laplacian-consistent kernel g with $\sigma = 2$. However, we now wish to apply the linear operator T_h to both sides of equation (2.7), and in doing so we do not want the error term to increase. Since $g \geq 0$, this is not a problem:

$$|T_h O(h^{3/2})| \leq \int_{\mathbb{R}^N} |O(h^{3/2})| g_h(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y} \leq \int_{\mathbb{R}^N} Ch^{3/2} g_h(\mathbf{x}-\mathbf{y})\,\mathrm{d}\mathbf{y} = Ch^{3/2}.$$

With this in hand, we can apply T_h to both sides of (2.7) and obtain

$$T_h^2 u(t,\mathbf{x}) = T_h u(t+h,\mathbf{x}) + O(h^{3/2}). \tag{2.8}$$

If we write equation (2.7) with $t+h$ in place of t and substitute the expression for $T_h u(t+h,\mathbf{x})$ in equation (2.8), we have

$$T_h^2 u(t,\mathbf{x}) = u(t+2h,\mathbf{x}) + 2O(h^{3/2}). \tag{2.9}$$

We can iterate this process and get

$$T_h^n u(t,\mathbf{x}) = u(t+nh,\mathbf{x}) + nO(h^{3/2}) \tag{2.10}$$

with the same constant C in the estimate $|O(h^{3/2})| \leq Ch^{3/2}$ as long as $t+nh \in [t_1,t_2]$. To ensure that this happens, we take $t = t_1$ and $h = \tau/n$. Then

$$T_h^n u(t_1,\mathbf{x}) = u(t_1+\tau,\mathbf{x}) + O\Big(\Big(\frac{\tau}{n}\Big)^{1/2}\Big) \tag{2.11}$$

and we obtain (2.4). If we could take $t_1 = 0$, this would end the proof. This is not possible because all of the O terms were based on a fixed interval $[t_1,t_2]$. However, we have taken t_1 small enough to finish the proof .

Step 2 : getting rid of t_1.
Since $\int_{\mathbb{R}^N} g(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$, $\|g_h\|_{L^1(\mathbb{R}^N)} = 1$, and thus

$$\|g_h^{n*} * v\|_{\mathcal{F}} \leq \|v\|_{\mathcal{F}}.$$

If we take $v = u(t_1,\cdot) - u_0$, then this inequality and condition (1) imply that

$$\|T_h^n u(t_1,\cdot) - T_h^n u_0\|_{\mathcal{F}} < \varepsilon. \tag{2.12}$$

Relations (2.12) and (2.11) imply that

$$\|T_h^n u_0 - u(t_1+\tau,\cdot)\|_{\mathcal{F}} < 2\varepsilon. \tag{2.13}$$

This inequality and condition (2) show that

$$\|T_h^n u_0 - u(\tau,\cdot)\|_{\mathcal{F}} < 3\varepsilon \tag{2.14}$$

for $n > n_0$ and $h = \tau/n$. This proves the theorem in the case $h = \tau/n$.

Conclusion. It is a simple matter to obtain the more general result. Again, by Theorem 1.10, there is a $\delta = \delta(\tau,\varepsilon)$ such that $|nh-\tau| < \delta$ implies that $\|u(nh,\cdot) - u(\tau,\cdot)\|_{\mathcal{F}} < \varepsilon$ and that $nh \in [t_1,t_2]$ (by condition (3)). Combining this with (2.14) shows that

$$\|T_h^n u_0 - u(nh,\cdot)\|_{\mathcal{F}} < 4\varepsilon$$

if $n > n_0$ and $|nh-\tau| < \delta$, and this completes the proof. $\square$

2.3 Directional averages and directional heat equations

In this section, we list easy extensions of Theorem 2.2. They analyze local averaging processes which take averages at each point in a singular neighborhood made of a segment. In that way, we will make appear several nonlinear generalizations of the Laplacian which will accompany us throughout the book. Consider a C^2 function from $\mathbb{R}^N$ into $\mathbb{R}$ and a vector $\mathbf{z} \in \mathbb{R}^N$ with $|\mathbf{z}| = 1$. We wish to compute the mean value of u along a segment of the line through $\mathbf{x}$ parallel to the vector $\mathbf{z}$. To do this, we define the operator $T_h^{\mathbf{z}}$, $h \in [-1, 1]$, by

$$T_h^{\mathbf{z}} u(\mathbf{x}) = \frac{1}{2h} \int_{-h}^{h} u(\mathbf{x} + s\mathbf{z})\, \mathrm{d}s.$$

This operator is the directional counterpart of the isotropic operator M_h defined by equation (1.1). We use Taylor's formula to expand u at the point $\mathbf{x}$ along the line through $\mathbf{x}$ parallel to the vector $\mathbf{z}$:

$$u(\mathbf{x} + s\mathbf{z}) = u(\mathbf{x}) + sDu(\mathbf{x}) \cdot \mathbf{z} + \frac{s^2}{2} D^2 u(\mathbf{x})(\mathbf{z}, \mathbf{z}) + o(s^2). \tag{2.15}$$

By averaging both sides of (2.15) over $s \in [-h, h]$, we obtain the next result.

Proposition 2.4.

$$T_h^{\mathbf{z}} u(\mathbf{x}) = u(\mathbf{x}) + \frac{h^2}{6} D^2 u(\mathbf{x})(\mathbf{z}, \mathbf{z}) + o(h^2).$$

Proposition 2.4 is similar to to Theorem 2.2, and it suggests that iterations of the operator $T_h^{\mathbf{z}}$ are associated with the directional heat equation

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \frac{1}{6} D^2 u(t, \mathbf{x})(\mathbf{z}, \mathbf{z}) \tag{2.16}$$

in the same way that the iterations of the operator T_h in Theorem 2.3 are associated with the ordinary heat equation. If $\mathbf{z}$ is fixed, then the operator $T_h^{\mathbf{z}}$ and equation (2.16) act on u along each line in $\mathbb{R}^N$ parallel to $\mathbf{z}$ separately; there is no "cross talk" between lines. Exercise 2.5 formalizes and clarifies these comments when $\mathbf{z}$ is fixed. However, Proposition 2.4 is true when $\mathbf{z}$ is a function of $\mathbf{x}$. This means that we are able to approximate the directional second derivative by taking directional averages where $\mathbf{z}$ varies from point to point. The main choices considered in the book are $\mathbf{z} = Du/|Du|$ and $\mathbf{z} = Du^\perp/|Du|$, where $Du = (u_x, u_y)$ and $Du^\perp = (-u_y, u_x)$. Then by Proposition 2.4 we have the following limiting relations:

- Average in the direction of the gradient. By choosing $z = Du/|Du|$,

  $$\frac{1}{|Du|^2} D^2 u(Du, Du) = 6 \lim_{h \to 0} \frac{T_h^{Du/|Du|} u - u}{h^2}.$$

  We will interpret this differential operator as Haralick's edge detector in section 3.1.

- Average in the direction orthogonal to the gradient. By choosing $z = Du/|Du^\perp|$,

$$\frac{1}{|Du|^2} D^2u(Du^\perp, Du^\perp) = 6 \lim_{h\to 0} \frac{T_h^{Du^\perp/|Du|}u - u}{h^2}.$$

  This differential operator appears as the second term of the curvature equation. (See Chapter 12.)

Although we have not written them as such, the limits are pointwise in both cases.

2.4 Exercises

Exercise 2.2. We will denote the characteristic function of a set $A \subset \mathbb{R}^N$ by $\mathbf{1}_A$. Thus, $\mathbf{1}_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $\mathbf{1}_A(\mathbf{x}) = 0$ otherwise. Consider the kernel $g = (1/\pi)\mathbf{1}_{D(0,1)}$, where $D(0,1)$ is the disk of radius one centered at zero. In this case, g is a radial function and it is clearly Laplacian consistent. For $N = 2$, let $A = [-1/2, 1/2] \times [-1/2, 1/2]$. Then $g = \mathbf{1}_A$ is not radial. Show that it is, however, Laplacian consistent. If we take $B = [-1,1]\times[-1/2,1/2]$, then $g = (1/2)\mathbf{1}_B$ is no longer Laplacian consistent because it does not satisfy condition (iv). Show that this kernel does, however, satisfy a relation similar to (2.2). ∎

Exercise 2.3. The aim of the exercise is to prove roughly that radial functions with fast decay are Laplacian consistent. Assume $g \in L^1(\mathbb{R}^N)$ is radial with finite first second moments, $\int_{\mathbb{R}^N} |\mathbf{x}|^k |g(\mathbf{x})|\,\mathrm{d}\mathbf{x} < +\infty$, $k = 0,1,2,3$ and such that $\int_{\mathbb{R}^N} \mathbf{x}_i^2 g(\mathbf{x})\,\mathrm{d}\mathbf{x} > 0$. Show that g satisfies conditions (ii) and (iii) of Definition 2.1 and that, for suitably chosen $a, b \in \mathbb{R}$, the rescaled function $\mathbf{x} \mapsto ag(\mathbf{x}/b)$ satisfies conditions (i) and (iv), where σ can be taken to be an arbitrary positive number. ∎

Exercise 2.4. The aim of the exercise is to illustrate by simple examples what happens to the iterated filter g^{n*}, $n \in \mathbb{N}$ when g does not satisfy some of the requirements of the Laplacian consistency (Definition 2.1). We recall the notation (2.1), $g_h(\mathbf{x}) = \frac{1}{h^{N/2}}\, g\Big(\frac{\mathbf{x}}{h^{1/2}}\Big)$.

1) Take on $\mathbb{R}$, $g(x) = 1$ on $[-1,1]$, $g(x) = 0$ otherwise. Which one of the assumptions $(i) - (v)$ is not satisfied in Definition 2.1 ? Compute $g_{\frac{1}{n}}^{n*} * u$, where $u = 1$ on $\mathbb{R}$. Conclude : the iterated filter blows up.

2) Take on $\mathbb{R}$, $g(x) = 1$ on $[0,1]$, $g(x) = 0$ otherwise. Which one of the assumptions $(i) - (v)$ is not satisfied in Definition 2.1 Compute $g^{n*} * u$, where $u(x) = x$ on $\mathbb{R}$. Conclude : the iterated filter “drifts”.

3) Assume that the assumptions $(i) - (v)$ hold, except (iii). By a simple adaptation of its proof, draw a more general form of Theorem 2.2.

4) Perform the same analysis as in 3) when all assumptions hold but (iv).

5) Take the case of dimension $N = 1$ and assume that (i) hold but (ii) does not hold. Set $g_h(\mathbf{x}) = \frac{1}{h}\, g\Big(\frac{\mathbf{x}}{h}\Big)$ and give a version of Theorem 2.2 in that case (make an order 1 Taylor expansion of u). ∎

Exercise 2.5. Let $\mathbf{z}$ be a fixed vector in $\mathbb{R}^N$ with $|\mathbf{z}| = 1$ and let u_0 be in $\mathcal{F}$. Define a one-dimensional kernel g by $g(s) = \frac{1}{2}\mathbf{1}_{[-1,1]}(s)$.

(i) Show that g is Laplacian consistent. Compute the variance σ of g.

(ii) Show that
$$u(t,\mathbf{x}) = \int_{\mathbb{R}} u_0(\mathbf{x}+s\mathbf{z})G_t(s)\,\mathrm{d}s$$
is a solution of the directional heat equation
$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = D^2u(t,\mathbf{x})(\mathbf{z},\mathbf{z}), \quad u(0,\mathbf{x}) = u_0(\mathbf{x}). \tag{2.17}$$
Give an example to show that $u(t,\cdot)$ is not necessarily C^2. This being the case, how does one interpret the right-hand side of (2.17)?

(iii) Let $g_h(s) = (6h)^{-1/2}g(s/(6h)^{-1/2})$ and $T_hu(\mathbf{x}) = \int_{\mathbb{R}} u(\mathbf{x}+s\mathbf{z})g_h(s)\,\mathrm{d}s$. By applying Theorem 2.3 for $N=1$, show that, for each $t>0$,
$$T_h^n u_0 \to u(t,\cdot) \text{ in } \mathcal{F} \text{ as } n\to+\infty \text{ and } nh\to t. \ \blacksquare \tag{2.18}$$

Exercise 2.6. The Weickert equation can be viewed as a variant of the curvature equation [280]. It uses a nonlocal estimate of the direction orthogonal to the gradient for the diffusion direction. This direction is computed as the direction v of the eigenvector corresponding to the smallest eigenvalue of $k*(Du\otimes Du)$, where $(\boldsymbol{y}\otimes\boldsymbol{y})(\boldsymbol{x}) = (\boldsymbol{x}\cdot\boldsymbol{y})\boldsymbol{y}$. Prove that if the convolution kernel is removed, then this eigenvector is simply $Du^{\perp}$. So the equation writes
$$\frac{\partial u}{\partial t} = u_{\eta\eta}, \tag{2.19}$$
where η denotes the coordinate in the direction v. $\blacksquare$

Exercise 2.7. Suppose that $u\in C^2(\mathbb{R})$. Assuming that $u'(x)\neq 0$, show that
$$u''(x) = \lim_{h\to 0}\frac{1}{h^2}\Big(\max_{s\in[-h,h]} u(x+s) + \min_{s\in[-h,h]} u(x+s) - 2u(x)\Big). \tag{2.20}$$
What is the value of the right-hand side of (2.20) if $u'(x)=0$?

Now consider $u\in C^2(\mathbb{R}^2)$. We wish to establish an algorithm similar to (2.20) to compute the second derivative of u in the direction of the gradient $Du=(u_x,u_y)$. For this to make sense, we must assume that $Du(\mathbf{x})\neq 0$. With these assumptions, we know from (2.20) that
$$u_{\xi\xi}(\mathbf{x}) = \frac{\partial^2 v}{\partial\xi^2}(\mathbf{x},0) = \lim_{h\to 0}\frac{1}{h^2}\Big(\max_{s\in[-h,h]} u(\mathbf{x}+s\mathbf{z}) + \min_{s\in[-h,h]} u(\mathbf{x}+s\mathbf{z}) - 2u(\mathbf{x})\Big), \tag{2.21}$$
where $v(\mathbf{x},\xi) = u(\mathbf{x}+\xi\mathbf{z})$ and $\mathbf{z} = Du/|Du|$. The second part of the exercise is to prove that, in fact,
$$u_{\xi\xi}(\mathbf{x}) = \lim_{h\to 0}\frac{1}{h^2}\Big(\max_{\mathbf{y}\in D(0,h)} u(\mathbf{x}+\mathbf{y}) + \min_{\mathbf{y}\in D(0,h)} u(\mathbf{x}+\mathbf{y}) - 2u(\mathbf{x})\Big), \tag{2.22}$$
where $D(0,h)$ is the disk of radius h centered at the origin. Intuitively, (2.22) follows from (2.21) because the gradient indicates the direction of maximal change in $u(\mathbf{x})$, so in the limit as $h\to 0$, taking max and min in the direction of the gradient is equivalent to taking max and min in the disk. The point of the exercise is to formalize this. $\blacksquare$

2.5 Comments and references

Asymptotics. Our proof that iterated and rescaled convolutions of a Laplacian-consistent kernel tend asymptotically to the Gaussian is a version of the De Moivre–Laplace formula, or the central limit theorem, adapted to image processing [45]. This result is particularly relevant to image analysis, since it implies that iterated linear smoothing leads inevitably to convolution with the Gaussian, or equivalently, to the application of the heat equation. We do not wish to imply, however, that the Gaussian is the only important kernel for image processing. The Gaussian plays a significant role in our form of image analysis, but there are other kernels that, because of their spectral and algebraic properties, have equally important roles in other aspects of signal and image processing. This is particularly true for wavelet theory which combines recursive filtering and sub-sampling.

Directional diffusion. Directional diffusion has a long history that began when Hubel and Wiesel showed the existence of direction-sensitive cells in the visual areas of the neocortex [146]. There has been an explosion of publication on directional linear filters, beginning, for example, with influential papers such as that by Daugman [85]. We note again that Gabor's contribution to directional filtering is described in [183].

Chapter 3

Linear Scale Space and Edge Detection

The general analysis framework in which an image is associated with smoothed versions of itself at several scales is called *scale space*. Following the results of Chapter 2, a linear scale space must be performed by applying the heat equation to the image. The main aim of this smoothing is to find out *edges* in the image. We shall first explain this doctrine. In the second section, we discuss experiments and several serious objections to such an image representation.

3.1 The edge detection doctrine

One of the uses of linear theory in two dimensions is *edge detection*. The assumption of the edge detection doctrine is that relevant information is contained in the traces produced in an image by the apparent contours of physical objects. If a black object is photographed against a white background, then one expects the silhouette of the object in the image to be bounded by a closed curve across which the light intensity u_0 varies strongly. We call this curve an *edge*. At first glance, it would seem that this edge could be detected by computing the gradient Du_0, since at a point $\mathbf{x}$ on the edge, $|Du_0(\mathbf{x})|$ should be large and $Du(\mathbf{x})$ should point in a direction normal to the boundary of the silhouette. It would therefore appear that finding edges amounts to computing the gradient of u_0 and determining the points where the gradient is large. This conclusion is unrealistic for two reasons:

(a) There may be many points where the gradient is large due to small oscillations in the image that are not related to real objects. Recall that digital images are always noisy, and thus there is no reason to assume the existence or computability of a gradient.

(b) The points where the gradient exceeds a given threshold are likely to form regions and not curves.

As we emphasized in the Introduction, objection (a) is dealt with by smoothing the image. We associate with the image u_0 smoothed versions $u(t, \cdot)$, where the scale parameter t indicates the amount of smoothing. In the classical linear theory, this smoothing is done by convolving u_0 with the Gaussian G_t.

One way that objection (b) has been approached is by redefining edge points. Instead of just saying an edge point is a point $\mathbf{x}$ where $|Du_0(\mathbf{x})|$ exceeds a threshold, one requires the gradient to satisfy a maximal property. We illustrate this in one dimension. Suppose that $u \in C^2(\mathbb{R})$ and consider the points where $|u'(x)|$ attains a local maximum. At some of these points, the second derivative u'' changes sign, that is, $\operatorname{sign}(u''(x-h)) \neq \operatorname{sign}(u''(x+h))$ for sufficiently small h. These are the points where u'' *crosses zero*, and they are taken to be the edge points. Note that this criterion avoids classifying a point x as an edge point if the gradient is constant in an interval around x. Marr and Hildreth generalized this idea to two dimensions by replacing u'' with the Laplacian Δu, which is the only isotropic linear differential operator of order two that generalizes u'' [199]. Haralick's edge detector is different but in the same spirit [135]. Haralick gives up linearity and defines edge points as those points where the gradient has a local maximum in the direction of the gradient. In other words, an edge point $\mathbf{x}$ satisfies $g'(0) = 0$, where $g(t) = |Du(\mathbf{x}+tDu(\mathbf{x})|/|Du(\mathbf{x})|$. This implies that $D^2u(\mathbf{x})(Du(\mathbf{x}), Du(\mathbf{x})) = 0$ (see Exercise 3.2). We are now going to state these two algorithms formally. They are illustrated in Figures 3.2 and 3.3, respectively.

Algorithm 3.1 (Edge detection: Marr–Hildreth zero-crossings).

(1) Create the multiscale images $u(t, \cdot) = G_t * u_0$ for increasing values of t.

(2) At each scale t, compute all the points where $Du \neq 0$ and Δu changes sign. These points are called zero-crossings of the Laplacian, or simply zero-crossings.

(3) (Optional) Eliminate the zero-crossings where the gradient is below some prefixed threshold.

(4) track back from large scales to fine scales the "main edges" detected at large scales.

Algorithm 3.2 (Edge detection: The Haralick–Canny edge detector).

(1) As before, create the multiscale images $u(t, \cdot) = G_t * u_0$ for increasing values of t.

(2) At each scale t, find all points $\mathbf{x}$ where $Du(\mathbf{x}) \neq 0$ and $D^2u(\mathbf{x})(\mathbf{z}, \mathbf{z})$ crosses zero, $\mathbf{z} = Du/|Du|$. At such points, the function $s \mapsto u(\mathbf{x} + s\mathbf{z})$ changes from concave to convex, or conversely, as s passes through zero.

(3) At each scale t, fix a threshold $\theta(t)$ and retain as *edge points at scale t* only those points found above that satisfy $|Du(\mathbf{x})| > \theta(t)$. The backtracking step across scales is the same as for Marr–Hildreth.

In practice, edges are computed for a finite number of dyadic scales, $t = 2^n$, $n \in \mathbb{Z}$.

3.1.1 Discussion and critique

The Haralick–Canny edge detector is generally preferred for its accuracy to the Marr–Hildreth algorithm. Their use and characteristics are, however, essentially

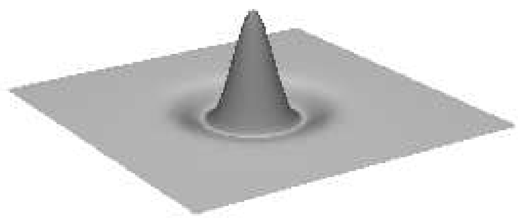

Figure 3.1: A three-dimensional representation of the Laplacian of the Gaussian. This convolution kernel, which is a wavelet, is used to estimate the Laplacian of an image at different scales of linear smoothing.

the same. There are also many variations—attempted improvements—of the algorithms we have described, and the following discussion adapts easily to these related edge detection schemes. The first thing to notice is that, by Proposition 1.5, $u(t,\cdot) = G_t * u_0$ is a C^∞ function for each $t > 0$ if $u_0 \in \mathcal{F}$. Thus we can indeed compute second order differential operators applied to $u(t,\cdot) = G_t * u_0$, $t > 0$. In the case of linear operators like the Laplacian or the gradient, the task is facilitated by the formula proved in the mentioned proposition. For example, we have $\Delta u(t,\mathbf{x}) = \Delta(G_t * u_0)(\mathbf{x}) = (\Delta G_t) * u_0(\mathbf{x})$, where in dimension two (Figure 3.1),

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

In the same way, Haralick's edge detector makes sense, because u is C^∞, at all points where $Du(\mathbf{x}) \neq 0$. If $Du(\mathbf{x}) = 0$, then $\mathbf{x}$ cannot be an edge point, since u is "flat" there. Thus, thanks to the filtering, there is no theoretical problem with computing edge points. There are, however, practical objections to these methods, which we will now discuss.

Linear scale space

The first serious problems are associated with the addition of an extra dimension: Having many images $u(t,\cdot)$ at different scales t confounds our understanding of the image and adds to the cost of computation. We no longer have an absolute definition of an edge. We can only speak of edges at a certain scale. Conceivably, a way around this problem would be to track edges across scales. In fact, it has been observed in experiments that the "main edges" persist under convolution as t increases, but they lose much of their spatial accuracy. On the other hand, filtering with a sharp low-pass filter, that is, with t small, keeps these edges in their proper positions, but eventually, as t becomes very small, even these main edges can be lost in the crowd of spurious edge signals due to noise and texture. The scale space theory of Witkin proposes to identify the main edges at some scale t and then to track them backward as t decreases [289]. In theory, it would seem that this method could give an accurate location of the main edges. In practice, any implementation of these ideas is computationally costly due to the problems involved with multiple thresholdings and following edges across scales. In fact, tracking edges across scales is incompatible with having thresholds for the gradients, since such thresholds may remove edges at

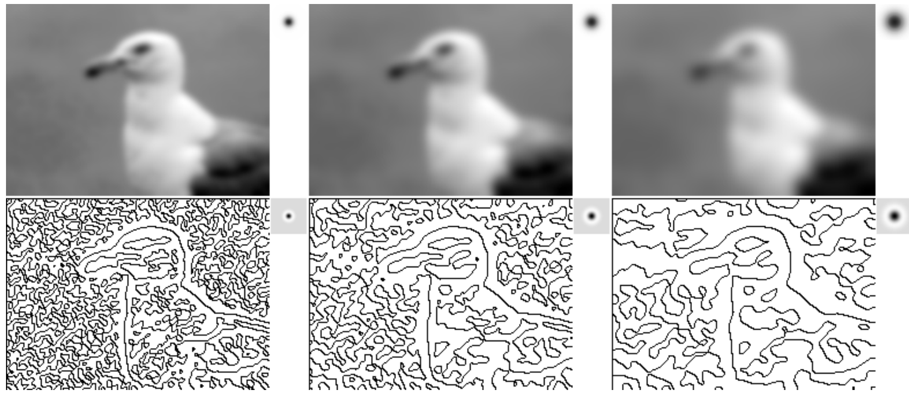

Figure 3.2: Zero-crossings of the Laplacian at different scales. This figure illustrates the original scale space theory as developed by David Marr [198]. To extract more global structure, the image is convolved with Gaussians whose variances are powers of two. One computes the Laplacian of the smoothed image and displays the lines along which this Laplacian changes sign: the zero-crossings of the Laplacian. According to Marr, these zero-crossings represent the "raw primal sketch" of the image, or the essential information on which further vision algorithms should be based. Above, left to right: the results of smoothing and the associated Gaussian kernels at scales 1, 2, and 4. Below, left to right: the zero-crossings of the Laplacian and the corresponding kernels, which are the Laplacians of the Gaussians used above.

certain scales and not at others. The conclusion is that one should trace all zero-crossings across scales without considering whether they are true edges or not. This makes matching edges across scales very difficult. For example, experiments show that zero-crossings of sharp edges that are sparse at small scales are no longer sparse at large scales. (Figure 3.4 shows how zero-crossings can be created by linear smoothing.) The Haralick–Canny detector suffers from the same problems, as is well demonstrated by experiments.

Other problems with linear scale space are illustrated in Figures 3.5 and 3.6. Figure 3.5 illustrates how linear smoothing can create new gray levels and new extrema. Figure 3.6 shows that linear scale space does not maintain the inclusion between objects. The shape inclusion principal will be discussed in Chapter 21.

We must conclude that the work on linear edge detection has been an attempt to build a theory that has not succeeded. After more than thirty years of activity, it has become clear that no robust technology can be based on these ideas. Since edge detection algorithms depend on multiple thresholds on the gradient, followed by "filling-the-holes" algorithms, there can be no scientific agreement on the identification of edge points in a given image. In short, the problems associated with linear smoothing followed by edge detection have not been resolved by the idea of chasing edges across scales.

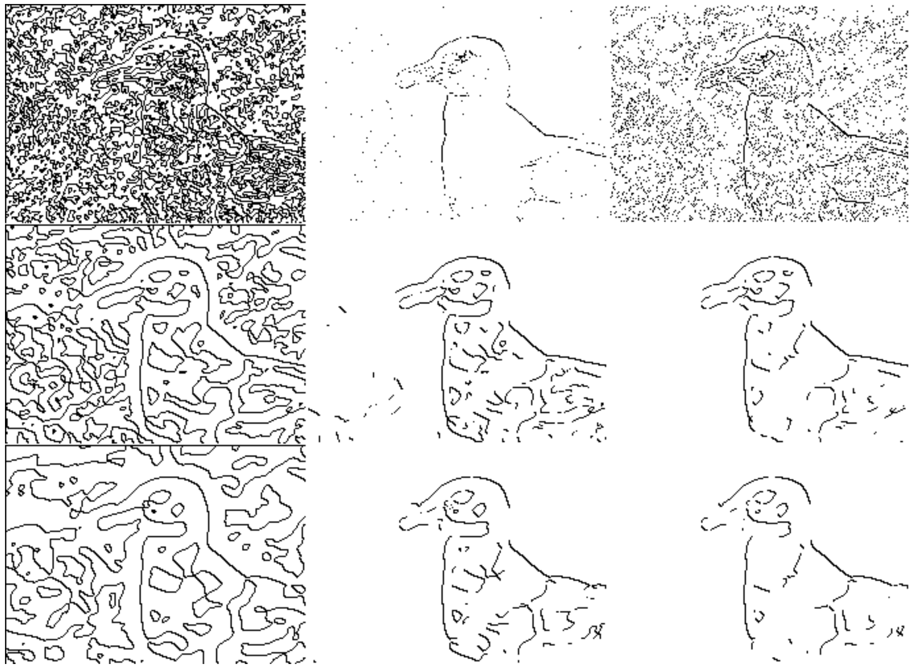

Figure 3.3: Canny's edge detector. These images illustrate the Canny edge detector. Left column: result of the Canny filter without the threshold on the gradient. Middle column: result with a visually "optimal" scale and an image-dependent threshold (from top to bottom: 15, 0.5, 0.6). Right column: result with a fixed gradient threshold equal to 0.7. Note that such an edge detection theory depends on no fewer than two parameters that must be fixed by the user: smoothing scale and gradient threshold .

Figure 3.4: Zero-crossings of the Laplacian of a synthetic image. Left to right: the original image; the image linearly smoothed by convolution with a Gaussian; the sign of the Laplacian of the filtered image (the gray color corresponds to values close to 0, black to clear-cut negative values, white to clear-cut positive values); the zero-crossings of the Laplacian. This experiment clearly shows a drawback of the Laplacian as edge detector.

Contrast invariance

As already mentioned in the Introduction, a central theme of the book is that the use of contrast-invariant operators will solve some of the technical problems associated with linear smoothing and other linear image operators. The

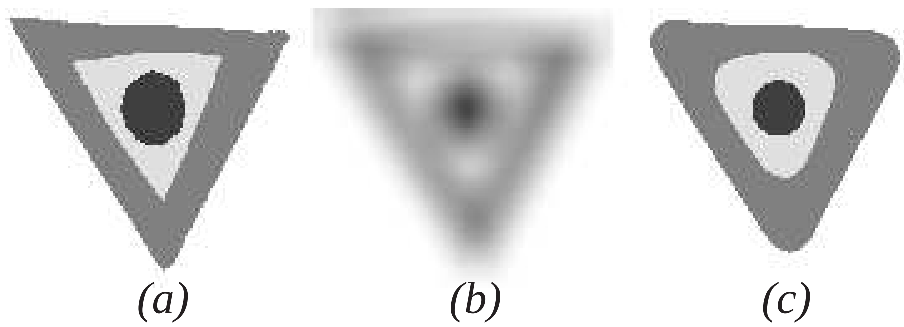

Figure 3.5: The heat equation creates structure. This experiment shows that linear scale space can create new structures and thus increase the complexity of an image. Left to right: The original synthetic image (a) contains three gray levels. The black disk is a regional and absolute minimum. The "white" ring around the black disk is a regional and absolute maximum. The outer gray ring has a gray value between the other two and is a regional minimum. The second image (b) shows what happens when (a) is smoothed with the heat equation: New local extrema have appeared. Image (c) illustrates the action on (a) of a contrast-invariant local filter, the iterated median filter, which is introduced in Chapter 10.

development of these ideas starts in Chapter 3.

Recall from section I.3 that an (image) operator $u \mapsto Tu$ is contrast invariant if T commutes with all nondecreasing functions g, that is, if

$$g(Tu) = T(g(u)). \tag{3.1}$$

If image analysis is to be robust, it must be invariant under changes in lighting that produce contrast changes. It must also be invariant under the nonlinear response of the sensors used to capture an image. These, and perhaps other, contrast changes are modeled by g. If g is strictly increasing, then relation (3.1) ensures that the filtered image $Tu = g^{-1}(T(g(u)))$ does not depend on g. A problem with linear theory is that linear smoothing, that is, convolution, is not generally contrast invariant:

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

In the same way, the operator T_t that maps u_0 into the solution of the heat equation, $u(t, \cdot)$ is not generally contrast invariant. In fact, if g is C^2, then

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

and

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

Exercise 3.1. Prove this last relation. Prove that if $g(s) = as + b$ then $g(u)$ satisfies the heat equation if u does. ∎

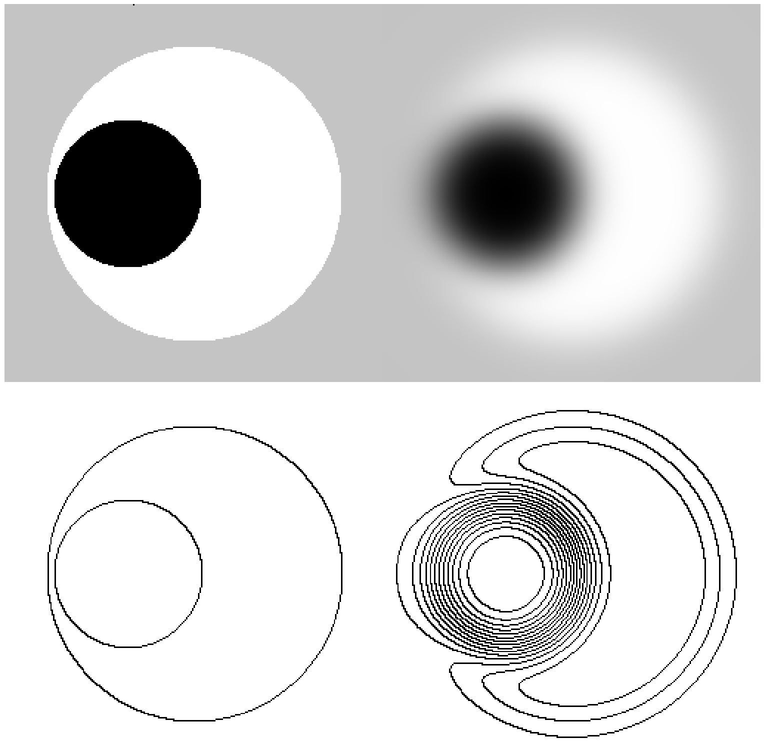

Figure 3.6: Violation of the inclusion by the linear scale space. Top, left: an image that contains a black disk enclosed by a white disk. Top, right: At a certain scale, the black and white circles mix together. Bottom, left: The boundaries of the two circles. Bottom, right: After smoothing with a certain value of t, the inclusion that existed for very small t in no longer preserved. We display the level lines of the image at levels multiples of 16.

3.2 Exercises

Exercise 3.2. Define an edge point $\mathbf{x}$ in a smooth image u as a point $\mathbf{x}$ at which $g(t)$ attains a maximum, where

$$g(t) = |Du\left(\mathbf{x} + t\frac{Du(\mathbf{x})}{|Du(\mathbf{x})|}\right)|.$$

Prove by differentiating $g(t)$ that edge points satisfy $D^2u(\mathbf{x})(Du(\mathbf{x}), Du(\mathbf{x})) = 0$ ■

Exercise 3.3. Construct simple functions u, g, and k such that $g(k * u) \neq k * (g(u))$. ■

Exercise 3.4. Consider the Perona–Malik equation in divergence form:

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(|Du|)Du), \tag{3.2}$$

where $g(s) = 1/(1+\lambda^2 s^2)$. It is easily checked that we have a diffusion equation when $\lambda|Du| \leq 1$ and an inverse diffusion equation when $\lambda|Du| > 1$. To see this, consider the second derivative of u in the direction of Du,

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

and the second derivative of u in the orthogonal direction,

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

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 $\Delta u = u_{\xi\xi} + u_{\eta\eta}$. Prove that the Perona–Malik equation then becomes

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

Interpret the local behavior of the equation as a heat equation or a reverse heat equation according to the size of $|Du|$ compared to λ^{-1}. ■

3.3 Comments and references

Scale space. The term “scale space” was introduced by Witkin in 1983. He suggested tracking the zero-crossings of the Laplacian of the smoothed image across scales [289]. Yuille and Poggio proved that these zero-crossings can be tracked for one-dimensional signals [292]. Hummel and Moniot [149, 153] and Yuille and Poggio [293] analyzed the conjectures of Marr and Witkin according to which an image is completely recoverable from its zero-crossings at different scales. Mallat formulated Marr’s conjecture as an algorithm in the context of wavelet analysis. He replaced the Gaussian with a two-dimensional cubic spline, and he used both the zero-crossings of the smoothed images and the nonzero values of the gradients at these points to reconstruct the image. This algorithm works well in practice, and the conjecture was that these zero-crossings and the values of the gradients determined the image. A counterexample given by Meyer shows that this is not the case. Perfect reconstruction is possible in the one-dimensional case for signals with compact support if the smoothing kernel is the Tukey window, $k(x) = 1 + \cos x$ for $|x| \le \pi$ and zero elsewhere. An account of the Mallat conjecture and these examples can be found in [159]. Koenderink presents a general and insightful theory of image scale space in [171].

Gaussian smoothing and edge detection. The use of Gaussian filtering in image analysis is so pervasive that it is impossible to point to a “first paper.” It is, however, safe to say that David Marr’s famous book, *Vision* [198], and the original paper by Hildreth and Marr [199] have had an immeasurable impact on edge detection and image processing in general. The term “edge detection” appeared as early as 1959 in connection with television transmission [161]. The idea that the computation of derivatives of an image necessitates a previous smoothing has been extensively developed by the Dutch school of image analysis [42, 115]. See also the books by Florack [110], Lindeberg [182], and Romeny [268], and the paper [103]. Haralick’s edge detector [135], as implemented by Canny [51], is probably the best known image analysis operator. A year after Canny’s 1986 paper, Deriche published a recursive implementation of Canny’s criteria for edge detection [89].

Chapter 4

Four Algorithms to Smooth a Shape

In this short but important chapter, we discuss algorithms whose aim it is to smooth shapes. Shape must be understood as a rough data which can be extracted from an image, either a subset of the plane, or the curve surrounding it. Shape smoothing is directed at the elimination of spurious, often noisy, details. The smoothed shape can then be reduced to a compact and robust code for recognition. The choice of the right smoothing will make us busy throughout the book. A good part of the solution stems from the four algorithms we describe and their progress towards more robustness, more invariance and more locality. What we mean by such qualities will be progressively formalized. We will discuss two algorithms which directly smooth *sets*, and two which smooth Jordan curves. One of the aims of the book is actually to prove that both approaches, different though they are, eventually yield the *very same process*, namely a curvature motion.

4.1 Dynamic shape

In 1986, Koenderink and van Doorn defined a *shape* in $\mathbb{R}^N$ to be a closed subset X of $\mathbb{R}^N$ [174]. They then proposed to smooth the shape by applying the heat equation $\partial u/\partial t - \Delta u = 0$ directly to $\mathbf{1}_X$, the characteristic function of X. Of course, the solution $G_t * \mathbf{1}_X$ is not a characteristic function. The authors defined the evolved shape at scale t to be

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

The value $1/2$ is chosen so the following simple requirement is satisfied: Suppose that X is the half-plane $X = \{(x, y) \mid (x, y) \in \mathbb{R}^2, x \geq 0\}$. The requirement is that this half plane doesn't move,

$$X = X_t = \{(x, y) \mid G_t * \mathbf{1}_X(x, y) \geq \lambda\},$$

and this is true only if $\lambda = 1/2$. There are at least two problems with dynamic shape evolution for image analysis. The first concerns nonlocal interactions, as illustrated in Figure 4.1. Here we have two disks that are near one another.

Figure 4.1: Nonlocal interactions in the dynamic shape method. Left to right: Two close disks interact as the scale increases. This creates a new, qualitatively different, shape. The change of topology, at the scale where the two disks merge into one shape, also entails the appearance of a singularity (a cusp) on the shape(s) boundaries.

The evolution of the union of both disks, considered as a single shape, is quite different from the evolution of the disks separately. A related problem, also illustrated in Figure 4.1, is the creation of singularities. Note how a singularity in orientation and the curvature of the boundary of the shape develops at the point where the two disks touch. Figure 4.2 further illustrates the problems associated with the dynamic shape method.

4.2 Curve evolution using the heat equation

We consider shapes in $\mathbb{R}^2$ whose boundaries can be represented by a finite number of simple closed rectifiable Jordan curves. Thus, each curve we consider can be represented by a continuous mapping $f : [0,1] \to \mathbb{R}^2$ such that f is one-to-one on $(0,1)$ and $f(0) = f(1)$, and each curve has a finite length. We also assume that these curves do not intersect each other. We will focus on smoothing one of these Jordan curves, which we call C_0. We assume that C_0 is parameterized by $s \in [0,L]$, where L is the length of the curve. Thus, C_0 is represented as $\mathbf{x}_0(s) = (x(s), y(s))$, where s is the length of the curve between $\mathbf{x}_0(0)$ and $\mathbf{x}_0(s)$.

At first glance, it might seem reasonable to smooth C_0 by smoothing the coordinate functions x and y separately. If this is done linearly, we have seen from Theorem 2.3 that the process is asymptotic to smoothing with the heat equation. Thus, one is led naturally to consider the vector heat equation

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s) \tag{4.1}$$

with initial condition $\mathbf{x}(0,s) = \mathbf{x}_0(s)$. If $\mathbf{x}(t,s) = (x(t,s), y(t,s))$ is the solution of (4.1), then we know from Proposition 1.9 that

$$\inf_{s\in[0,L]} x_0(s) \le x(t,s) \le \sup_{s\in[0,L]} x_0(s),$$
$$\inf_{s\in[0,L]} y_0(s) \le y(t,s) \le \sup_{s\in[0,L]} y_0(s),$$

for $s \in [0,L]$ and $t \in [0,+\infty)$. Thus, the evolved curves C_t remain in the rectangle that held C_0. Also, we know from Proposition 1.5 that the coordinate functions $x(t,\cdot)$ and $y(t,\cdot)$ are C^∞ for $t > 0$. There are, however, at least two reasons that argue against smoothing curves this way:

Figure 4.2: Nonlocal behavior of shapes with the dynamic shape method. This image displays the smoothing of two irregular shapes by the dynamic shape method (Koenderink–van Doorn). Top left: initial image, made of two irregular shapes. From left to right, top to bottom: dynamic shape smoothing with increasing Gaussian variance. Notice how the shapes merge more and more. We do not have a separate analysis of each shape but rather a "joint analysis" of the two shapes. The way the shapes merge is of course sensitive to the initial distance between the shapes. Compare with Figure 4.4.

(1) When $t > 0$, s is no longer a length parameter for the evolved curve C_t.

(2) Although $x(t, \cdot)$ and $y(t, \cdot)$ are C^∞ for $t > 0$, this does not imply that the curves C_t have similar smoothness properties. In fact, it can be seen from Figure 4.3 that it is possible for an evolved curve to cross itself and it is possible for it to develop singularities.

How is this last mentioned phenomenon possible ? It turns out that one can parameterize a curve with corners or cusps with a very smooth parameterization: see Exercise 4.1.

In image processing, we say that a process that introduces new features, such as described in item (2) above, is not *causal*. (This informal definition should not be confused with the use of "causality," as it is used, for example, when speaking about filters: A filter F is said to be causal, or realizable, if the equality of two signals s_0 and s_1 up to time t_0 implies that $Fs_0(t) = Fs_1(t)$ for the same period.)

4.3 Restoring locality and causality

Our main objective is to redefine the smoothing processes so they are local and do not create new singularities. This can be done by alternating a small-scale

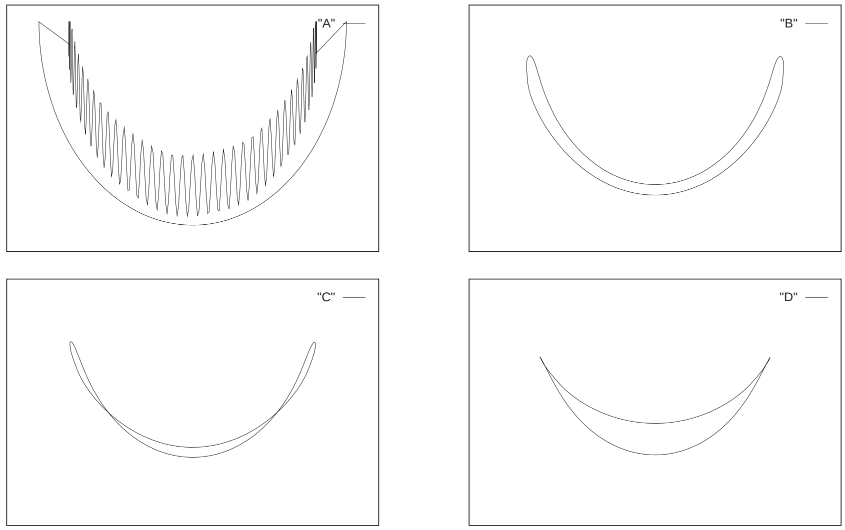

Figure 4.3: Curve evolution by the heat equation. The coordinates of the curves are parameterized by the arc length and then smoothed as real functions of the length using the heat equation. From A to D: the coordinates are smoothed with an increasing scale. Each coordinate function therefore is C^∞; the evolving curve can, however, develop self-crossings (as in C) or singularities (as in D).

linear convolution with a natural renormalization process.

4.3.1 Localizing the dynamic shape method

In the case of dynamic shape analysis, we define an alternate dynamic shape algorithm as follows:

Algorithm 4.1 (The Merriman–Bence–Osher algorithm).

(1) Convolve the characteristic function of the initial shape X_0 with G_h, where h is small.

(2) Define $X_1 = \{\mathbf{x} \mid G_h * \mathbf{1}_{X_0} \geq 1/2\}$.

(3) Set $X_0 = X_1$ and go back to (1).

This is an iterated dynamic shape algorithm. The dynamic shape method itself is an example of a *median filter*, which will be defined in Chapter 10. The Merriman–Bence–Osher algorithm is thus an *iterated median filter* (see Figure 4.4). We will see in Chapters 14 and 15 that median filters have asymptotic properties that are similar to those expressed in Theorem 2.3. In the case of median filters, the associated partial differential equation will be a curvature motion equation (defined in Chapter 12).

Figure 4.4: The Merriman–Bence–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 the same initial data (top, left) as in Figure 4.2. From left to right, top to bottom: smoothing with increasing scales. Notice that the shapes remain separate. In fact, their is no interaction between the evolving shapes. Each one evolves as if the other did not exist.

4.3.2 Renormalized heat equation for curves

In 1992, Mackworth and Mokhtarian noticed the loss of causality when the heat equation was applied to curves [189]. Their method to restore causality looks, at least formally, like the remedy given for the nonlocalization of the dynamic shape method. Instead of applying the heat equation for relatively long times (or, equivalently, convolving the curve $\mathbf{x}$ with the Gaussian G_t for large t), they use the following algorithm:

Algorithm 4.2 (Renormalized heat equation for curves).

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

(2) Let L_n denote the length of the curve $\mathbf{x}_n$ obtained after n iterations and let s_n denote its length parameter. For $n \geq 1$, write $\tilde{\mathbf{x}}_{n+1}(s_n) = G_h * \mathbf{x}_n(s_n)$. Then reparameterize $\tilde{\mathbf{x}}_{n+1}$ by its length parameter $s_{n+1} \in [0, L_{n+1}]$, and denote it by $\mathbf{x}_{n+1}$.

(3) Iterate.

This algorithm is illustrated in Figure 4.5. It should be compared with Figure 4.3.

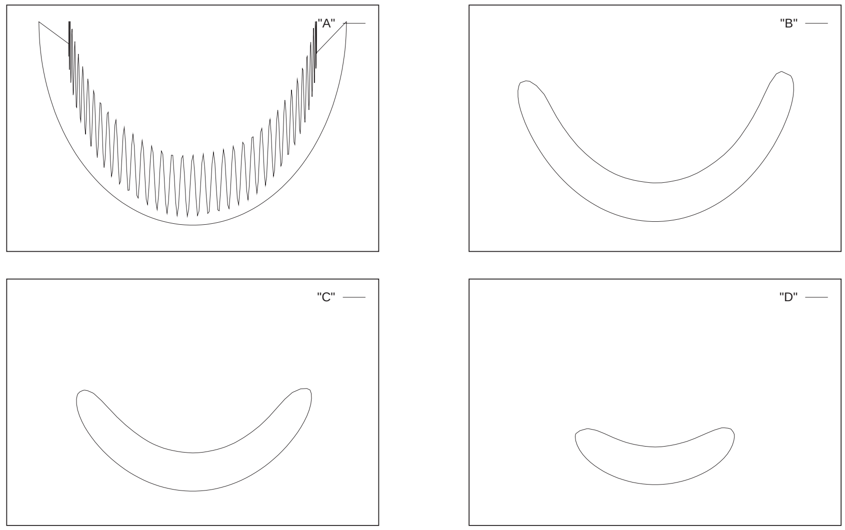

Figure 4.5: Curve evolution by the renormalized heat equation (Mackworth–Mokhtarian). After each smoothing step, the coordinates of the curve are reparameterized by the arc length of the smoothed curve. From A to D: the curve is smoothed with an increasing scale. Note that, in contrast with the linear heat equation (Figure 4.3), the evolving curve shows no singularities and does not cross itself.

Theorem 4.1. *Let* $\mathbf{x}$ *be a* C^2 *curve parameterized by its length parameter* $s \in [0, L]$. *Then for small* h,

$$G_h * \mathbf{x}(s) - \mathbf{x}(s) = h\frac{\partial^2 \mathbf{x}}{\partial s^2} + o(h). \tag{4.2}$$

This theorem is easily checked, see Exercise 4.2

In view of (4.2) and what we have seen regarding asymptotic limits in Theorem 2.3 and Exercise 2.5, it is reasonable to conjecture that, in the asymptotic limit, Algorithm 4.2 will yield the solution of following evolution equation:

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\partial^2 \mathbf{x}}{\partial s^2}, \tag{4.3}$$

where $\mathbf{x}_0 = \mathbf{x}(0, \cdot)$. It is important to note that (4.3) is *not* the heat equation (4.1). Indeed, from Algorithm 4.2 we see that s must denote the length parameter of the evolved curve $\mathbf{x}(t, \cdot)$ at time t. In fact $\partial^2\mathbf{x}/\partial s^2$ has a geometric interpretation as a curvature vector. We will study this nonlinear curve evolution equation in Chapter 12.

4.4 Exercises

Exercise 4.1. Construct a C^∞ mapping $f : [0, 1] \to \mathbb{R}^2$ such that the image of $[0, 1]$ is a square. This shows that a curve can have a C^∞ parameterization without being smooth. ■

Exercise 4.2. Prove Theorem 4.1. If $\mathbf{x}$ is a C^3 function of s, then the result follows directly from Theorem 2.2. The result holds, however, for a C^2 curve. ■

4.5 Comments and references

Dynamic shape, curve evolution, and restoring causality. Our account of the dynamic shape method is based on the well-known paper by Koenderink and van Doorn in which they introduced this notion [174]. The curve evolution by the heat equation is from the first 1986 version of curve analysis proposed by Mackworth and Mokhtarian [188]. See also the paper by Horn and Weldon [143]. There were model errors in the 1986 paper [188] that were corrected by the authors in their 1992 paper [189]. There, they also proposed the correct intrinsic equation. However, this 1992 paper contains several inexact statements about the properties of the intrinsic equation. The correct theorems and proofs can be found in a paper by Grayson written in 1987 [128]. The algorithm that restores causality and locality to the dynamic shape method was discovered by Merriman, Bence, and Osher, who devised this algorithm for a totally different reason: They were looking for a clever numerical implementation of the mean curvature equation [203].

Topological change under smoothing. We have included several figures that illustrate how essential topological properties of an image change when the image is smoothed with the Gaussian. Damon has made a complete analysis of the topological behavior of critical points of an image under Gaussian smoothing [83]. This analysis had been sketched in [291].

Part II

Contrast-Invariant Image Analysis

Chapter 5

Contrast-Invariant Classes of Functions and Their Level Sets

This chapter is about one of the major technological contributions of mathematical morphology, namely the representation of images by their upper level sets. As we shall see in this chapter, this leads to a handy contrast invariant representation of images.

Definition 5.1. *Let $u \in \mathcal{F}$. The level set of u at level $0 \leq \lambda \leq 1$ is denoted by $\mathcal{X}_\lambda u$ and defined by*

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

Strictly speaking, we have called level sets what should more properly be called upper level sets. Several level sets of a digital image are shown in Figure 5.1 and all of the level sets of a synthetic image are illustrated in Figure 5.2. The reconstruction of an image from its level sets is illustrated in Figure 5.3. Two important properties of the level sets of a function follow directly from the definition. The first is that the level sets provide a complete description of the function. Indeed, we can reconstruct u from its level sets $\mathcal{X}_\lambda u$ by the formula

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

This formula is called *superposition principle* as u is being reconstructed by "superposing" its level sets.

Exercise 5.1. Prove the superposition principle. ■

The second important property is that level sets of a function are globally invariant under contrast changes. We say that two functions u and v have the same level sets globally if for every λ there is μ such that $\mathcal{X}_\mu v = \mathcal{X}_\lambda u$, and conversely. Now suppose that a contrast change $g : \mathbb{R} \to \mathbb{R}$ is continuous and increasing. Then it is not difficult to show that $v = g(u)$ and u have the same level sets globally.

Exercise 5.2. Check this last statement for any function u and any continuous increasing contrast change g. ■

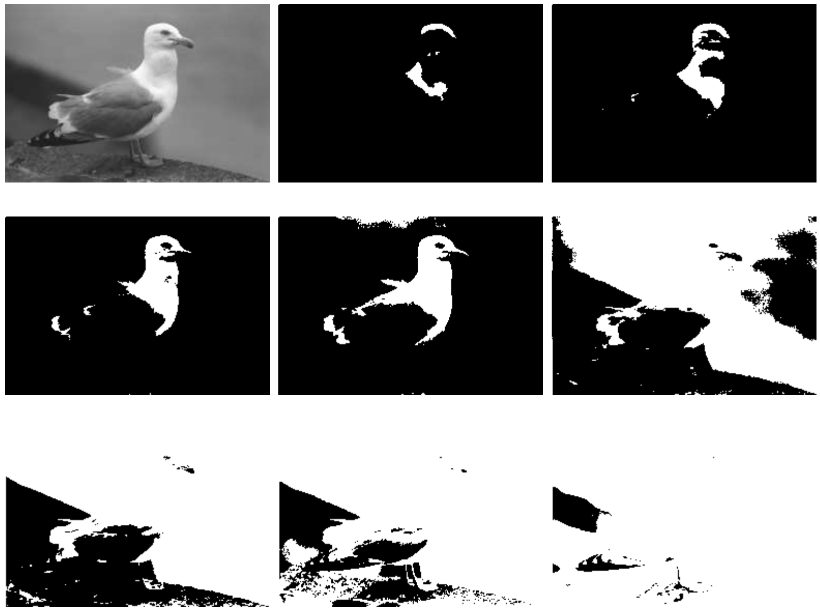

Figure 5.1: Level sets of a digital image. Left to right, top to bottom: We first show an image with range of gray levels from 0 to 255. Then we show eight level sets in decreasing order from $\lambda = 225$ to $\lambda = 50$, where the grayscale step is 25. Notice how essential features of the shapes are contained in the boundaries of level sets, the level lines. Each level set (which appears as white) is contained in the next one, as guaranteed by Proposition 5.2.

Conversely, we shall prove that if the level sets of a function $v \in \mathcal{F}$ are level sets of u, then there is a continuous contrast change g such that $v = g(u)$. This justifies the attention we will dedicate to level sets, as they turn out to contain all of the contrast invariant information about u.

5.1 From an image to its level sets and back

In the next proposition, for a sake of generality, we consider bounded measurable functions on S_N, not just functions in $\mathcal{F}$.

Proposition 5.2. *Let X_λ denote the level sets $\mathcal{X}_\lambda u$ of a bounded measurable function $u : S_N \to \mathbb{R}$. Then the sets X_λ satisfy the following two structural properties:*

(*i*) *If $\lambda > \mu$, then $X_\lambda \subset X_\mu$. In addition, there are two real numbers $\lambda_{max} \geq \lambda_{min}$ so that $X_\lambda = S_N$ for $\lambda < \lambda_{min}$, $X_\lambda = \emptyset$ for $\lambda > \lambda_{max}$.*

(*ii*) *$X_\lambda = \bigcap_{\mu<\lambda} X_\mu$ for every $\lambda \in \mathbb{R}$.*

Conversely, if $(X_\lambda)_{\lambda\in\mathbb{R}}$ is a family of sets of $\mathcal{M}$ that satisfies (i) and (ii), then the level sets of the function u defined by superposition principle,

$$u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in X_\lambda\} \tag{5.1}$$

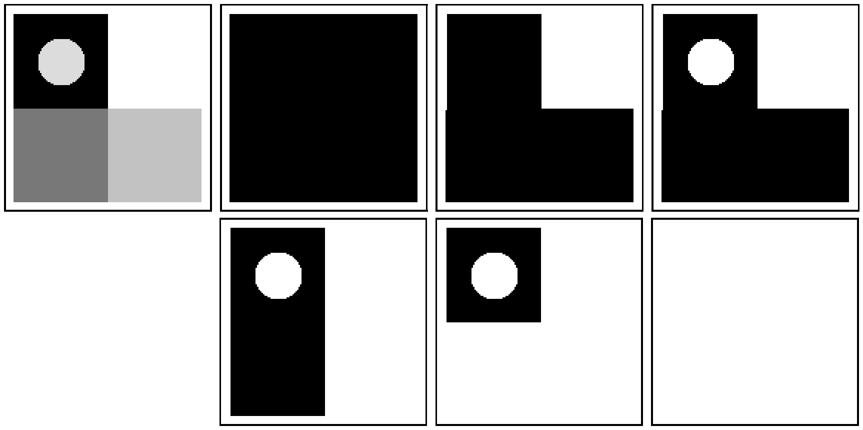

Figure 5.2: A simple synthetic image and all of its level sets (in white) with decreasing levels, from left to right and from top to bottom.

satisfy $\mathcal{X}_\lambda u = X_\lambda$ *for all* $\lambda \in \mathbb{R}$ *and* $\lambda_{min} \le u \le \lambda_{max}$.

Proof. The first part of Relation (i) follows directly from the definition of upper level sets. The second part of (i) works with $\lambda_{min} = \inf u$ and $\lambda_{max} = \sup u$. The relation (ii) follows from the equivalence $u(\mathbf{x}) \ge \lambda \Leftrightarrow u(\mathbf{x}) \ge \mu$ for every $\mu < \lambda$.

Conversely, take a family of subsets $(X_\lambda)_{\lambda\in\mathbb{R}}$ satisfying (i) and (ii) and define u by the superposition principle. Let us show that $X_\lambda = \mathcal{X}_\lambda u$. Take first $\mathbf{x} \in X_\lambda$. Then it follows from the definition of u that $u(\mathbf{x}) \ge \lambda$, and hence $\mathbf{x} \in \mathcal{X}_\lambda u$. Thus, $X_\lambda \subset \mathcal{X}_\lambda u$. Conversely, let $\mathbf{x} \in \mathcal{X}_\lambda u$. Then $u(\mathbf{x}) = \sup\{\nu \mid \mathbf{x} \in X_\nu\} \ge \lambda$. Consider any $\mu < \lambda$. Then there exists a μ' such that $\mu < \mu' \le \sup\{\nu \mid \mathbf{x} \in X_\nu\}$ and $\mathbf{x} \in X_{\mu'}$. It follows from (i) that $\mathbf{x} \in X_\mu$. Since μ was any number less that λ, we conclude by using (ii) that $\mathbf{x} \in \bigcap_{\mu<\lambda} X_\mu = X_\lambda$. It is easily checked that $\lambda_{min} \le u \le \lambda_{max}$. □

Exercise 5.3. Check the last statement of the preceding proof, that $\lambda_{min} \le u \le \lambda_{max}$. ■

5.2 Contrast changes and level sets

Practical aspects of contrast changes are illustrated in Figures 5.4, 5.5, 5.6, and 5.7, which illustrate how insensitive our perception of images is to contrast changes, even when they are flat on some interval. When this happens, some information on the image is even lost, as several grey levels melt together.

Definition 5.3. *Any nondecreasing continuous surjection* $g : \mathbb{R} \to \mathbb{R}$ *will be called a contrast change.*

Exercise 5.4. Remark that $g(s) \to \pm\infty$ as $s \to \pm\infty$. Check that if $u \in \mathcal{F}$ and g is a contrast change, then $g(u) \in \mathcal{F}$. ■

In case g is increasing, g has an inverse contrast change g^{-1}. In case g is flat on some interval, we shall be happy with a pseudo-inverse for g.

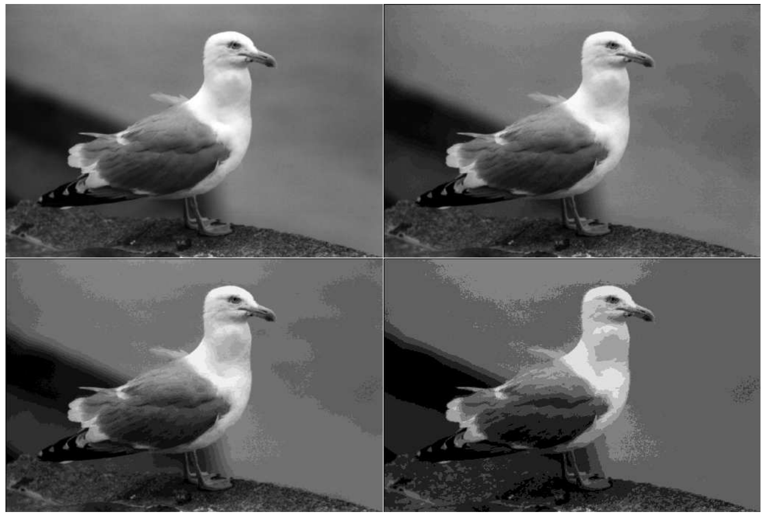

Figure 5.3: Reconstruction of an image from its level sets: an illustration of Proposition 3.2. We use four different subsets of the image's level sets to give four reconstructions. Top, left: all level sets; top, right: all level sets whose gray level is a multiple of 8; bottom, left: multiples of 16; bottom, right: multiples of 32. Notice the relative stability of the image shape content under these drastic quantizations of the gray levels.

Definition 5.4. *The pseudo-inverse of any contrast change $g : \mathbb{R} \to \mathbb{R}$ is defined by*

$$g^{(-1)}(\lambda) = \inf\{r \in \mathbb{R} \ \mid g(r) \geq \lambda\}.$$

Exercise 5.5. Check that g^{-1} is finite on $\mathbb{R}$ and tends to $\pm\infty$ as $s \to \pm\infty$. Give an example of g such that g^{-1} is not continuous. ■

Exercise 5.6. Compute and draw $g^{(-1)}$ for the function $g(s) = \max(0, s)$. Notice that such a function is ruled out by our conditions at infinity for contrast changes. ■

Lemma 5.5. *Let $g : \mathbb{R} \to \mathbb{R}$ be a contrast change. Then for every $\lambda \in \mathbb{R}$, $g(g^{(-1)})(\lambda) = \lambda$ and*

$$g(s) \geq \lambda \text{ if and only if } s \geq g^{(-1)}(\lambda). \tag{5.2}$$

Proof. The first relation follows immediately from the continuity of g. If $g(s) \geq \lambda$, then $s \geq g^{(-1)}(\lambda)$ by the definition of $g^{(-1)}(\lambda)$. Conversely, if $s \geq g^{(-1)}(\lambda)$, then $g(s) \geq g(g^{(-1)}(\lambda)) = \lambda$ and thus $g(s) \geq \lambda$. □

Theorem 5.6. *Let $u \in \mathcal{F}$ and g be a contrast change. Then any level set of $g(u)$ is a level set of u. More precisely, for $\lambda \in \mathbb{R}$,*

$$\mathcal{X}_\lambda g(u) = \mathcal{X}_{g^{(-1)}(\lambda)} u. \tag{5.3}$$

Proof. The proof is read directly from Lemma 5.5 by taking $s = u$. □

The next result is a converse statement to Theorem 5.6.

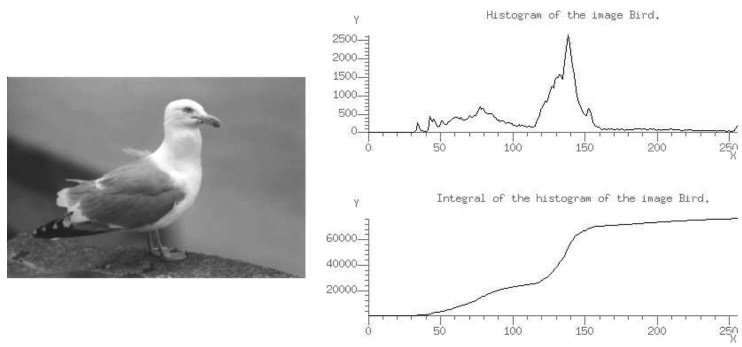

Figure 5.4: The histogram of the image Bird. For each $i \in \{0, 1, \ldots, 255\}$, we display (above, right) the function $h(i) = \text{Card}\,\{\mathbf{x} \mid u(\mathbf{x}) = i\}$. The function below is given by $g(i) = \text{Card}\,\{\mathbf{x} \mid u(\mathbf{x}) \le i\}$, an integral of h. It provides an indication about the overall contrast of the image and about the contrast change imposed by the sensors. The pseudo inverse function $g^{(-1)}$ can be used as a contrast change to create an image $g^{(-1)}(u)$ with a flat histogram.

Theorem 5.7. *Let u and $v \in \mathcal{F}$ such that every level set of v is a level set of u. Then $v = g(u)$ for some contrast change g.*

Proof. One can actually give an explicit formula for g, namely, for every $\mu \in u(S_N)$,

$$g(\mu) = \sup\{\lambda \in v(S_N) \mid \mathcal{X}_\mu u \subset \mathcal{X}_\lambda v\}. \tag{5.4}$$

For $\mu \notin u(S_N)$, we can easily extend g into an nondecreasing function such that $g(\pm\infty) = \pm\infty$). (Take (e.g.) g piecewise affine). Note that $\nu > \mu$ implies that $g(\nu) \ge g(\mu)$. Let us first show that $\inf v \le g(\mu) \le \sup v$. Set

$$\Lambda := \{\lambda \mid \mathcal{X}_\mu u \subset \mathcal{X}_\lambda v\}.$$

Λ is not empty because $\mathcal{X}_{\inf v} = S_N$ and therefore $\inf v \in \Lambda$. Thus $g(\mu) = \sup \Lambda \ge \inf v$. On the other hand $\mathcal{X}_{\sup v + \varepsilon} v = \emptyset$ for every $\varepsilon > 0$. Since $\mu \in u(S_N)$, $\mathcal{X}_\mu u \ne \emptyset$ and therefore $g(\mu) = \sup \Lambda \le \sup v$.

Step 1: Proof that $v(\mathbf{x}) \ge g(u(\mathbf{x}))$. By Proposition 5.2($i$) Λ has the form $(-\infty, \sup \Lambda)$ or $(-\infty, \sup \Lambda]$. But by Proposition 5.2(ii), $\mathcal{X}_{\sup \Lambda} v = \bigcap_{\lambda < \sup \Lambda} \mathcal{X}_\lambda v$, and this implies by the definition of Λ that $g(\mu) = \sup \Lambda \in \Lambda$. Thus,

$$\mathcal{X}_\mu u \subset \mathcal{X}_{g(\mu)} v. \tag{5.5}$$

Given $\mathbf{x} \in S_N$, let $\mu = u(\mathbf{x})$ in (5.5). Then,

$$\mathcal{X}_{u(\mathbf{x})} u \subset \mathcal{X}_{g(u(\mathbf{x}))} v.$$

Since $\mathbf{x} \in \mathcal{X}_{u(\mathbf{x})} u$, we conclude that $\mathbf{x} \in \mathcal{X}_{g(u(\mathbf{x}))} v = \{\mathbf{y} \mid v(\mathbf{y}) \ge g(u(\mathbf{x}))\}$.

Step 2: Proof that $v(\mathbf{x}) \le g(u(\mathbf{x}))$. Given $\mathbf{x} \in S_N$, we translate the assumption with $\lambda = v(\mathbf{x})$ as follows: There exists a $\mu(\mathbf{x}) \in \mathbb{R}$ such that

$$\mathcal{X}_{v(\mathbf{x})} v = \{\mathbf{y} \mid u(\mathbf{y}) \ge \mu(\mathbf{x})\} = \mathcal{X}_{\mu(\mathbf{x})} u. \tag{5.6}$$

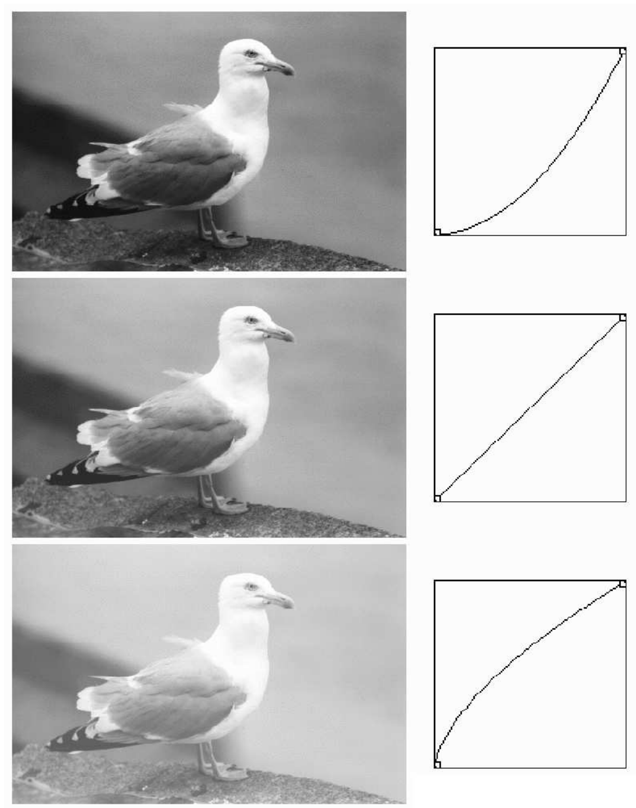

Figure 5.5: Contrast changes and an equivalence class of images. The three images have exactly the same level sets and level lines, but their level sets are mapped onto three different gray-level scales. The graphs on the right are the graphs of the contrast changes $u \mapsto g(u)$ that have been applied to the initial gray levels. The first one is concave; it enhances the darker parts of the image. The second one is the identity; it leaves the image unaltered. The third one is convex; it enhances the brighter parts of the image. Software allows one to manipulate the contrast of an image to obtain the best visualization. From the image analysis viewpoint, image data should be considered as an equivalence class under all possible contrast changes.

Since $\mathbf{x} \in \mathcal{X}_{v(\mathbf{x})}v$, we know that $\mathbf{x} \in \mathcal{X}_{\mu(\mathbf{x})}u$. Thus, $u(\mathbf{x}) \geq \mu(\mathbf{x})$, and $\mathcal{X}_{u(\mathbf{x})}u \subset \mathcal{X}_{\mu(\mathbf{x})}u = \mathcal{X}_{v(\mathbf{x})}v$. This last relation implies by the definition of g that $v(\mathbf{x}) \leq g(u(\mathbf{x}))$.

Step 3: Proof that g is continuous. Recall that the image of a connected set by a continuous function is connected. Thus $u(S_N)$ is an interval of $\mathbb{R}$ and so is $v(S_N)$. Since $g(u) = v$, $g(u(S_N)) = v(S_N)$ is an interval. Now, a nondecreasing function is continuous on an interval if and only if its range is connected. Thus g is continuous on $u(S_N)$ and so is its extension to $\mathbb{R}$. □

Exercise 5.7. Prove the last statement in the theorem, namely that "a nondecreasing function is continuous on an interval if and only if its range is connected". ■

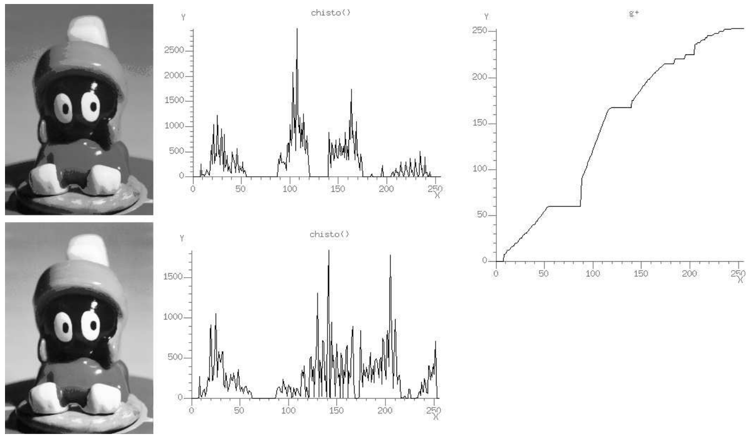

Figure 5.6: The two images (left) have the same set of level sets. The contrast change that maps the upper image onto the lower image is displayed on the right. It corresponds to one of the possible g functions whose existence is stated in Corollary 3.14. The function g may be locally constant on intervals where the histogram of the upper image is zero (see top, middle graph). Indeed, on such intervals, the level sets are invariant.

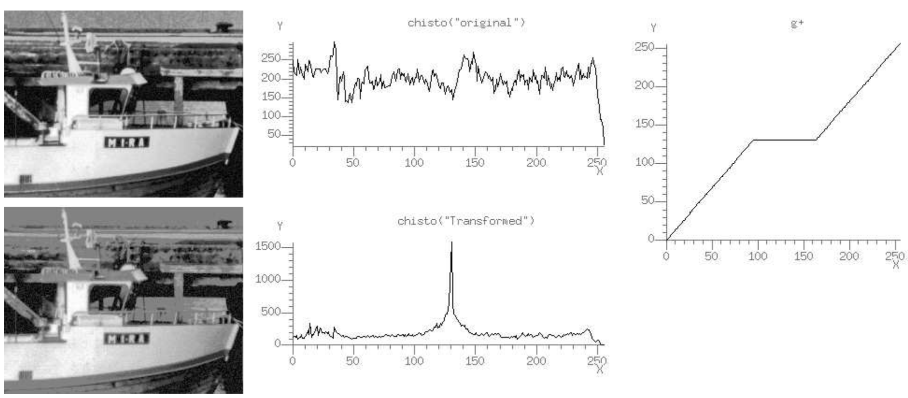

Figure 5.7: The original image (top, left) has a strictly positive histogram (all gray levels between 0 and 255 are represented). Therefore, if any contrast change g that is not strictly increasing is applied, then some data will be lost. Every level set of the transformed image $g(u)$ is a level set of the original image; however, the original image has more level sets than the transformed image.

Exercise 5.8. By reading carefully the steps 1 and 2 of the proof of Theorem 5.7, check that this theorem applies with u and v just bounded and measurable on S_N. Then one has still has $v = g(u)$ with g defined in the same way. Of course g is still nondecreasing but not necessarily continuous. Find a simple example of functions u and v such that g is not continuous. ∎

5.3 Exercises

Exercise 5.9. This exercise gives a way to compute the function g such that $v = g(u)$ defined in the proof of Theorem 5.7 in terms of the repartition functions of u and v. Let G be a Gauss function defined on $\mathbb{R}^N$ such that $\int_{\mathbb{R}^N} G(\mathbf{x})d\mathbf{x} = 1$. For every measurable subset of $\mathbb{R}^N$, set $|A|_G := \int_A G(\mathbf{x})d\mathbf{x}$. Let u be a bounded continuous function on $\mathbb{R}^N$. We can associate with u its *repartition function* $h_u(\lambda) := |\mathcal{X}_\lambda u|_G$. Show that $h_u : \lambda \in [\inf u, \sup u] \to h_u(\lambda)$ is strictly decreasing. Show that it can have jumps but is left-continuous, that is $h_u(\lambda) = \lim_{\mu\uparrow\lambda} h_u(\mu)$. Define for every non increasing function h a pseudo inverse by $h^{((-1))}(\mu) := \sup\{\lambda \mid h(\lambda) \geq \mu\}$. Show that $h^{((-1))}$ is non increasing and that $h^{((-1))} \circ h(\mu) \geq \mu$, and that if h is left-continuous, $h \circ h^{((-1))}(\mu) \geq \mu$. Using (5.4) prove that $g = h_v^{((-1))} \circ h_u$.
Hint: prove that $g(\mu) = \sup\{\lambda \mid |\mathcal{X}_\mu u|_G \leq |\mathcal{X}_\lambda v|_G\}$. ∎

Exercise 5.10. Let u be a real-valued function. If $(\mu_n)_{n\in\mathbb{N}}$ is an increasing sequence that tends to λ, prove that

$$\mathcal{X}_\lambda u = \bigcap_{n\in\mathbb{N}} \mathcal{X}_{\mu_n} u \tag{5.7}$$

$$\{\mathbf{x} \mid u(\mathbf{x}) > \lambda\} = \bigcup_{\mu>\lambda} \mathcal{X}_\mu u. \tag{5.8}$$

∎

5.4 Comments and references

Contrast invariance and level sets. It was Wertheimer who noticed that the actual local values of the gray levels in an image could not be relevant information for the human visual system [287]. Contrast invariance is one of the fundamental model assumptions in mathematical morphology. The two basic books on this subject are Matheron [202] and Serra [253, 255]. See also the fundamental paper by Serra [254]. Ballester et al. defined an "image intersection" whose principle is to keep all pieces of bilevel sets common to two images [31]. (A bilevel set is of the form $\{\mathbf{x} \mid \lambda \leq u(\mathbf{x}) \leq \mu\}$.) Monasse and Guichard developed a *fast level set transform* (FLST) to associate with every image the inclusion tree of connected components of level sets [207]. They show that the inclusion trees of connected upper and lower level sets can be fused into a single inclusion tree; among other applications, this tree can be used for image registration. See Monasse [206].

Contrast changes. The ability to vary the contrast (to apply a contrast change) of a digital image is a very useful tool for improving image visualization. Professional image processing software has this capability, and it is also found in popular software for manipulating digital images. For more about contrast changes that preserve level sets, see [63]. Many reference on contrast-invariant operators are given at the end of Chapter 7.

Chapter 6

Specifying the contrast of images

Midway image equalization means any method giving to a pair of images a similar histogram, while maintaining as much as possible their previous grey level dynamics. The comparison of two images, in order to extract a mutual information, is one of the main themes in computer vision. The pair of images can be obtained in many ways: they can be a stereo pair, two images of the same object (a painting for example), multi-channel images of the same region, images of a movie, etc. This comparison is perceptually greatly improved if both images have the same grey level dynamics. In addition, a lot of image comparison algorithms, based on grey level, take as basic assumption that intensities of corresponding points in both images are equal. As it is well known by experts in stereo vision, this assumption is generally false for stereo pairs and deviations from this assumption cannot even be modeled by affine transforms [78]. Consequently, if we want to compare visually and numerically two images, it is useful to give them first the same dynamic range and luminance.

In all of this applicative chapter the images $u(\mathbf{x})$ and $v(\mathbf{x})$ are defined on a domain which is the union of M pixels. The area of each pixel is equal to 1. The images are discrete in space and values: they attain values in a finite set l and they are constant on each pixel of the domain. We shall call such images *discrete images*. The piecewise constant interpolation is a very bad image interpolation. It is only used here for a fast handling of image histograms. For other scopes, better interpolation methods are of course necessary.

Definition 6.1. *Let u be a discrete image. We call cumulative histogram of u the function $H_u : \mathbb{L} \to \mathbb{M} := [0, M] \cap \mathbb{N}$ defined by*

$$H_u(l) := \text{meas}(\{\mathbf{x} \mid u(\mathbf{x}) \leq l\}).$$

This cumulative histogram is the integral of the *histogram* of the image, the function $h(l) = \text{meas}(\{\mathbf{x} \mid u(\mathbf{x}) = l\})$. Figures 5.4, 5.6 and the first line of Figure 6.1. show the histograms of some images and their cumulative histograms. In fact Figure 5.7 shows first the histogram and then the modified histogram after a contrast change has been applied. These experiments illustrate the robustness of image relevant information to contrast changes and even to the removal of some level sets, when the contrast change is flat on an interval. Such experiments suggest that one can *specify* the histogram of a given image by applying the adequate contrast change. Before proceeding, we have to define the pseudo-inverses of a discrete function.

Proposition 6.2. *Let $\varphi : \mathbb{L} \to \mathbb{M}$ be a nondecreasing function from a finite set of values into another. Define two pseudo-inverse functions for φ :*

$$\varphi^{(-1)}(l) := \inf\{s \mid \varphi(s) \geq l\} \text{ and } \varphi^{((-1))}(l) := \sup\{s \mid \varphi(s) \leq l.\}$$

Then one has the following equivalences:

$$\varphi(s) \geq l \Leftrightarrow s \geq \varphi^{(-1)}(l), \quad \varphi(s) \leq l \Leftrightarrow s \leq \varphi^{((-1))}(l) \tag{6.1}$$

and the identity

$$(\varphi^{(-1)})^{((-1))} = \varphi. \tag{6.2}$$

Proof. The implication $\varphi(s) \geq l \Rightarrow s \geq \varphi^{(-1)}(l)$ is just the definition of $\varphi^{(-1)}$. The converse implication is due to the fact that the infimum on a a finite set is attained. Thus $\varphi(\varphi^{(-1)}(l)) \geq l$ and therefore $s \geq \varphi^{(-1)}(l) \Rightarrow \varphi(s) \geq l$. The identity (6.2) is a direct consequence of the equivalences (6.1). Indeed,

$$s \leq (\varphi^{(-1)})^{((-1))}(l) \Leftrightarrow \varphi^{(-1)}(s) \leq l \Leftrightarrow s \leq \varphi(l).$$

□

Exercise 6.1. Prove that if φ is increasing, $\varphi^{(-1)} \circ \varphi(l) = l$ and $\varphi^{((-1))} \circ \varphi(l) = l$. If φ is surjective, $\varphi \circ \varphi^{(-1)} = l$ and $\varphi \circ \varphi^{((-1))}(l) = l$. ■

Proposition 6.3. *Let φ be a discrete contrast change and set $\tilde{u} := \varphi(u)$. Then*

$$H_{\tilde{u}} = H_u \circ \varphi^{((-1))}.$$

Proof. By (6.1), $\tilde{u} \leq l \Leftrightarrow u \leq \varphi^{((-1))}(l)$. Thus by the definitions of H_u and $H_{\tilde{u}}$,

$$H_{\tilde{u}}(l) = \text{meas}(\{\mathbf{x} \mid \tilde{u} \leq l\}) = \text{meas}(\{\mathbf{x} \mid u(\mathbf{x}) \leq \varphi^{((-1))}(l)\}) = H_u \circ \varphi^{((-1))}(l).$$

□

Let $G : \mathbb{L} \to \mathbb{M} := [0, 1, \ldots, M]$ be any discrete nondecreasing function. Can we find a contrast change $\varphi : \mathbb{L} \to \mathbb{L}$ such that the cumulative histogram of $\varphi(u)$, $H_{\varphi(u)}$ becomes equal to G? Not quite: if for instance u is constant its cumulative histogram is a one step function and Proposition 6.3 implies that $H_{\varphi(u)}$ will also be a one step function. More generally if u attains k values, then $\varphi(u)$ attains k values or less. Hence its cumulative histogram is a step function with $k+1$ steps. Yet, at least formally, the functional equation given by Proposition 6.3, $H_u \circ \varphi^{-1} = G$, leads to $\varphi = G^{-1} \circ H_u$. We know that we cannot get true inverses but we can involve pseudo-inverses. Thus, we are led to the following definition:

Proposition 6.4. *Let $G : \mathbb{L} \to \mathbb{M}$ be a nondecreasing function. We call specification of u on the cumulative histogram G the image*

$$\tilde{u} := G^{((-1))} \circ H_u(u).$$

Exercise 6.2. Prove that if G and H are one to one, then the cumulative histogram of $\tilde{u}$ is G. ■

Definition 6.5. *Let, for* $l \in [0, L] \cap I\!N$, $G(l) = \lfloor \frac{M}{L} l \rfloor$, *where* $\lfloor r \rfloor$ *denotes the largest integer smaller than* r. *Then* $\tilde{u} := G^{((-1))} \circ H_u(u)$ *is called the* uniform equalization of u. *If* v *is another discrete image and one takes* $G = H_v$, $\tilde{u} := H_v^{((-1))} \circ H_u(u)$ *is called the* specification of u on v.

When H_u is one to one, any specified cumulative histogram G. Otherwise, the above definitions do the best that can be expected and are actually quite efficient. For instance in the "marshland experiment" (Figure 6.1) the equalized histogram and its cumulative histogram are displayed on the second row. The cumulative histogram is very close to its goal, the linear function. The equalized histogram does not look flat but a sliding average of it would look almost flat. Yet it is quite dangerous to specify the histogram of an image with an arbitrary

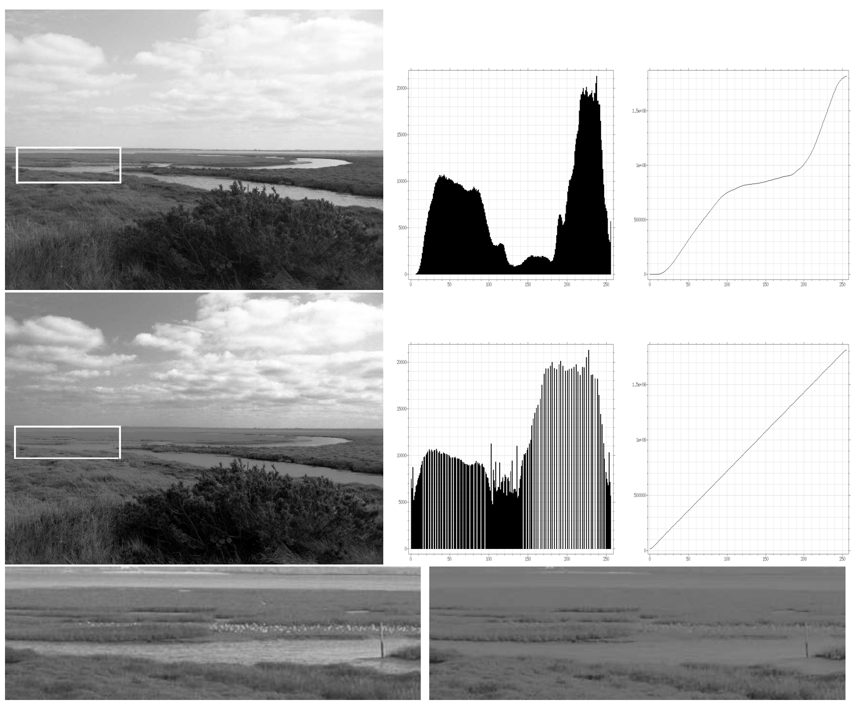

Figure 6.1: *First row: Image* u*, the corresponding grey level histogram* h_u*, and the cumulative histogram* H_u*. Second row: Equalized image* $H_u(u)$*, its histogram and its cumulative histogram. In the discrete case, histogram equalization flattens the histogram as much as possible. We see on this example that image equalization can be visually harmful. In this marshland image, after equalization, the water is no more distinguishable from the vegetation. The third row shows a zoom on the rectangular zone, before and after equalization.*

histogram specification. This fact is illustrated in Figures 6.1 and 6.2 where a uniform equalization erases existing textures by making them too flat (Figure 6.1) but also enhances the quantization noise in low contrasted regions and produces artificial edges or textures (see Figure 6.2).

Figure 6.2: *Effect of histogram equalization on the quantization noise. On the left, the original image. On the right, the same image after histogram equalization. The effect of this equalization on the dark areas (the piano, the left part of the wall), which are low contrasted, is perceptually dramatic. We see many more details but the quantization noise has been exceedingly amplified.*

6.1 Midway equalization

We have seen that if one specifies u on v, then u inherits roughly the histogram of v. It is sometimes more adequate to bring the cumulative histograms of u and v towards a cumulative histogram which would be "midway" between both. Indeed, if we want to compare visually and numerically two images, it is useful to give them first the same dynamic range and luminance. Thus we wish:

- From two images u and v, construct by contrast changes two images $\tilde{u}$ and $\tilde{v}$, which have a similar cumulative histogram.
- This common cumulative histogram h should stand "midway" between the previous cumulative histograms of u and v, and be as close as possible to each of them. This treatment must avoid to favor one cumulative histogram rather than the other.

Definition 6.6. *Let u and v be two discrete images. Set*

$$\Phi := \frac{1}{2}\left(H_u^{(-1)} + H_v^{(-1)}\right).$$

We call midway cumulative histogram *of u and v the function*

$$G := \Phi^{((-1))} = \left(\frac{1}{2}(H_u^{(-1)} + H_v^{(-1)})\right)^{((-1))} \tag{6.3}$$

and "midway specifications" of u and v the functions $\tilde{u} := \Phi \circ H_u(u)$ and $\tilde{v} := \Phi \circ H_v(v)$.

Exercise 6.3. Prove the last statement by application of Proposition 6.4. ■

Exercise 6.4. Let u and v be two constant images, whose values are a and b. Prove that their "midway" function is the right one, namely a function w which is constant and equal to $\frac{a+b}{2}$. ■

Exercise 6.5. Prove that if we take as a definition of the midway histogram

$$G := \left(\frac{1}{2}(H_u^{((-1))} + H_v^{((-1))})\right)^{(-1)},$$

then for two constant images $u = a$ and $v = b$ the midway image is constant and equal to $[1/2(a+b) - 1]$. ! This proves that Definition 6.6 is better. ■

Exercise 6.6. Prove that if u is a discrete image and f and g two nondecreasing functions, then the midway image of $f(u)$ and $g(u)$ is $\frac{f(u)+g(u)}{2}$. ■

Exercise 6.7. If we want the "midway" cumulative histogram H to be a compromise between H_u and H_v, the most elementary function that we could imagine is their average, which amounts to average their histograms as well. However, the following example proves that this idea is not judicious at all.

Consider two images whose histograms are "crenel" functions on two disjoint intervals, for instance $u(\mathbf{x}) := ax$, $v(\mathbf{x}) = bx + c$. Compute a, b, c in such a way that h_u and h_v have disjoint supports. Then compute the specifications of u and v on the mean cumulative histogram $G := \frac{H_u+H_v}{2}$. Compare with their specifications on the midway cumulative histogram. ■

6.2 Experimenting midway equalization on image pairs

Results on a stereo pair

The top of Figure 6.3 shows a pair of aerial images in the region of Toulouse. Although the angle variation between both views is small, and the photographs are taken at nearly the same time, we see that the lightning conditions vary significantly (the radiometric differences can also come from a change in camera settings). The second line shows the result of the specification of the histogram of each image on the other one. The third line shows both images after equalization.

If we scan some image details, as illustrated on Figure 6.4, the damages caused by a direct specification become obvious. Let us specify the darker image on the brightest one. Then the information loss, due to the reduction of dynamic range, can be detected in the brightest areas. Look at the roof of the bright building in the top left corner of the image (first line of Figure 6.4): the chimneys project horizontal shadows on the roof. In the specified image, these shadows have almost completely vanished, and we cannot even discern the presence of a chimney anymore. In the same image after equalization, the shadows are still entirely recognizable, and their size reduction remains minimal. The second line of Figure 6.4 illustrates the same phenomenon, observed in the bottom center of the image. The structure present at the bottom of the image has completely disappeared after specification and remains visible after midway equalization. These examples show how visual information can be lost by specification and how midway algorithms reduce significantly this loss.

Multi-Channel images

The top of Figure 6.5 shows two pieces of multi-channel images of Toulouse. The first one is extracted from the blue channel, and the other one from the infrared

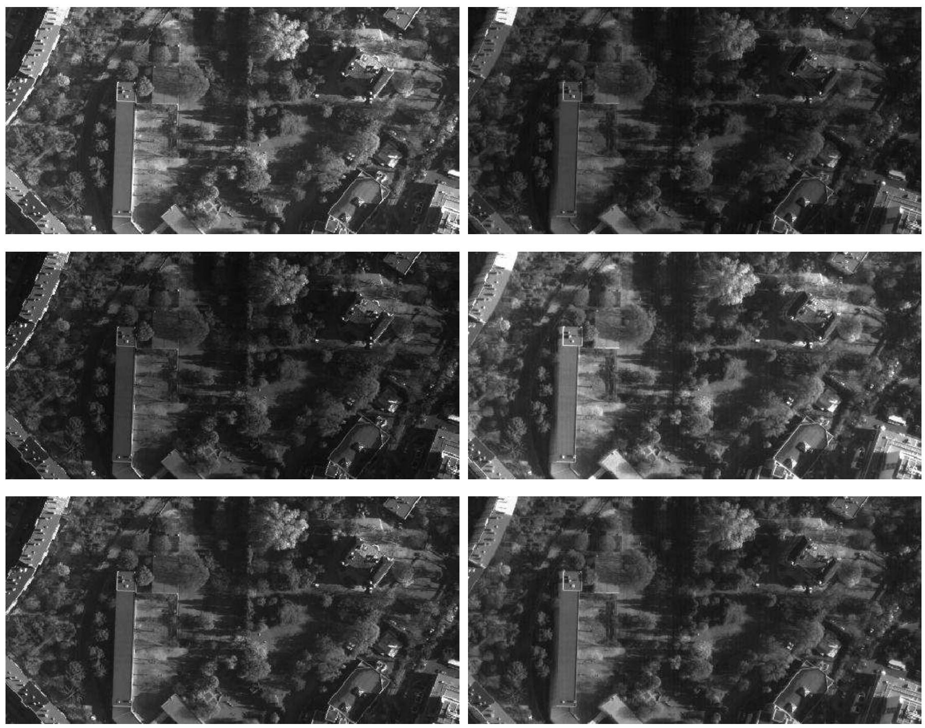

Figure 6.3: Stereo pair: two pieces of aerial images of a region of Toulouse. Same images after specification of their histograms on each other (left: the histogram of the first image has been specified on the second, and right: the histogram of the second image has been specified on the first). Stereo pair after midway equalization.

Figure 6.4: *Extracts from the stereo pair shown on Figure 6.3. From left to right: in the original image, in the specified one, in the original image after midway equalization. Notice that no detail is lost in the midway image, in contrast with the middle image.*

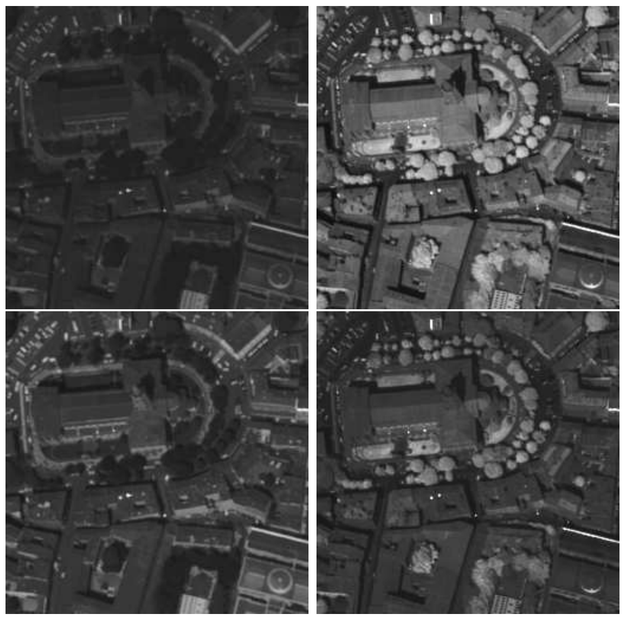

Figure 6.5: *First line: two images of Toulouse (blue and infrared channel). Second line: same images after midway equalization.*

channel. The second and third line of the same figure show the same images after midway equalization. The multichannel images have the peculiarity to present contrast inversions : for instance, the trees appear to be darker than the church in the blue channel, and are naturally brighter than the church in the infrared channel. The midway equalization being limited to increasing contrast changes, it obviously cannot handle these contrast inversions. In spite of these contrast inversions, the results remain visually good, which underlines the robustness of the method gives globally a good equalization.

Photographs of the same painting

The top of Figure 6.6 shows two different snapshots of the same painting, *Le Radeau de la Méduse*[1], by Théodore Géricault (small web public versions). The

[1]Muse du Louvre, Paris.

second one is brighter and seems to be damaged at the bottom left. The second line shows the same couple after midway equalization. Finally, the last line of Figure 6.6 shows the difference between both images after equalization. We see clear differences around the edges, due to the fact that the original images are not completely similar from the geometric point of view.

6.2.1 Movie equalization

One can define a midway cumulative histogram to an arbitrary number of images. This is extremely useful for the removal of flicker in old movies. Flicker has multiple causes, physical, chemical or numerical. The overall contrast of successive images of the same scene in a movie oscillates, some images being dark and others bright. Our main assumption is that image level sets are globally preserved from one image to the next, even if their level evolves. This leads to the adoption of a movie equalization method preserving globally all level sets of each image. We deduce from Theorem 5.7 in the previous chapter that the correction must be a global contrast change on each image. Thus the only left problem is to *specify* a common cumulative histogram (and therefore a common histogram) to all images of a given movie scene. Noticing that the definition of G in (6.3) for two images simply derives from a mean, its generalization is easy. Let us denote $u(t, \mathbf{x})$ the movie (now a discrete time variable has been added) and by H^t the cumulative histogram function of $\mathbf{x} \to u(t, \mathbf{x})$ at time t. Since flicker is localized in time, the idea is to define a time dependent cumulative histogram function K_t^h which will the "midway" cumulative histogram of the cumulative histograms in an interval $[t-h, t+h]$. Of course the linear scale space theory of Chapter 2 applies here. The ideal average is gaussian. Hence the following definition.

Definition 6.7. *Let $u(t, \mathbf{x})$ be a movie and denote by H_t the cumulative histogram of $u(t) : \mathbf{x} \to u(t, \mathbf{x})$. Consider a discrete version of the 1-D gaussian $G_h(t) = \frac{1}{(4\pi h)^{\frac{1}{2}}} e^{-\frac{t^2}{4h}}$. Set*

$$\Phi_{(t,l)} := \int G_h(t-s)(H_s^{(-1)})(l)ds.$$

We call "midway gaussian cumulative histogram at scale h" of the movie $u(t, \mathbf{x})$ the time dependent cumulative histogram

$$\mathbb{G}_{(t,l)} := \Phi_{(t,l)}^{((-1))} = \left(\int G_h(t-s)(H_s^{(-1)})(l)ds \right)^{((-1))} \tag{6.4}$$

and "midway specification" of the movie $u(t)$ the function $\tilde{u}(t) := \Phi \circ H_{u(t)}(u(t))$. If $H_{u(t)}$ is surjective, then $\tilde{u}(t)$ has $G_{(t,l)}$ as common cumulative histogram.

Notice that this is a straightforward extension of Definition 6.6.

The implementation and experimentation is easy. We simply show in Figure 6.7 three images of Chaplin' s film *His New Job*, taken at equal intervals of time. This extract of the film suffers from a severe real flicker. This flicker is corrected at the scale where, after gaussian midway equalization, the image mean becomes nearly constant through the sequence. The effects of this equalization are usually excellent. They are easily extended to color movies by processing each channel independently.

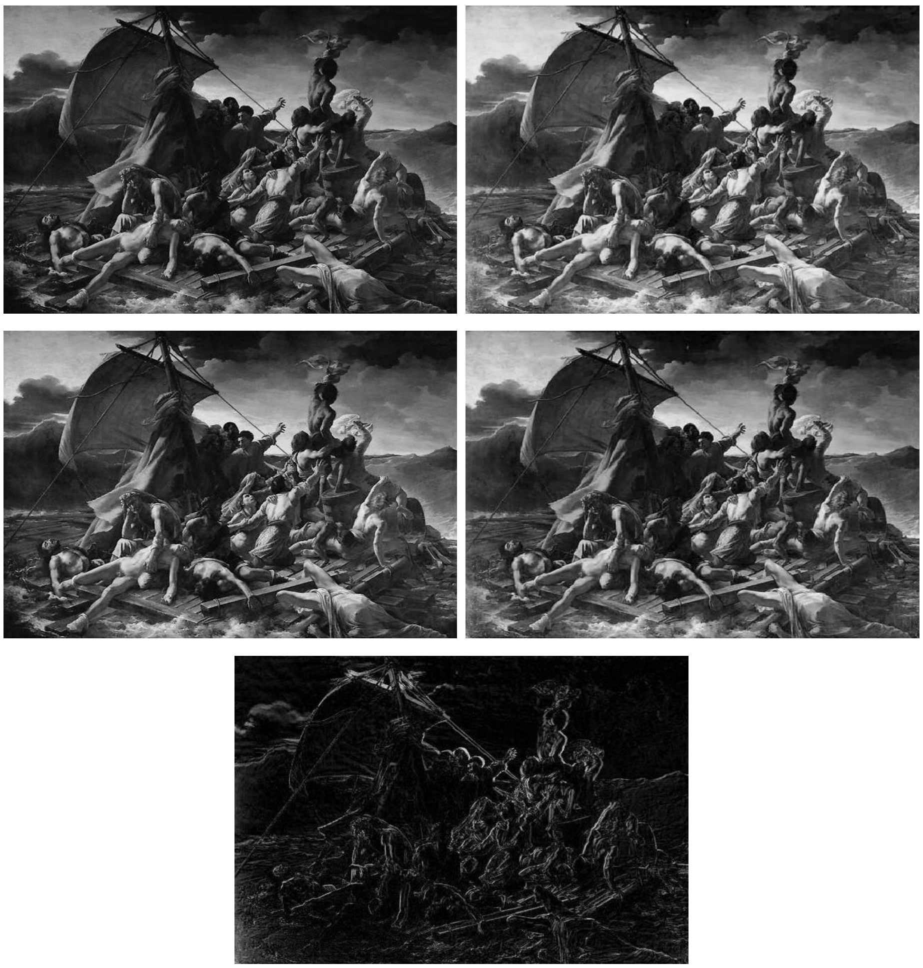

Figure 6.6: Two shots of the *Radeau de la Méduse, by Géricault.* The same images after midway equalization. Image of the difference between both images after equalization. The boundaries appearing in the difference are mainly due to the small geometric distortions between the initial images.

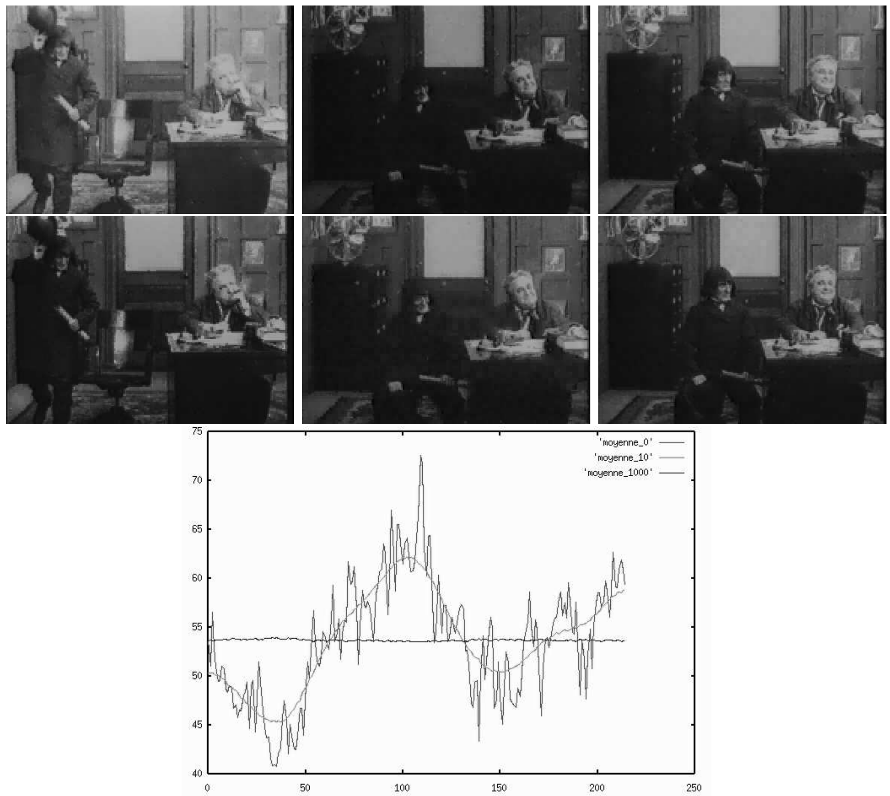

Figure 6.7: *(a) Three images of Chaplin' s film His New Job, taken at equal intervals of time. This extract of the film suffers from a severe real flicker. (b) Same images after the Scale-Time Equalization at scale* $s = 100$. *The flicker observed before has globally decreased. (c) Evolution of the mean of the current frame in time and at three different scales. The most oscillating line is the mean of the original sequence. The second one is the mean at scale* $s = 10$. *The last one, almost constant, corresponds to the large scale* $s = 1000$. *As expected the mean function is smoothed by the heat equation.*

6.3 Comments and references

Histogram specification As we have seen *histogram specification* [127] can be judicious if both images have the same kind of dynamic range. For the same reason as in equalization, this method can also product contouring artifacts. The midway theory is essentially based on Julie Delons' PhD and papers [87], [88] where she defines two midway histogram interpolation methods. One of them, the *square root* method involves the definition of a square root of any nondecreasing function g, namely a function g such that $f \circ f = g$. Assume that u and v come from the same image (this intermediate image is unknown), up to two contrast changes f and f^{-1}. The function f is unknown, but satisfies formally the equality $H_u \circ f = H_v \circ f^{-1}$. Thus

$$H_u{}^{-1} \circ H_v = f \circ f.$$

It follows that the general method consists in building an increasing function f such that $f \circ f = H_u{}^{-1} \circ H_v$ and replacing v by $f(v)$ and u by $f^{-1}(u)$. This led Delon [**?**] to call this new histogram midway method, the "square root" equalization. The midway interpolation developed in this chapter uses mainly J. Delon's second definition of the midway cumulative histogram as the harmonic mean of the cumulative histograms of both images. This definition is preferable to the square root. Indeed, both definitions yield very similar results but the harmonic mean extends easily to an arbitrary number of images and in particular to movies [88]. The Cox, Roy and Hingorani algorithm defined in [78] performs a midway equalization. They called their algorithm "Dynamic histogram warping" and its aim is to give a common cumulative histogram (and therefore a common histogram) to a pair of images. Although their method is presented as a dynamic algorithm, there is a very simple underlying formula, which is the harmonic mean of cumulative histograms discovered by Delon [87].

Chapter 7

Contrast-Invariant Monotone Operators

A function operator T is monotone if $u \geq v \Rightarrow Tu \geq Tv$. A set operator $\mathcal{T}$ is monotone if $X \subset Y$ implies $\mathcal{T}X \subset \mathcal{T}Y$. We are mainly interested in monotone function operators, since they are nonlinear generalizations of linear smoothing using a nonnegative convolution kernel. We have already argued that for image analysis to be robust, the operators must also be contrast invariant. The overall theme here will be to develop the equivalence between monotone contrast-invariant function operators and monotone set operators. This equivalence is based on one of the fundamentals of mathematical morphology described in Chapter 5: A real-valued function is completely described by its level sets.

This allows one to process an image u by processing separately its level sets by some monotone set operator $\mathcal{T}$ and defining the processed image by the *superposition principle*

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

Such an operator is called in digital technology a *stack filter*, since it processes an image as a stack of level sets. Conversely, we shall associate with any contrast invariant monotone function operator T a monotone set operator by setting

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

Such a construction is called *a level set extension of* T.

Several questions arise, which will be all answered positively once the functional framework is fixed: Are stack filters contrast invariant? Conversely, is any monotone contrast invariant operator a stack filter? Is any monotone set operator the level set extension of its stack filter?

In Section 7.1 we shall make definitions precise and give some remarkable conservative properties of contrast invariant monotone operators. Section 7.2 is devoted to stack filters and shows that they are monotone and contrast invariant. Section 7.3 defines the level set extension and shows the converse statement: Any contrast invariant monotone operator is a stack filter. Section 7.4 applies this construction to a remarkable denoising stack filter due to Vincent and Serra, the area opening.

7.1 Contrast-invariance

7.1.1 Set monotone operators

We will be mostly dealing with function operators T defined on $\mathcal{F}$ and set operators $\mathcal{T}$ defined on $\mathcal{L}$, but sometimes also defined on $\mathcal{M}$. We denote by $\mathcal{D}(\mathcal{T})$ the domain of $\mathcal{T}$. Now, all set operators we shall consider in practice are defined first on subsets of $\mathbb{R}^N$.

Definition 7.1. *Let $\mathcal{T}$ a monotone operator defined on a set of subsets of $\mathbb{R}^N$. We call* standard extension of $\mathcal{T}$ *to S_N the operator, still denoted by $\mathcal{T}$, defined by*

$$\mathcal{T}(X) = \mathcal{T}(X \setminus \{\infty\}) \cup (X \cap \{\infty\}).$$

In other terms if X doesn't contain ∞, $\mathcal{T}(X)$ is already defined and if X contains ∞, $\mathcal{T}(X)$ contains it too. Thus a standard extension satisfies $\infty \in \mathcal{T}X \Leftrightarrow \infty \in X$.

Remark 7.2. *Let us examine the case where $\mathcal{T}$ is initially defined on $\mathcal{C}$, the set of all closed subsets of $\mathbb{R}^N$. There are only two kinds of sets in $\mathcal{L}$, namely*

- *compact sets of $\mathbb{R}^N$*
- *sets of the form $X = C \cup \{\infty\}$, where C is a closed set of $\mathbb{R}^N$.*

Thus the standard extension of $\mathcal{T}$ extends $\mathcal{T}$ to $\mathcal{L}$, the set of all closed (and therefore compact) subsets of S_N.

All of the usual monotone set operators used in shape analysis satisfy a small list of standard properties which it is best to fix now. Their meaning will come obvious in examples.

Definition 7.3. *We say that a set operator $\mathcal{T}$ defined on its domain $D(\mathcal{T})$ is* standard monotone *if*

- $X \subset Y \implies \mathcal{T}X \subset \mathcal{T}Y$;
- $\infty \in \mathcal{T}X \Longleftrightarrow \infty \in X$;
- $\mathcal{T}(\emptyset) = \emptyset$, $\mathcal{T}(S_N) = S_N$;
- $\mathcal{T}(X)$ *is bounded in* $\mathbb{R}^N$ *if* X *is;*
- $\mathcal{T}(X)^c$ *is bounded in* $\mathbb{R}^N$ *if* X^c *is.*

Definition 7.4. *Let $\mathcal{T}$ be a monotone set operator on its domain $\mathcal{D}(\mathcal{T})$. We call* dual domain *the set*

$$\mathcal{D}(\tilde{\mathcal{T}}) := \{X \subset S_N \mid X^c \in \mathcal{D}(\mathcal{T})\}.$$

We call dual *of $\mathcal{T}$ the operator $X \to \tilde{\mathcal{T}}X = (\mathcal{T}(X^c))^c$, defined on $\mathcal{D}(\tilde{\mathcal{T}})$.*

Proposition 7.5. *$\mathcal{T}$ is a standard monotone operator if and only if $\tilde{\mathcal{T}}$ is.*

Exercise 7.1. Prove it! ■

7.1.2 Monotone function operators

Function operators are usually defined on $\mathcal{F}$, the set of continuous functions having some limit $u(\infty)$ at infinity. We shall always assume that this limit is preserved by T, that is, $Tu(\infty) = u(\infty)$. Think that images are usually compactly supported. Thus $u(\infty)$ is the "color of the frame" for a photograph. There is no use in changing this color.

Definition 7.6. *We say that a function operator $T : \mathcal{F} \to \mathcal{F}$ is standard monotone if for all $u, v \in \mathcal{F}$,*

$$u \geq v \implies Tu \geq Tv; \quad Tu(\infty) = u(\infty). \tag{7.1}$$

Exercise 7.2. Is the operator T defined by $(Tu)(\mathbf{x}) = u(\mathbf{x}) + 1$ standard monotone?
■

Recall from Chapter 5 that any nondecreasing continuous surjection $g : \mathbb{R} \to \mathbb{R}$ is called a contrast change.

Definition 7.7. *A function operator $T \colon \mathcal{F} \to \mathcal{F}$ is said to be contrast invariant if for every $u \in \mathcal{F}$ and every contrast change g,*

$$g(Tu) = Tg(u). \tag{7.2}$$

Checking contrast invariance with increasing contrast changes will make our life simpler.

Lemma 7.8. *A monotone operator is contrast invariant if and only if it commutes with strictly increasing contrast changes.*

Proof. Let g be a contrast change. We can find strictly increasing continuous functions g_n and $h_n : \mathbb{R} \to \mathbb{R}$ such that $g_n(s) \to g(s)$, $h_n(s) \to g(s)$ for all s and $g_n \leq g \leq h_n$ (see Exercise 7.12.) Thus, by using the 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 $T(g(u)) = g(Tu)$. □

Let us give some notable properties entailed by the monotonicity and the contrast invariance.

Lemma 7.9. *Let T be standard monotone contrast invariant operator. Then for every constant function $u \equiv c$ one has $Tu \equiv c$.*

Proof. Let g be a contrast change such that $g(s) = s$ for $\inf Tu \leq s \leq \sup Tu$. Since $Tu(\infty) = u(\infty) = c$, this implies that $\inf Tu \leq c \leq \sup Tu$ and therefore $g(c) = c$, which means $g(u) = u$.. By the contrast invariance we therefore obtain $Tu = Tg(u) = g(Tu) \equiv c$. □

We have indicated several times the importance of image operators being contrast invariant. In practice, image operators are also translation invariant. For $\mathbf{x} \in \mathbb{R}^N$ we are going to use the notation $\tau_{\mathbf{x}}$ to denote the translation operator for both sets and functions: For $X \in \mathcal{M}$, $\tau_{\mathbf{x}} X = \{\mathbf{x} + \mathbf{y} \mid \mathbf{y} \in X\}$, and for $u \in \mathcal{F}$, $\tau_{\mathbf{x}} u$ is defined by $\tau_{\mathbf{x}} u(\mathbf{y}) = u(\mathbf{y} - \mathbf{x})$. Since elements of $\mathcal{M}$ can contain ∞, we specify that $\infty \pm \mathbf{x} = \infty$ when $\mathbf{x} \in \mathbb{R}^N$. This implies that $\tau_{\mathbf{x}} u(\infty) = u(\infty)$.

Definition 7.10. *A set operator $\mathcal{T}$ is said to be translation invariant if its domain is translation invariant and if for all $X \in \mathcal{D}(\mathcal{T})$ and $\mathbf{x} \in \mathbb{R}^N$,*

$$\tau_{\mathbf{x}} \mathcal{T} X = \mathcal{T} \tau_{\mathbf{x}} X.$$

A function operator T is said to be translation invariant if for all $u \in \mathcal{F}$ and $\mathbf{x} \in \mathbb{R}^N$,

$$\tau_{\mathbf{x}} T u = T \tau_{\mathbf{x}} u.$$

We say that a function operator T commutes with the addition of constants if $u \in \mathcal{F}$ and $c \in \mathbb{R}$ imply $T(u + c) = Tu + c$.

Contrast-invariant operators clearly commute with the addition of constants: Consider the contrast change defined by $g(s) = s + c$.

Lemma 7.11. *Let T be a translation-invariant monotone function operator on $\mathcal{F}$ that commutes with the addition of constants. If $u \in \mathcal{F}$ is K-Lipschitz on $\mathbb{R}^N$, namely $|u(\mathbf{x}) - u(\mathbf{y})| \le K|\mathbf{x} - \mathbf{y}|$ for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^N$, then so is Tu.*

Proof. For any $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^N$, and $\mathbf{z} \in S_N$, we have

$$u(\mathbf{y} + \mathbf{z}) - K|\mathbf{x} - \mathbf{y}| \le u(\mathbf{x} + \mathbf{z}) \le u(\mathbf{y} + \mathbf{z}) + K|\mathbf{x} - \mathbf{y}|. \tag{7.3}$$

These inequalities work for $\mathbf{z} = \infty$ because $u(\mathbf{y} + \infty) = u(\mathbf{x} + \infty) = u(\infty)$. Thus we can write them as inequalities between functions on S_N:

$$\tau_{-\mathbf{y}} u - K|\mathbf{x} - \mathbf{y}| \le \tau_{-\mathbf{x}} u \le \tau_{-\mathbf{y}} u + K|\mathbf{x} - \mathbf{y}|. \tag{7.4}$$

Since T is monotone, we can apply T to the functions in (7.4) and preserve the inequalities, which yields

$$T(\tau_{-\mathbf{y}} u - K|\mathbf{x} - \mathbf{y}|) \le T(\tau_{-\mathbf{x}} u) \le T(\tau_{-\mathbf{y}} u + K|\mathbf{x} - \mathbf{y}|).$$

Now use the fact that T commutes with the addition of constants the translation invariance of T to obtain

$$\tau_{-\mathbf{y}}(Tu) - K|\mathbf{x} - \mathbf{y}| \le \tau_{-\mathbf{x}}(Tu)) \le T(\tau_{-\mathbf{y}} u) + K|\mathbf{x} - \mathbf{y}|).$$

Taking the values of these functions at 0 yields

$$Tu(\mathbf{y}) - K|\mathbf{x} - \mathbf{y}| \le Tu(\mathbf{x}) \le Tu(\mathbf{y}) + K|\mathbf{x} - \mathbf{y}|,$$

which is the announced result. □

We say that an operator is monotone on a set of functions if $u \ge v \Rightarrow Tu \ge Tv$. Clearly all above proofs do not depend upon the fact that the operator is standard, but just upon its translation invariance and monotonicity. Thus, by considering the proof of Lemma 7.11 and the definition of uniform continuity (Definition 0.3), one obtains the following generalizations.

Corollary 7.12. *Assume that T is a translation-invariant monotone operator on a set of uniformly continuous functions, that commutes with the addition of constants. Then Tu is uniformly continuous on $\mathbb{R}^N$ with the same modulus of continuity. In particular if u is L-Lipschitz on $\mathbb{R}^N$, then so is Tu.*

Exercise 7.3. Prove corollary 7.12.
■

7.2 Stack filters

Definition 7.13. *We say that a function operator T is obtained from a monotone set operator $\mathcal{T}$ as a* stack filter *if*

$$Tu(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{T}(\mathcal{X}_\lambda u)\} \tag{7.5}$$

for every $\mathbf{x} \in S_N$.

The relation (7.5) has practical implications. It means that Tu can be computed by applying $\mathcal{T}$ separately to each characteristic function of the level sets $\mathcal{X}_\lambda u$. This leads to the following *stack filter* algorithm.

$$\begin{array}{ccccccc} & & \mathcal{X}_\lambda u & \to & \mathcal{T}(\mathcal{X}_\lambda u) & & \\ & \nearrow & & & & \searrow & \\ u & & \vdots & & & & Tu(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{T}(\mathcal{X}_\lambda u)\}. \\ & \searrow & & & & \nearrow & \\ & & \mathcal{X}_\mu u & \to & \mathcal{T}(\mathcal{X}_\mu u) & & \end{array}$$

The image u is decomposed into the stack of level sets. Each level set is processed independently by the monotone operator $\mathcal{T}$. This yields a new stack of sets $\mathcal{T}(\mathcal{X}_\lambda u)$ and Formula (7.5) always defines a function Tu. Now, this construction will be perfect only if

$$\mathcal{X}_\lambda(Tu) = \mathcal{T}(\mathcal{X}_\lambda u). \tag{7.6}$$

Definition 7.14. *When (7.6) holds, we say that T "commutes with thresholds", or that T and $\mathcal{T}$ satisfy the "commutation with threshold" property.*

Of course, this commutation can hold only if $\mathcal{T}$ sends $\mathcal{L}$ into itself. A further condition which turns out to be necessary is introduced in the next definition.

Definition 7.15. *We say that a monotone set operator $\mathcal{T} : \mathcal{L} \to \mathcal{L}$ is* upper semicontinuous *if for every sequence of compact sets $X_n \in \mathcal{D}(\mathcal{T}) = \mathcal{L}$ such that $X_{n+1} \subset X_n$, we have*

$$\mathcal{T}(\bigcap_n X_n) = \bigcap_n \mathcal{T}(X_n). \tag{7.7}$$

Exercise 7.4. Show that a monotone operator $\mathcal{T} : \mathcal{L} \to \mathcal{L}$ is upper semicontinuous if and only if it satisfies, for every family $(X_\lambda)_{\lambda \in \mathbb{R}} \subset \mathcal{L}$ such that $X_\lambda \subset X_\mu$ for $\lambda > \mu$, the relation $\mathcal{T}(\bigcap_\lambda X_\lambda) = \bigcap_\lambda \mathcal{T}(X_\lambda)$. ■

Exercise 7.5. Show that a monotone operator on $\mathcal{L}$ is upper semicontinuous if and only if it satisfies (7.7) for every sequence of compact sets X_n such that $X_{n+1} \subset X_n^\circ$. Hint: Since S_N is the unit sphere in $\mathbb{R}^{N+1}$, one can endow it with the euclidian distance d in $\mathbb{R}^{N+1}$. Given a nondecreasing sequence Y_n in $\mathcal{L}$, set $X_n = \{\mathbf{x}, d(\mathbf{x}, Y_n) \leq \frac{1}{n}\}$. Then apply (7.7) to X_n and check that $\bigcap_n X_n = \bigcap_n Y_n$. ■

Exercise 7.6. Show that a monotone operator $\mathcal{T} : \mathcal{L} \to \mathcal{L}$ is upper semicontinuous if and only if it satisfies, for every family $(X_\lambda)_{\lambda \in \mathbb{R}} \subset \mathcal{L}$ such that $X_\lambda \subset X_\mu^\circ$ for $\lambda > \mu$, the relation $\mathcal{T}(\bigcap_\lambda X_\lambda) = \bigcap_\lambda \mathcal{T}(X_\lambda)$. ■

Theorem 7.16. *Let $\mathcal{T} \colon \mathcal{L} \to \mathcal{M}$ be a translation invariant standard monotone set operator. Then the associated stack filter T is translation invariant, contrast invariant and standard monotone from $\mathcal{F}$ into itself. If, in addition, $\mathcal{T}$ is upper semicontinuous, then T commutes with thresholds.*

Proof that T is monotone. One has

$$u \leq v \Leftrightarrow (\forall \lambda,\ \mathcal{X}_\lambda u \subset \mathcal{X}_\lambda v).$$

Since $\mathcal{T}$ is monotone, we deduce that

$$\forall \lambda,\ \mathcal{T}(\mathcal{X}_\lambda u) \subset \mathcal{T}(\mathcal{X}_\lambda v)$$

which by (7.5) implies $Tu \leq Tv$.

Proof that T is contrast invariant.
By Lemma 7.8 we can take g strictly increasing and therefore a bijection from $\mathbb{R}$ to $\mathbb{R}$. We notice that :
For $\lambda > g(\sup u)$, $\mathcal{X}_\lambda g(u) = \emptyset$ and therefore $\mathcal{T}(\mathcal{X}_\lambda g(u)) = \emptyset$.
For $\lambda < g(\inf u)$, $\mathcal{X}_\lambda g(u) = S_N$ and therefore $\mathcal{T}(\mathcal{X}_\lambda g(u)) = S_N$.
Thus using (7.5) we can restrict the range of λ in the definition of $T(g(u))(\mathbf{x})$:

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

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

$$= \sup\{g(\mu),\ \mathbf{x} \in \mathcal{T}(\mathcal{X}_\mu u)\} = g(Tu(\mathbf{x})).$$

Proof that Tu belongs to $\mathcal{F}$.
T is by construction translation invariant. By Corollary 7.12, Tu is uniformly continuous on $\mathbb{R}^N$. Let us prove that $Tu(\mathbf{x}) \to u(\infty)$ as $\mathbf{x} \to \infty$. We notice that for $\lambda > u(\infty)$, $\mathcal{X}_\lambda u$ is bounded. Since $\mathcal{T}$ is standard monotone $\mathcal{T}(\mathcal{X}_\lambda u)$ is bounded too. Now, by (7.5), $Tu(\mathbf{x}) \leq \lambda$ if $\mathbf{x} \in \mathcal{T}(\mathcal{X}_\lambda u)^c$. This last condition is satisfied if $\mathbf{x}$ is large enough and we deduce that $\limsup_{\mathbf{x} \to \infty} Tu(\mathbf{x}) \leq u(\infty)$. In the same way notice that $(\mathcal{X}_\lambda u)^c$ is bounded if $\lambda < u(\infty)$. So by the same argument, we also get $\liminf_{\mathbf{x} \to \infty} Tu(\mathbf{x}) \geq u(\infty)$. $\mathcal{T}$ being standard, it is easily checked using (7.5) that $Tu(\infty) = u(\infty)$. Thus, Tu is continuous on $\mathbb{R}^N$ and at ∞ and therefore on S_N.

Proof that T commutes with thresholds, when $\mathcal{T}$ is upper semicontinuous.
Let us show that the sets $Y_\lambda = \mathcal{T}(\mathcal{X}_\lambda u)$ satisfy the properties (i) and (ii) in

Proposition 5.2. By the monotonicity of $\mathcal{T}$, $Y_\lambda \subset Y_\mu$ for $\lambda > \mu$. Since $\mathcal{T}(\emptyset) = \emptyset$, we have

$$Y_\lambda = \mathcal{T}(\mathcal{X}_\lambda u) = \mathcal{T}(\emptyset) = \emptyset$$

for $\lambda > \max u$ and, in the same way $Y_\lambda = S_N$ for $\lambda < \min u$. So Tu has the same bounds as u. This proves Property (i). As for Property (ii), we have for every λ, using the upper semicontinuity and exercise 7.4,

$$Y_\lambda = \mathcal{T}(\mathcal{X}_\lambda u) = \mathcal{T}(\bigcap_{\mu<\lambda} \mathcal{X}_\mu u) = \bigcap_{\mu<\lambda} \mathcal{T}(\mathcal{X}_\mu u) = \bigcap_{\mu<\lambda} Y_\mu.$$

So by applying the converse statement of Proposition 5.2, we deduce that

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

□

Exercise 7.7. Check that $Tu(\infty) = u(\infty)$, as claimed in the former proof. ■

The upper semicontinuity of $\mathcal{T}$ is necessary to ensure the commutation with thresholds. See Exercise 7.21. The assumption that $\mathcal{T}$ sends bounded sets of $\mathbb{R}^N$ on bounded sets of $\mathbb{R}^N$ and complementary sets of bounded sets onto complementary sets of bounded sets also is necessary to ensure that Tu is continuous at ∞: see Exercise 7.16.

7.3 The level set extension

Our aim here is just the converse as in the former section. We wish to associate a standard monotone set operator $\mathcal{T}$ from $\mathcal{L}$ to $\mathcal{L}$ with any contrast invariant standard monotone function operator T, in such a way that the whole machinery works, namely both operators satisfy the commutation with threshold property $\mathcal{T}(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(Tu)$ and T is the stack filter of $\mathcal{T}$.

Lemma 7.17. *Let $u \le 0$ and $v \le 0 \in \mathcal{F}$ and assume that $\mathcal{X}_0 u = \mathcal{X}_0 v \,(\neq \emptyset)$. Then there is a contrast change h such that $h(0) = 0$ and $u \ge h(v)$.*

Proof. Define

$$\tilde{h}(r) = \begin{cases} \min\{u(\mathbf{x}) \mid \mathbf{x} \in \mathcal{X}_r v\} & \text{if } \min v \le r \le 0; \\ r & \text{if } r > 0; \\ \min u - \min v + r & \text{if } r \le \min v. \end{cases}$$

Notice that $\tilde{h}(0) = 0$ and that $\tilde{h}$ is nondecreasing. The following relation holds for all $\mathbf{x} \in \mathbb{R}^N$ by the definition of $\tilde{h}$ and because $u(\mathbf{x})$ belongs to the set $\{u(\mathbf{y}) \mid v(\mathbf{y}) \ge v(\mathbf{x})\}$:

$$u(\mathbf{x}) \ge \min\{u(\mathbf{y}) \mid v(\mathbf{y}) \ge v(\mathbf{x})\} = \tilde{h}(v(\mathbf{x})).$$

We now use the compactness in S_N of the level sets of v to show that $\tilde{h}$ is continuous at zero. Let $(r_k)_{k\in\mathbb{N}}$ be an arbitrary increasing sequence tending to zero. Choose $\mathbf{x}_k \in \mathcal{X}_{r_k} v$ such that $\tilde{h}(r_k) = u(\mathbf{x}_k)$. This is possible because u is continuous and the $\mathcal{X}_{r_k} v$ are compact and nonempty. Since $\tilde{h}$ is nondecreasing, $\tilde{h}(r_k) \to \tilde{h}^-(0)$.

Let $\mathbf{x}$ be any accumulation point of the set $\{\mathbf{x}_k\}_{k\in\mathbb{N}}$. Since the $\mathcal{X}_{r_k} v$ are compact, all the accumulation points of the set $\{x_k\}_{k\in\mathbb{N}}$ are contained in $\mathcal{X}_0 v = \bigcap_{k\in\mathbb{N}} \mathcal{X}_{r_k} v$. This means that $u(\mathbf{x}) = 0$. But $\lim u(\mathbf{x}_k) = u(\mathbf{x})$ by the continuity of u, and we conclude that $\tilde{h}^-(0) = 0$. At this point $\tilde{h}$ satisfies the announced requirements for h, except that it is not always continuous for all $r < 0$. This is easily fixed by choosing a continuous nondecreasing function h such that $\tilde{h} \geq h$ and $h(0) = 0$. One way to do this is to take $h(r) = (1/|r|) \int_{2r}^{r} \tilde{h}(s)\,\mathrm{d}s$ for $r < 0$. Then $u(\mathbf{x}) \geq \tilde{h}(v(\mathbf{x})) \geq h(v(\mathbf{x}))$ as announced. □

Exercise 7.8. Prove that $h(r) = (1/|r|) \int_{2r}^{r} \tilde{h}(s)\,\mathrm{d}s$ is indeed nondecreasing and continuous for $r \leq 0$ and that $\tilde{h} \geq h$. Find examples of functions u and v defined on S_1 for which $\tilde{h}$ is not continuous. ■

Definition 7.18 (and proposition (Evans-Spruck)). [1] *Given a contrast invariant monotone operator T on $\mathcal{F}$, we call level set extension of T the set operator defined in the following way : for any $X \in \mathcal{L}$, take $u \leq 0$ such that $\mathcal{X}_0 u = X$ and set*

$$\mathcal{T}(X) = \mathcal{X}_0 T(u).$$

Then $\mathcal{T}(X)$ does not depend upon the particular choice of u.

Proof. The proof follows directly from Lemma 7.17: Take u and $v \in \mathcal{F}$ such that $u \leq 0$, $v \leq 0$, and $\mathcal{X}_0 u = \mathcal{X}_0 v$. Let h be a contrast change such that $h(0) = 0$ and $u \geq h(v)$. Since T is monotone and contrast invariant one has by Lemma 7.1.2 $Tu \leq 0$, and $Tu \geq Th(v) = h(Tv)$. Using the fact that $h(0) = 0$, we obtain that $Tv(\mathbf{x}) = 0$ implies that $Tu(\mathbf{x}) = 0$. By interchanging the roles of u and v, $Tu(\mathbf{x}) = 0$ implies that $Tv(\mathbf{x}) = 0$. We conclude that $\mathcal{X}_0 Tu = \mathcal{X}_0 Tv$. □

Exercise 7.9. Definition 7.18 would'nt be complete if we did not prove that for any $X \in \mathcal{L}$ we can find $u \leq 0$ in $\mathcal{F}$ such that $\mathcal{X}_0 u = X$. Hint: Since S_N is the unit sphere in $\mathbb{R}^{N+1}$, one can endow it with the euclidian distance d in $\mathbb{R}^{N+1}$. Use the distance function $d(\mathbf{x}, X)$ to define u. This distance function is continuous: see Exercise 7.18. ■

Theorem 7.19 (Evans–Spruck). *Let T be a contrast-invariant monotone operator on $\mathcal{F}$ and $\mathcal{T}$ its level set extension on $\mathcal{L}$. Then $\mathcal{T}$ is monotone, T and $\mathcal{T}$ satisfy the commutation with thresholds $\mathcal{T}\mathcal{X}_\lambda u = \mathcal{X}_\lambda Tu$ for all $\lambda \in \mathbb{R}$, T is the stack filter associated with $\mathcal{T}$ and $\mathcal{T}$ is upper semicontinuous on $\mathcal{L}$. In addition, if T is standard, then so is $\mathcal{T}$.*

Proof. **Commutation with thresholds:** Given u and λ, let g be a continuous contrast change such that $g(s) = \min(s, \lambda) - \lambda$ on the range of u, which is a compact interval of $\mathbb{R}$. We then have $\mathcal{X}_0 g(u) = \mathcal{X}_\lambda u$. Using this relation, the level set extension and the contrast invariance of T,

$$\mathcal{T}(\mathcal{X}_\lambda u) = \mathcal{T}(\mathcal{X}_0 g(u)) = \mathcal{X}_0(T(g(u))) = \mathcal{X}_0(g(Tu)) = \mathcal{X}_\lambda(Tu).$$

[1] What we are doing here is related to the scheme originally introduced by Osher and Sethian as a numerical method for front propagation [224]. We briefly described their work in the Introduction (see page 26).

Proof of the stack filter property: This is an immediate consequence of the superposition principle and the commutation with thresholds :

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

Proof that $\mathcal{T}$ is upper semicontinuous on $\mathcal{L}$: By the result of Exercise 7.5, it is enough to consider a sequence $(X_n)_{n\geq 1}$ in $\mathcal{L}$ such that $X_{n+1} \subset X_n^\circ$. By Lemma 7.20 below there is a function $u \in \mathcal{F}$ such that $\mathcal{X}_{1-\frac{1}{n}}u = X_n$ and $\mathcal{X}_1 u = \bigcap_n X_n$. Finally, using twice the just proven commutation of thresholds,

$$\mathcal{T}(\bigcap_n X_n) = \mathcal{T}(\mathcal{X}_1 u) = \mathcal{X}_1(Tu) = \bigcap_n \mathcal{X}_{1-\frac{1}{n}}Tu = \bigcap_n \mathcal{T}(\mathcal{X}_{1-\frac{1}{n}}u) = \bigcap_n \mathcal{T}(X_n).$$

Proof that $\mathcal{T}$ is standard if T is: Recall that T is standard if $Tu(\infty) = u(\infty)$. By using the commutation with thresholds, all of the standard properties for $\mathcal{T}$ are straightforward. For instance, taking some $u \in \mathcal{F}$,

$$\mathcal{T}(\emptyset) = \mathcal{T}(\mathcal{X}_{\max u+1}u) = \mathcal{X}_{\max u+1}Tu = \emptyset.$$

Indeed, by the monotonicity, the contrast invariance, and Lemma 7.1.2, $u \leq C \Rightarrow Tu \leq C$.
In the same way, let $X \in \mathcal{L}$ and u a function such that $\mathcal{X}_0 u = X$. If X is bounded, then $u(\infty) < 0$, so that $Tu(\infty) = u(\infty) < 0$. Thus $\mathcal{T}(X) = \mathcal{X}_0 Tu$ is bounded. If $X^c = \{\mathbf{x} \mid u(\mathbf{x}) < 0\}$ is bounded, then $Tu(\infty) = u(\infty) \geq 0$. Thus $\mathcal{T}(X)^c = (\mathcal{X}_0 Tu)^c$ is bounded. Finally by the commutation with thresholds,

$$\infty \in X \Leftrightarrow u(\infty) \geq 0 \Leftrightarrow Tu(\infty) \geq 0 \Leftrightarrow \infty \in \mathcal{X}_0(Tu) = \mathcal{T}(X).$$

□

Exercise 7.10. Prove that the level set extension $\mathcal{T}$ is monotone. The argument is not given in the above proof. ■

Lemma 7.20. *Let $(X_n)_{n\geq 1}$ be a sequence in $\mathcal{L}$ such that $X_{n+1} \subset X_n^\circ$. There is a function $u \in \mathcal{F}$ such that $\mathcal{X}_{1-\frac{1}{n}}u = X_n$ for $n \geq 1$ and $\mathcal{X}_1 u = \bigcap_{n\geq 1} X_n$.*

Proof. Let us use the euclidian distance d of $\mathbb{R}^{N+1}$ restricted to S_N considered as a subset of $\mathbb{R}^{N+1}$. Set $u(\mathbf{x}) = 1$ if $\mathbf{x} \in \bigcap_n X_n$,

$$u(\mathbf{x}) = (1 - \frac{1}{n})\frac{d(\mathbf{x}, X_{n+1})}{d(\mathbf{x}, X_n^c) + d(\mathbf{x}, X_{n+1})} + (1 - \frac{1}{n+1})\frac{d(\mathbf{x}, X_n^c)}{d(\mathbf{x}, X_n^c) + d(\mathbf{x}, X_{n+1})}$$

for $\mathbf{x} \in X_n \setminus X_{n+1}$ and $n \geq 1$, $u(\mathbf{x}) = -\sup(-1, -d(\mathbf{x}, X_1))$ if $\mathbf{x} \notin X_1$. It is easily checked that u belongs in $\mathcal{F}$ and satisfies the announced properties. □

7.4 A first application: the extrema killer

This section is devoted to the study of operators that remove "peaks," or extreme values, from an image. Such peaks are often created by impulse noise, that is, local destruction of pixel values and their replacement by a random value. Old movies present this kind of noise and it also occurs by transmission failure in satellite imaging. The operators we study are called *area opening*, or *extrema killer* operators, and they have been shown to be very effective at removing this kind of noise. The action of these operators is illustrated in Figures 7.1 and 7.2.

The following definitions are standard, but we include them here for completeness.

Definition 7.21. *Consider a closed subset X of S_N. X is disconnected if it can be written as $X = (A \cap X) \cup (B \cap X)$, where A and B are disjoint open sets and both $A \cap X$ and $B \cap X$ are not empty. X is connected if it is not disconnected. The connected component of $\mathbf{x}$ in X, denoted by $cc(\mathbf{x}, X)$, is the maximal connected subset of X that contains $\mathbf{x}$.*

We wish to define a denoising operator on $\mathcal{L}$; since some sets therein contain ∞, we need an extension of the Lebesgue measure on $\mathbb{R}^N$ to S_N. This is immediately fixed by setting $\text{meas}(\{\infty\}) = +\infty$. The only property of this extended measure that we need to check is following:

Lemma 7.22. *if Y_n is a nonincreasing sequence of compact sets of S_N, then $meas(\cap_n Y_n) = \lim_n meas(Y_n)$.*

Proof. If the compact sets Y_n do not contain ∞ for n large enough, then they are bounded in $\mathbb{R}^N$ for n large and the result just follows from Lebesgue theorem. If instead the sets Y_n all contain ∞, then $\cap_n Y_n$ contains it too and all sets have infinite measure. $\square$

Definition 7.23. *Let $a > 0$ a scale parameter and denote for every $X \in \mathcal{L}$ by X_i its connected components, so that $X = \bigcup_i X_i$. We call small component killer the operator on $\mathcal{L}$ which removes from X all connected components with area stricly less than a :*

$$\mathcal{T}_a X = \bigcup_{meas(X_i) \geq a} X_i. \tag{7.8}$$

Theoretically, X can have an uncountable number of components; take, for example, the Cantor set. However, X can have only a countable number of components with positive measure. The assumption $\text{meas}(\{\infty\}) = +\infty$ implies that all connected components of X containing ∞ stay in $\mathcal{T}_a X$. We are going to prove that the small component killer is upper semicontinuous and this uses some elementary topological lemmas.

Lemma 7.24. *Consider an arbitrary nonincreasing sequence of nonempty compact sets $(Y_n)_{n \in \mathbb{N}}$ of S_N and its limit $Y = \bigcap_{n \in \mathbb{N}} Y_n$. Then Y is not empty and compact. In addition, for any open set Z that contains Y, there is an index n_0 such that $Y_n \subset Z$ for all $n \geq n_0$.*

Proof. The first property is a classical property of compact sets. Assume by contradiction that the second property is not true. Then $Y_n \cap (S_N \setminus Z) \neq \emptyset$ infinitely often. This implies that $(Y_n \cap (S_N \setminus Z))_{n \in \mathbb{N}}$ is a nonincreasing sequence of nonempty compact sets. But this means that $Y \cap (S_N \setminus Z) \neq \emptyset$, which is a contradiction. $\square$

Lemma 7.25. *Let $(Y_n)_{n \in \mathbb{N}}$ be a nonincreasing sequence of nonempty compact subsets of S_N and consider the intersection $Y = \bigcap_{n \in \mathbb{N}} Y_n$. If the Y_n are connected, then Y is connected.*

Proof. We know that Y is not empty and compact. Suppose, by contradiction, that Y is not connected. Then we can represent Y by $Y = (Y \cap Z_1) \cup (Y \cap Z_2)$, where Z_1 and Z_2 are disjoint open sets, $Y \cap Z_1 \neq \emptyset$, and $Y \cap Z_1 \neq \emptyset$. Since $Y \subset Z_1 \cup Z_2$, by Lemma 7.24 there exists an n_0 such that $Y_n \subset Z_1 \cup Z_2$ for all $n \geq n_0$, and for these n we have

$$Y_n = Y_n \cap (Z_1 \cup Z_2) = (Y_n \cap Z_1) \cup (Y_n \cap Z_2).$$

Furthermore, $Y_n \cap Z_1 \neq \emptyset$ and $Y_n \cap Z_1 \neq \emptyset$. This contradicts the fact that the Y_n are connected. $\square$

Exercise 7.11. Show that $\mathcal{T}_a$ is idempotent: ${\mathcal{T}_a}^2 X = \mathcal{T}_a X$ and that it is a contraction mapping: $\mathcal{T}_a X \subset X$. ■

With the extrema killer we have a prime example of a theory that begins with a set operator $\mathcal{T}_a$ defined on $\mathcal{L}$.

Lemma 7.26. *The small component killer $\mathcal{T}_a$ is upper semicontinuous on $\mathcal{L}$.*

Proof. We first prove that $\mathcal{T}_a$ is monotone. Assume $X \subset Y$. Then for every $\mathbf{x} \in X$, $cc(\mathbf{x}, X) \subset cc(\mathbf{x}, Y)$. If $\text{meas}\,(cc(\mathbf{x}, X)) \geq a$, then $\text{meas}\,(cc(\mathbf{x}, Y)) \geq a$, and we conclude that $\mathcal{T}_a X \subset \mathcal{T}_a Y$. Now let $(X_n)_n$ be any nonincreasing sequence of nonempty compact sets and $X = \cap_n X_n$. We wish to show that $\mathcal{T}_a X = \bigcap_n \mathcal{T}_a X_n$. By monotonicity of $\mathcal{T}_a$,

$$\mathcal{T}_a X \subset \bigcap_n \mathcal{T}_a(X_n).$$

Let us show the converse inclusion. Let $\mathbf{x} \in \cap_n \mathcal{T}_a(X_n)$. Then $Y_n := cc(x, X_n)$ has measure larger than a for all n. In addition if $m < n$ then $Y_n \subset Y_m$. By Lemmas 7.24 and 7.25, $Y := \cap_n Y_n$ is a connected compact set that contains $\mathbf{x}$. In addition by Lemma 7.22, $\text{measure}(Y) = \lim_n \text{measure}(Y_n) \geq a$. Since $Y = \cap_n Y_n \subset \cap_n X_n = X$, we have $cc(\mathbf{x}, X) \supseteq Y$ and therefore $\mathbf{x} \in \mathcal{T}_a(X)$. $\square$

We can now build a stack filter from $\mathcal{T}_a$.

Definition 7.27 (and proposition). *The stack filter T_a of $\mathcal{T}_a$ is called a maxima killer. $\mathcal{T}_a$ and T_a satisfy the commutation with thresholds. As a consequence, no connected component of a level set of $T_a u$ has measure less than a. Furthermore, T_a is standard monotone, translation and contrast invariant from $\mathcal{F}$ into $\mathcal{F}$.*

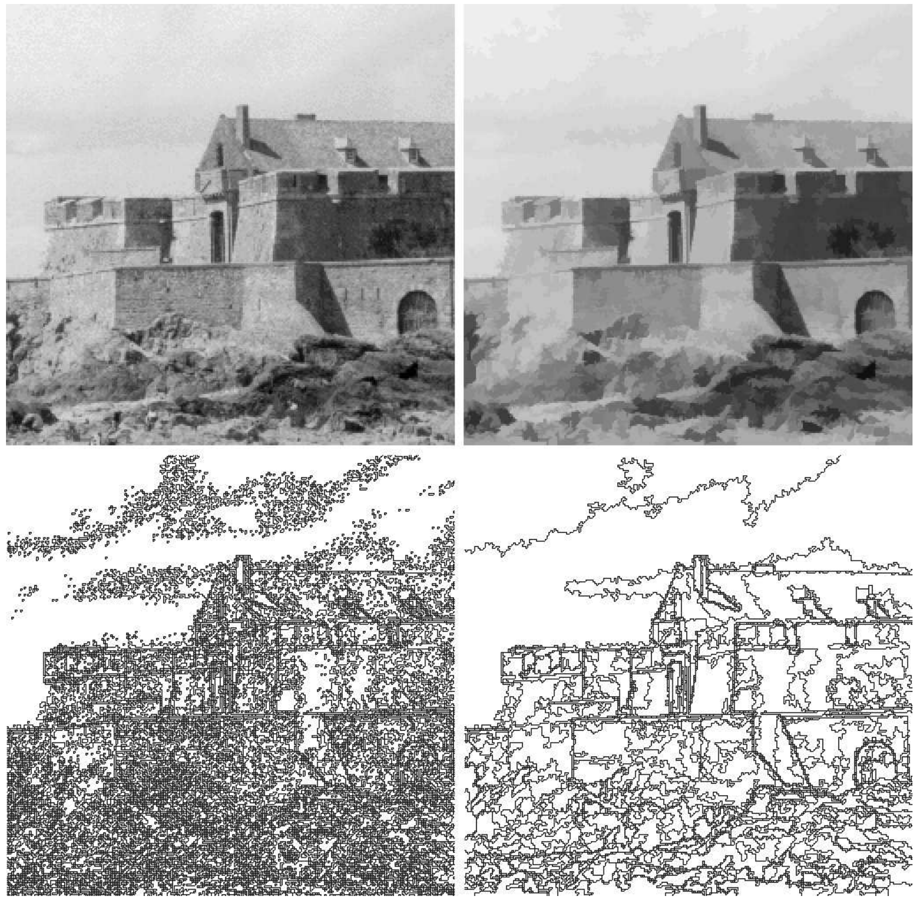

Figure 7.1: Extrema killer: maxima killer followed by minima killer. The extrema killer removes all connected components of upper and lower level sets with area less than some threshold, which here equals 20 pixels. Notice how texture disappears in the second image. All other features seem preserved. On the second row, we see for both the original and the processed image the level lines at 16 equally spaced levels. The level lines on the right hand side are a subset of the level lines of the left hand. All level lines surrounding extremal regions with area smaller than 20 have been removed and the other ones are untouched.

Proof. We just have to check that all assumptions of Theorem 7.16 are satisfied. $\mathcal{T}_a$ is obviously translation invariant, monotone and is upper semicontinuous by Lemma 7.26. It satisfies $\mathcal{T}_a(\emptyset) = \emptyset$, $\mathcal{T}_a(S_N) = S_N$. $\mathcal{T}_a(E)$ is compact if E is. Indeed, it is the union of a finite set of compact connected components. If E is bounded in $\mathbb{R}^N$, then so is $\mathcal{T}_aE \subset E$. $(\mathcal{T}_aE)^c$ is bounded in S_N if E^c is. Indeed, if E^c is bounded, then E has a connected component Y containing $B(0,R)^c$ for some $R > 0$. This connected component has infinite measure. Then $\mathcal{T}_a(E)$ still contains Y and $\mathcal{T}_a(E)^c$ is contained in $B(0,R)$. By construction, ∞ belongs to $\mathcal{T}_aX$ if and only if it belongs to X. Thus, $\mathcal{T}_a$ is standard monotone. $\square$

A maxima killer T_a cuts off the maxima of continuous functions, but it does nothing for the minima. We can immediately define a minima killer T_a^- as the dual operator of T_a,

$$T_a^- u = -T_a(-u).$$

A good denoising process is to alternate T_a and T_a^-, as illustrated in Figures 7.1 and 7.2 . We note, however, that T_a and T_a^- do not necessarily commute, as is shown in Exercise 7.17.

7.5 Exercises

Exercise 7.12. Let $g : \mathbb{R} \to \mathbb{R}$ be a contrast change. Construct increasing contrast changes g_n and h_n such that $g_n(s) \to g(s)$, $h_n(s) \to g(s)$ for all s and $g_n \le g \le h_n$. Hint : define first an increasing continuous function $f(s)$ on $\mathbb{R}$ such that $f(-\infty) = 0$ and $f(+\infty) = \frac{1}{n}$. ■

Exercise 7.13. Let $u : \mathbb{R}^N \to \mathbb{R}$. Show that $\tau_{\mathbf{x}} \mathcal{X}_\lambda u = \mathcal{X}_\lambda \tau_{\mathbf{x}} u$, $\mathbf{x} \in \mathbb{R}^N$. ■

Exercise 7.14. Prove that a translation invariant operator $\mathcal{T}$ from $\mathcal{L}$ to $\mathcal{L}$ satisfies one of the three possibilities : $\mathcal{T}(\{\infty\}) = \{\infty\}$, $\mathcal{T}(\{\infty\}) = S_N$ or $\mathcal{T}(\{\infty\}) = \emptyset$. ■

Exercise 7.15. Let T be a translation invariant standard monotone operator on $\mathcal{F}$. Prove the following statements:

(i) $Tu = c$ for every constant function $u : S_N \to c$.

(ii) $u \ge c$ implies $Tu \ge c$, and $u \le c$ implies $Tu \le c$.

(iii) If in addition T commutes with the addition of constants, $\sup_{\mathbf{x}\in\mathbb{R}^N} |Tu(\mathbf{x}) - Tv(\mathbf{x})| \le \sup_{\mathbf{x}\in\mathbb{R}^N} |u(\mathbf{x}) - v(\mathbf{x})|$.
(Hint: Write $-\sup|u(\mathbf{x}) - v(\mathbf{x})| \le u(\mathbf{x}) - v(\mathbf{x}) \le \sup|u(\mathbf{x}) - v(\mathbf{x})|$.)

■

Exercise 7.16.

1) In dimension 1, consider the set operator defined on $\mathcal{L}$ by $\mathcal{T}X = [\inf X, \infty]$ if $\inf(X \cap \mathbb{R}) \in \mathbb{R}$, $\mathcal{T}X = S_1$ if $\inf(X \cap \mathbb{R}) = -\infty$, $\mathcal{T}(\{\infty\}) = \{\infty\}$, $\mathcal{T}(\emptyset) = \emptyset$. Check that $\mathcal{T}$ satisfies all assumptions of Theorem 7.16 except one. Compute the stack filter associated with $\mathcal{T}$ and show that it satisfies all conclusions of the mentioned theorem except one : Tu does not belong to $\mathcal{F}$ and more specifically $Tu(\mathbf{x})$ is not continuous at ∞.

2) Consider the function operator on $\mathcal{F}$, $Tu(\mathbf{x}) = \sup_{\mathbf{x}\in S_N} u(\mathbf{x})$. Check that T is monotone, contrast invariant, and sends $\mathcal{F}$ to $\mathcal{F}$. Compute the level set extension $\mathcal{T}$ of T. ■

Exercise 7.17. Let $N = 1$ and take $u(x) = \sin x$ for $|x| \le 8\pi$, $u(x) = 0$ otherwise. Compute $T_a u$ and $T_a^- u$ and show that they commute on u if $a \le \pi$ and do not commute if $a > \pi$. Following the same idea, construct a function $u \in \mathcal{F}$ in dimension two such that $T_a T_a^- u \ne T_a^- T_a u$. ■

Exercise 7.18. Let X be a closed subset of a metric space endowed with a distance d and consider the distance function to X,

$$d(\mathbf{y}) = d(\mathbf{y}, X) = \inf_{\mathbf{x}\in X} d(\mathbf{x}, \mathbf{y}).$$

Show that d is 1-Lipschitz, that is, $|d(\mathbf{x}, X) - d(\mathbf{y}, X)| \le d(\mathbf{x}, \mathbf{y})$. ■

Exercise 7.19. In the following questions, we explain the necessity of the assumptions $\mathcal{T}(\emptyset) = \emptyset$, $\mathcal{T}(S_N) = S_N$ for defining function monotone operators from $\mathcal{F}$ to $\mathcal{F}$.

1) Set $\mathcal{T}(X) = X_0$ for all $X \in \mathcal{L}$, where $X_0 \neq \emptyset$ is a fixed set. Check that the associated stack filter satisfies $Tu(\mathbf{x}) = +\infty$ if $\mathbf{x} \in X_0$, $Tu(\mathbf{x}) = -\infty$ otherwise.

2) Let $\mathcal{T}$ be a monotone set operator, without further assumption. Show that its associated stack filter T is, however, monotone and commutes with all contrast changes. (We extend each contrast change g by setting $g(\pm\infty) = \pm\infty$.) ■

Exercise 7.20. Take an operator $\mathcal{T}$ satisfying the same assumptions as in Theorem 7.16, but defined on $\mathcal{M}$ and apply the arguments of the proof of Theorem 7.16. Check that the stack filter associated with $\mathcal{T}$ is a contrast invariant, translation invariant monotone operator on the set of all bounded measurable functions, $L^\infty(\mathbb{R}^N)$. If in addition $\mathcal{T}$ is upper semicontinuous on $\mathcal{M}$, then the commutation with thresholds holds. ■

Exercise 7.21. The upper semicontinuity is necessary to ensure that a monotone set operator defines a function operator such that the commutation with thresholds $\mathcal{X}_\lambda(Tu) = \mathcal{T}(\mathcal{X}_\lambda(u))$ holds for every λ. Let us choose for example the following set operator $\mathcal{T}$,

$$\mathcal{T}(X) = X \text{ if } \mathrm{meas}(X) > a \text{ and } \mathcal{T}(X) = \emptyset \text{ otherwise }.$$

(We use the Lebesgue measure on $\mathbb{R}^N$, with the completion $\mathrm{meas}(\{\infty\}) = +\infty$)

1) Prove that $\mathcal{T}$ is standard monotone.

2) Let u be the function from S_1 into S_1 defined by $u(x) = max(-|x|, -2a)$ for some $a > 0$, with $u(\infty) = -2a$. Check that u belongs to $\mathcal{F}$. Then, applying the stack filter T of $\mathcal{T}$, check that

$$T(u)(x) = \sup\{\lambda, x \in \mathcal{T}(\mathcal{X}_\lambda u)\} = max(\min(-|x|, -a/2), -2a).$$

3) Deduce that $\mathcal{X}_{-a/2}T(u) = [-a/2, a/2]$, $\mathcal{X}_{-a/2}u = [-a/2, a/2]$ and therefore

$$\mathcal{T}(\mathcal{X}_{-a/2}u) = \emptyset \neq \mathcal{X}_{-a/2}T(u),$$

which means that T does not commute with thresholds. ■

Exercise 7.22. Like in the preceding exercise, we consider here contrast invariant operators defined on all measurable bounded functions of $\mathbb{R}^N$. The aim of the exercise is to show that such operators send images with finite range into images with finite range. More precisely, denote by $R(u) = u(\mathbb{R}^N)$ the range of u. Then we shall prove that for every u, $R(Tu) \subset \overline{Ru}$. In particular, if $R(u)$ is finite, then the range of Tu is a finite subset of Ru. If u is binary, Tu is, etc. This shows that contrast invariant operators preserve sharp contrasts. A binary image is transformed into a binary image. So contrast invariant operators create no blur, as opposed to linear operators, which always create new intermediate grey levels.

1) Consider

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

where $d(s, X)$ denotes the distance from s to X, that is, $d(s, X) = \inf_{x \in X} |s - x|$. Show that g is a contrast change satisfying $g(s) = s$ for $s \in \overline{Ru}$ and $g(s) > s$ otherwise.

2) Check that $g(s) = s$ if and only if $s \in \overline{Ru}$. In particular, $g(u) = u$. Deduce from this and from the contrast invariance of T that for every $\mathbf{x} \in \mathbb{R}^N$, $Tu(\mathbf{x})$ is a fixed point of g. Conclude. ■

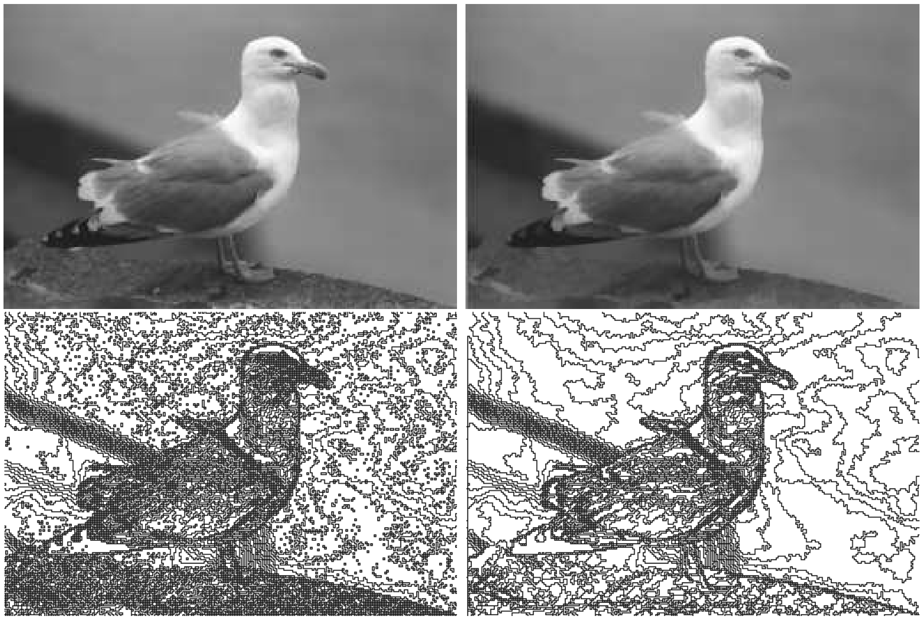

Figure 7.2: Extrema killer: maxima killer followed by minima killer. Above, left: original image. Above, right: image after extrema killer removed connected components of 20 pixels or less. Below: level lines (levels of multiples of 16) of the image before and after the application of the extrema killer.

7.6 Comments and references

Contrast invariance and stack filters. Image operators that commute with thresholds have been popular because, among other reasons, they are easily implemented in hardware (VLSI). This led to very simple patents being awarded in signal and image processing as late as 1987 [79]. These operators have been given four different names, although operators are equivalent: *stack filters* [50, 137, 286]; *threshold decomposition* [141]; *rank filters* [69, 166, 288]; and *order filters* [265]. The best known of these are the sup, inf, and median operators. The implementation of the last named has received much attention because of its remarkable denoising properties [109, 218, 290].

Maragos and Shafer [195, 196] and Maragos and Ziff [197] introduced the functional notation and established the link between stack filters and the Matheron formalism in "flat" mathematical morphology. The complete equivalence between contrast-invariant operators and stack filters, as developed in this chapter, does not seem to have appeared elsewhere; at least we do not know of other references. A related classification of rank filters with elegant and useful generalizations to the so-called *neighborhood filters* can be found in [166].

The extrema killer. The extrema killer is probably the most efficient denoising filter for images degraded by impulse noise, which is manifest by small spots. In spite of its simplicity, this filter has only recently seen much use. This is undoubtedly due to the nontrivial computations involved in searching for the connected components of upper and lower level sets. The first reference to the extrema killer that we know is [72]. The filter in its generality was defined by Vincent in [276]. This definition fits into the general theory of connected filters developed by Salembier and Serra [246]. Masnou defined a variant called the *grain filter* that is both contrast invariant and invariant under reverse contrast changes [201]. Monasse and Guichard developed a fast implementation of this filter based on the so-called *fast level set transform* [207].

We will develop in Chapter 19 a theory of scale space that is based on a family of image smoothing operators T_t, where t is a scale parameter. We note here that the family $(T_a)_{a\in\mathbb{R}^+}$ of extrema killers does not constitute a scale space because it does not satisfy one of the conditions, namely, what we call the local comparison principle. That this is so, is the content of Exercise 19.1.

Chapter 8

Sup-Inf Operators

The main contents of this chapter are two representation theorems: one for translation-invariant monotone set operators and one for functions operators that are monotone, contrast invariant, and translation invariant. If T is a function operator satisfying these three conditions, then it has a "sup-inf" representation of the form

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

where $\mathcal{B}$ is a family of subsets of $\mathcal{M}(S_N)$, the set of all measurable subsets of S_N. This theorem is a nonlinear version of the Riesz theorem that states that a continuous linear translation-invariant operator from $L^2(\mathbb{R}^N)$ to $C^0(\mathbb{R}^N)$ can be represented as a convolution

$$Tu(\mathbf{x}) = \int_{\mathbb{R}^N} u(\mathbf{x}-\mathbf{y})k(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

In this case, the kernel $k \in L^2(\mathbb{R}^N)$ is called the impulse response. In the same way, $\mathcal{B}$ is an impulse response for the nonlinear operator.

8.1 Translation-invariant monotone set operators

Recall that a set of $\mathcal{M}$ can contain ∞. We have specified that $\mathbf{x}+\infty = \infty$ for every $\mathbf{x} \in S_N$. As a consequence, for any subset B of S_N, $\infty + B = \{\infty\}$.

Definition 8.1. *We say that a subset $\mathcal{B}$ of $\mathcal{M}$ is standard if it is not empty and satisfies*

(i) $\forall R > 0,\ \exists R' > 0,\ (\mathbf{x}+B \subset B(0,R)$ *and* $B \in \mathcal{B}) \Rightarrow \mathbf{x} \in B(0,R')$;

(ii) $\forall R > 0,\ \exists R' > 0,\ \mathbf{x} \in B(0,R')^c \Rightarrow (\exists B \in \mathcal{B},\ \mathbf{x}+B \subset B(0,R)^c)$.

Exercise 8.1. Conditions (i) and (ii) look a bit sophisticated, but are easily satisfied. Check that Condition (i) is equivalent to

$$\forall R > 0,\ \exists C > 0,\ (B \in \mathcal{B},\ \text{and diameter}(B) \le R) \Rightarrow B \subset B(0,C).$$

Check that this condition is achieved (e.g.) if all elements of $\mathcal{B}$ contain 0. Check that Condition (ii) is achieved if $\mathcal{B}$ contains at least one bounded element B. ∎

Exercise 8.2. Show that if $\mathcal{B}$ contains $\emptyset$, then $\mathcal{B}$ is not standard. ■

Theorem 8.2 (Matheron). *Let $\mathcal{T}$ be a translation-invariant and standard monotone set operator. Consider the subset of $\mathcal{D}(\mathcal{T})$, $\mathcal{B} = \{B \in \mathcal{D}(\mathcal{T}) \mid 0 \in \mathcal{T}B\}$. Then $\mathcal{B}$ is standard and*

$$\mathcal{T}X = \{\mathbf{x} \in S_N \mid \mathbf{x} + B \subset X \textit{ for some } B \in \mathcal{B}\}. \tag{8.1}$$

Conversely, if $\mathcal{B}$ is any standard subset of $\mathcal{M}$, then formula (8.1) *defines a translation-invariant standard monotone set operator on $\mathcal{M}$.*

Definition 8.3. *In Mathematical Morphology, a set $\mathcal{B}$ such that* (8.1) *holds is called a set of* structuring elements of $\mathcal{T}$ *and $\mathcal{B} = \{X \in \mathcal{D}(\mathcal{T}) \mid 0 \in \mathcal{T}X\}$ is called* the canonical set of structuring elements of $\mathcal{T}$.

Proof of Theorem 8.2.

Proof of (8.1).
Let $\mathcal{B} = \{X \in \mathcal{D}(\mathcal{T}) \mid 0 \in \mathcal{T}X\}$. Then for any $\mathbf{x} \in \mathbb{R}^N$,

$$\begin{aligned}\mathbf{x} \in \mathcal{T}X &\overset{(1)}{\Longleftrightarrow} 0 \in \mathcal{T}X - \mathbf{x} \overset{(2)}{\Longleftrightarrow} 0 \in \mathcal{T}(X - \mathbf{x}) \overset{(3)}{\Longleftrightarrow} X - \mathbf{x} \in \mathcal{B} \\ &\overset{(4)}{\Longleftrightarrow} X - \mathbf{x} = B \text{ for some } B \in \mathcal{B} \overset{(5)}{\Longleftrightarrow} \mathbf{x} + B \subset X \text{ for some } B \in \mathcal{B}.\end{aligned}$$

The equivalence (2) follows from the translation invariance of $\mathcal{T}X$; (3) is just the definition of $\mathcal{B}$; and (4) is a restatement of (3). The implication from left to right in (5) is obvious. The implication from right to left in (5) is the point where the monotonicity of $\mathcal{T}$ is used: Since $B \subset X - \mathbf{x}$, it follows from the monotonicity of $\mathcal{T}$ that $X - \mathbf{x} \in \mathcal{B}$.
Let now $\mathbf{x} = \infty$. Since $\mathcal{T}$ is standard, $\mathcal{B}$ is not empty (it contains S_N) and we have

$$\infty \in \mathcal{T}X \Leftrightarrow \infty \in X \Leftrightarrow \exists B \in \mathcal{B},\ \infty + B \subset X,$$

because $\infty + S_N = \{\infty\}$.

Proof that $\mathcal{B}$ is standard if $\mathcal{T}$ is standard monotone.
Since $\mathcal{T}(S_N) = S_N$, $\mathcal{B}$ contains S_N and is therefore not empty. $\mathcal{T}$ sends bounded sets on bounded sets if and only if there is for every $R > 0$ some $R' > 0$ such that $\mathcal{T}(B(0,R)) \subset B(0,R')$. Using (8.1), this last relation is equivalent to $\{\mathbf{x} \mid \mathbf{x} + B \subset B(0,R)\} \subset B(0,R')$ which is (i). In the same way, $\mathcal{T}$ sends complementary sets of bounded sets on complementary sets of bounded sets if and only if (ii) holds.

Proof that (8.1) **defines a standard monotone set operator if $\mathcal{B}$ is standard.**
Using (8.1), it is a straightforward calculation to check that $\mathcal{T}$ is monotone and translation invariant, and that $\mathcal{T}(S_N) = S_N$, $\mathcal{T}(\emptyset) = \emptyset$. The equivalence $\infty \in \mathcal{T}X$ if and only if $\infty \in X$ follows from the fact that $\mathcal{B}$ is not empty. The argument of the preceding paragraph already proved that $\mathcal{T}$ sends bounded sets onto bounded sets and complementary sets of bounded sets onto complementary sets of bounded sets. □

Exercise 8.3. Check that if $\mathcal{T}$ is standard monotone, then its canonical set of structuring elements satisfies (ii). ■

$\mathcal{B}_0 = \{X \mid 0 \in \mathcal{T}X\}$ is not the only set that can be used to represent $\mathcal{T}$. A monotone operator $\mathcal{T}$ can have many such sets and here is their characterization.

Proposition 8.4. *Let $\mathcal{T}$ be a translation invariant standard monotone set operator and let $\mathcal{B}_0$ its canonical set of structuring elements. Then $\mathcal{B}_1$ is another standard set of structuring elements for $\mathcal{T}$ if and only if it satisfies*

(i) $\mathcal{B}_1 \subset \mathcal{B}_0$,

(ii) for all $B_0 \in \mathcal{B}_0$, there is $B_1 \in \mathcal{B}_1$ such that $B_1 \subset B_0$.

Proof. Assume that $\mathcal{T}$ is obtained from some set $\mathcal{B}_1$ by (8.1). Let $\mathcal{B}_0$ be the canonical set of structuring elements of $\mathcal{T}$. Then for every $B_1 \in \mathcal{B}_1$, $\mathcal{T}B_1 = \{\mathbf{x} \mid \mathbf{x} + B \subset B_1 \text{ for some } B \in \mathcal{B}_1\}$. It follows that $0 \in \mathcal{T}B_1$ and therefore $B_1 \in \mathcal{B}_0$. Thus $\mathcal{B}_1 \subset \mathcal{B}_0$. In addition, if $B_0 \in \mathcal{B}_0$, then $0 \in \mathcal{T}B_0$, which means that $0 \in \{\mathbf{x} \mid \mathbf{x}+B_1 \subset B_0 \text{ for some } B_1 \in \mathcal{B}_1\}$ that is $B_1 \subset B_0$ for some $B_1 \in \mathcal{B}_1$.

Conversely, let $\mathcal{B}_1$ satisfy (i) and (ii) and let

$$\mathcal{T}_1 X = \{\mathbf{x} \mid \exists B_1 \in \mathcal{B}_1,\ \mathbf{x} + B_1 \subset X\}.$$

Using (i), one deduces that $\mathcal{T}_1 X \subset \mathcal{T}X$ for every X and using (ii) yields the converse inclusion. Thus $\mathcal{B}_1$ is a structuring set for $\mathcal{T}$. The fact that $\mathcal{B}_1$ is standard is an obvious check using (i) and (ii). □

8.2 The Sup-Inf form

Lemma 8.5. *Let $T : \mathcal{F} \to \mathcal{F}$ be a standard monotone function operator, $\mathcal{T}$ a standard monotone translation invariant set operator and $\mathcal{B}$ a set of structuring elements for $\mathcal{T}$. If T and $\mathcal{T}$ satisfy the commutation of thresholds $\mathcal{T}\mathcal{X}_\lambda u = \mathcal{X}_\lambda T u$, then T has the "sup-inf" representation*

$$Tu(\mathbf{x}) = \sup_{B \in \mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}). \tag{8.2}$$

Proof. For $u \in \mathcal{F}$, set $\tilde{T}u(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(y)$. We shall derive the identity $T = \tilde{T}$ from the equivalence

$$\tilde{T}u(\mathbf{x}) \geq \lambda \Longleftrightarrow Tu(\mathbf{x}) \geq \lambda. \tag{8.3}$$

Assume first that $\mathbf{x} \in \mathbb{R}^N$. Then

$$\begin{aligned}
&Tu(\mathbf{x}) \geq \lambda \overset{(1)}{\Longleftrightarrow} Tu(\mathbf{x}) \geq \mu \text{ for all } \mu < \lambda \overset{(2)}{\Longleftrightarrow} \mathbf{x} \in \mathcal{X}_\mu Tu \text{ for all } \mu < \lambda \\
&\overset{(3)}{\Longleftrightarrow} \mathbf{x} \in \mathcal{T}\mathcal{X}_\mu u \text{ for all } \mu < \lambda \overset{(4)}{\Longleftrightarrow} \exists B \in \mathcal{B},\ \mathbf{x} + B \subset \mathcal{X}_\mu u \text{ for all } \mu < \lambda \\
&\overset{(5)}{\Longleftrightarrow} \text{There is a } B \in \mathcal{B} \text{ such that } \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \mu \text{ for all } \mu < \lambda \\
&\overset{(6)}{\Longleftrightarrow} \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \lambda \overset{(7)}{\Longleftrightarrow} \tilde{T}u(\mathbf{x}) \geq \lambda.
\end{aligned}$$

Equivalence (1) is just a statement about real numbers and (2) is the definition of a level set. It is at (3) that we replace $\mathcal{X}_\mu Tu$ with $\mathcal{T}\mathcal{X}_\mu u$. Equivalence (4) follows by the definition of $\mathcal{T}$ from $\mathcal{B}$ by (8.1). The equivalence (5) is the definition of the level set $\mathcal{X}_\mu u$. Equivalence (6) is another statement about real numbers, and (7) is the definition of $\tilde{T}$.
Assume now that $\mathbf{x} = \infty$. Since for all $B \in \mathcal{L}$, $\infty + B = \{\infty\}$, one obtains $\tilde{T}u(\infty) = u(\infty)$. By assumption $Tu(\infty) = u(\infty)$. This completes the proof of (8.2). □

From the preceding result, we can easily derive a general form for translation and contrast invariant standard monotone operators.

Theorem 8.6. *Let $T : \mathcal{F} \to \mathcal{F}$ be a translation and contrast invariant standard monotone operator. Then it has a "sup-inf" representation* (8.2) *with a standard set of structuring elements.*

Proof. By the level set extension (Theorem 7.19), T defines a unique upper semicontinuous standard monotone set operator $\mathcal{T} : \mathcal{L} \mapsto \mathcal{L}$. $\mathcal{T}$ is defined by the commutation of thresholds, $\mathcal{T}\mathcal{X}_\lambda u = \mathcal{X}_\lambda Tu$. By Lemma 8.5, the commutation with thresholds is enough to ensure that T has the sup-inf representation (8.2) for any set of structuring elements $\mathcal{B}$ of $\mathcal{T}$. □

Definition 8.7. *As a consequence of the preceding theorem, the canonical set of structuring elements of $\mathcal{T}$ will also be called* canonical set of structuring elements of T.

The next theorem closes the loop.

Theorem 8.8. *Given any standard subset $\mathcal{B}$ of $\mathcal{M}$, Equation* (8.2),

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

defines a contrast and translation invariant standard monotone function operator from $\mathcal{F}$ into itself.

Proof. By Theorem 7.16, it is enough to prove that T is the stack filter of $\mathcal{T}$, the standard monotone set operator associated with $\mathcal{B}$. Let us call T' this stack filter and let us check that $Tu(\mathbf{x}) \geq \lambda \Leftrightarrow T'u(\mathbf{x}) \geq \lambda$.
we have $T'u = \sup\{\lambda,\ \mathbf{x} \in \mathcal{T}(\mathcal{X}_\lambda u)\}$. Thus by (8.1),

$$T'u(\mathbf{x}) \geq \lambda \Leftrightarrow \forall \mu < \lambda,\ \exists B,\ \mathbf{x} + B \subset \mathcal{X}_\mu u.$$

On the other hand,

$$\begin{aligned} Tu(\mathbf{x}) = \sup_{B \in \mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x}+B} u \geq \lambda \quad &\Leftrightarrow \\ \forall \mu < \lambda,\ \exists B \in \mathcal{B},\ \inf_{\mathbf{y} \in \mathbf{x}+B} u \geq \mu \quad &\Leftrightarrow \\ \forall \mu < \lambda, \exists B \in \mathcal{B},\ \mathbf{x} + B \subset \mathcal{X}_\mu u. & \end{aligned}$$

Thus, $T = T'$. □

We end this section by showing that sup-inf operators can also be represented as inf-sup operators,

$$Tu(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

This is done, in the mathematical morphology terminology, by "duality". The dual operator of a function operator is defined by $\tilde{T}u = -T(-u)$. Notice that $\tilde{\tilde{T}} = T$.

Proposition 8.9. *If T is a standard monotone, translation invariant and contrast invariant operator, then so is $\tilde{T}$. As a consequence, T has a dual "inf-sup" form*

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

where $\tilde{\mathcal{B}}$ is any set of structuring elements for $\tilde{T}$

Proof. Setting $\tilde{g}(s) = -g(-s)$, it is easily checked that $\tilde{g}$ is a contrast change if and only if g is. One has by the contrast invariance of T,

$$\tilde{T}(g(u)) = -T(-g(u)) = -T(\tilde{g}(-u)) = -\tilde{g}(T(-u)) = g(-T(-u)) = g(\tilde{T}u).$$

Thus, $\tilde{T}$ is contrast invariant. The standard monotonicity and translation invariance of $\tilde{T}$ are obvious. Finally, if we have $\tilde{T}u(\mathbf{x}) = \sup_{B \in \tilde{\mathcal{B}}} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})$, then

$$Tu = -\sup_{B \in \tilde{\mathcal{B}}} \inf_{\mathbf{y} \in \mathbf{x}+B} (-u(\mathbf{y})) = -\sup_{B \in \tilde{\mathcal{B}}} (-\sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y})) = \inf_{B \in \tilde{\mathcal{B}}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

□

Exercise 8.4. Check the standard monotonicity and translation invariance of $\tilde{T}$. ■

8.3 Locality and isotropy

For linear filters, locality can be defined by the fact that the convolution kernel is compactly supported. This property is important, as it guarantees that the smoothed image is obtained by a local average. Morphological filters may need a locality property for the same reason.

Definition 8.10. *We say that a translation invariant function operator T on $\mathcal{F}$ is local if there is some $M \geq 0$ such that*

$$(u = u' \text{ on } B(0, M)) \Rightarrow Tu(0) = Tu'(0).$$

The point 0 plays no special role in the definition. By translation invariance it is easily deduced from the definition that for $\mathbf{x} \in \mathbb{R}^N$, the values of $Tu(\mathbf{x})$ only depend upon the restriction of u to $B(\mathbf{x}, M)$.

Proposition 8.11. *Let $T : \mathcal{F} \to \mathcal{F}$ be a contrast and translation invariant standard monotone operator and $\mathcal{B}$ a set of structuring elements for T. If T is local, then $\mathcal{B}_M = \{B \in \mathcal{B} \mid B \subset \overline{B(0,M)}\}$ also is a set of structuring elements for T. Conversely, if all elements of $\mathcal{B}$ are contained in $\overline{B(0,M)}$, then T is local.*

Proof. We prove the statement with the sup-inf form for T. Since T is local if and only if $\tilde{T}$ is, the same result will hold for the inf-sup form. So assume that some local T derives from $\mathcal{B}$ in the sup-inf form,

$$Tu(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in B} u(\mathbf{x}+\mathbf{y}). \tag{8.4}$$

Consider the new function $u_\varepsilon(\mathbf{x}) = u(\mathbf{x}) - \frac{1}{\varepsilon} d(\mathbf{x}, B(0,M))$, where we take for d a distance function on S_N, so that $u_\varepsilon \in \mathcal{F}$. Take any $B \in \mathcal{B}$ containing a point $\mathbf{z} \notin \overline{B(0,M)}$ and therefore not belonging to $\mathcal{B}_M$. Then $\inf_{\mathbf{y}\in B} u_\varepsilon(\mathbf{y}) \leq u(\mathbf{z}) - \frac{1}{\varepsilon} d(\mathbf{z}, B(0,M)) < Tu(0)$ for ε small enough. So we can discard such B's in the computation of $Tu(0)$ by (8.4). Since by the locality assumption $Tu(0) = Tu_\varepsilon(0)$, we obtain

$$Tu(0) = Tu_\varepsilon(0) = \sup_{B\in\mathcal{B}_M} \inf_{\mathbf{y}\in B} u(\mathbf{y}).$$

By the translation invariance of all all considered operators, this proves the direct statement. The converse statement is straightforward. □

We end this paragraph with a definition and an easy characterization of isotropic operators in the sup-inf form. In the next proposition, we actually consider a more general setting, namely the invariance of T under some geometric group of transformations of $\mathbb{R}^N$. Since we use to extend the set and function operators to S_N, we shall extend such transforms by setting $g(\infty) = \infty$.

Definition 8.12. *Let T (resp $\mathcal{T}$) be a standard monotone contrast and translation invariant function operator associated with some set of structuring elements $\mathcal{B}$ (resp. a standard monotone set operator associated with $\mathcal{B}$). We say that $\mathcal{B}$ is invariant under a group G of transformations of S_N onto S_N if, for all $g \in G$, $B \in \mathcal{B}$ implies $gB \in \mathcal{B}$. Define the operator I_g on functions $u : S_N \to \mathbb{R}$ by $I_g u(\mathbf{x}) = u(g\mathbf{x})$. If, for all $g \in G$, $TI_g = I_g T$ (resp. $\mathcal{T}g = g\mathcal{T}$), we say that T (resp. $\mathcal{T}$) is invariant under G. In particular, we say that T (resp. $\mathcal{T}$) is* isotropic *if it commutes with all linear isometries R of $\mathbb{R}^N$, and* affine invariant *if it commutes with all linear maps A with determinant 1.*

Proposition 8.13. *Let G be any group of linear maps : $g : \mathbb{R}^N \to \mathbb{R}^N$ extended to S_N by setting $g(\infty) = \infty$. If T (resp. $\mathcal{T}$) is translation invariant and invariant under G and $\mathcal{B}$ is a standard set of structuring elements for T (resp $\mathcal{T}$), then $G\mathcal{B} = \{gB \mid g \in G,\ B \in \mathcal{B}\}$ is another, G-invariant, standard set of structuring elements. Conversely, if $\mathcal{B}$ is a standard and G-invariant set of structuring elements for T (resp. $\mathcal{T}$), then this operator is G-invariant (and translation invariant.)*

Proof. All the verifications are straightforward. The only point to mention is that the considered groups are made of transforms sending bounded sets onto bounded sets and complementary sets of bounded sets onto complementary sets of bounded sets. □

Exercise 8.5. Prove carefully Proposition 8.13. ■

Some terminology.
It would be tedious to state theorems on operators on $\mathcal{F}$ with such a long list of requirements as *Standard Monotone, Translation and Contrast Invariant, Isotropic*. We shall keep the initials and call such operators *SMTCII operators.* All the examples we consider in this book are actually SMTCII operators. Not all are local, so we will specify it when needed. Operators can be still more invariant, in fact affine invariant, and we will specify it as well. Since all of these operators T have an inf-sup or a sup-inf form, we always take for $\mathcal{B}$ a standard structuring set reflecting the properties of T, that is, bounded in $B(0, M)$ when T is local and invariant by the same group as T. A last thing to specify is this: We have restricted our analysis to operators defined on $\mathcal{F}$. On the other hand, their inf-sup form permits to extend them on all measurable functions and we shall still denote the resulting operator by T. Tu can then assume the $-\infty$ and $+\infty$ values. All the same, it is an immediate check to see that this extension still is monotone and commutes with contrast changes:

Proposition 8.14. *Let T be a function operator in the* inf-sup *or* sup-inf *form associated with a standard set of structuring elements $\mathcal{B} \subset \mathcal{M}$. Then T is standard monotone and contrast invariant on the set of all bounded measurable functions of S_N.*

Exercise 8.6. Prove Proposition 8.14. ■

8.4 The who's who of monotone contrast invariant operators

The aim of this short section is to draw a synthetic picture of an equivalence chain built up in this chapter and in Chapter 7. We have constructed three kinds of objects,

- contrast and translation invariant standard monotone function operators $T:\ \mathcal{F} \to \mathcal{F}$;
- translation invariant standard monotone set operators $\mathcal{T}$ defined on $\mathcal{L}$;
- standard sets of structuring elements $\mathcal{B}$.

The results proven so far can be summarized in the following theorem.

Theorem 8.15. *Given any of the standard objects T, $\mathcal{T}$ and $\mathcal{B}$ mentioned above, one can pass to any other one by using one of the six formulae given below.*

$$\begin{array}{ll}
\mathcal{B} \to T, & Tu(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y});\\
\mathcal{B} \to \mathcal{T}, & \mathcal{T}X = \{\mathbf{x} \mid \exists B \in \mathcal{B},\ \mathbf{x}+B \subset X\};\\
\mathcal{T} \to T, & Tu(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{T}\mathcal{X}_\lambda u\};\\
T \to \mathcal{T}, & \mathcal{T}(\mathcal{X}_0 u) = \mathcal{X}_0(Tu);\\
\mathcal{T} \to \mathcal{B}, & \mathcal{B} = \{B \in \mathcal{L} \mid 0 \in \mathcal{T}B\};\\
T \to \mathcal{B}, & \text{by } T \to \mathcal{T} \text{ and } \mathcal{T} \to \mathcal{B}.
\end{array}$$

In addition, $\mathcal{B}$ can be bounded in some $B(0,M)$ if and only if T is local; T or $\mathcal{T}$ is G-invariant, for instance isotropic, if and only if it derives from some G-invariant (isotropic) $\mathcal{B}$. If an operator has the inf-sup or sup-inf form for some $\mathcal{B}$, it can be extended to all measurable functions on $\mathbb{R}^N$ into a monotone and contrast invariant operator.

Proof. Theorem 8.2 yields $\mathcal{T} \to \mathcal{B}$ and $\mathcal{B} \to \mathcal{T}$; Theorem 7.16 yields $\mathcal{T} \to T$; Theorem 8.6 yields $T \to \mathcal{T} \to \mathcal{B}$; Theorem 7.19 yields $T \to \mathcal{T}$. The final statements come from Propositions 8.11, 8.13 and 8.14. □

So we get a full equivalence between all objects, but we have left apart the commutation with thresholds property. When we define a set operator $\mathcal{T}$ from a function operator T by the level set extension, we know that $\mathcal{T} : \mathcal{L} \to \mathcal{L}$ is upper semicontinuous and that the commutation with thresholds $\mathcal{X}_\lambda(Tu) = \mathcal{T}(\mathcal{X}_\lambda u)$ holds. Conversely, if we define a function operator T as the stack filter of a standard monotone set $\mathcal{T}$, we do not necessarily have the commutation of thresholds ; this is true only if $\mathcal{T}$ is upper semicontinuous on $\mathcal{L}$ (see Theorem 7.16) and this upper semicontinuity property is not always granted for interesting monotone operators, particularly when they are affine invariant. Fortunately enough, the commutation with thresholds is "almost" satisfied for any stack filter as we state in Proposition 8.18 in the next section.

8.4.1 Commutation with thresholds almost everywhere

In this section we always assume the considered sets to belong to $\mathcal{M}$ and the considered functions to be Lebesgue measurable. We say that a set X is contained in a set Y almost everywhere if

$$\text{measure}(X \setminus Y) = 0,$$

where measure denotes the usual Lebesgue measure in $\mathbb{R}^N$. 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 measure$(\{\mathbf{x}, u(\mathbf{x}) \neq v(\mathbf{x})\}) = 0$.

Lemma 8.16. *Let $(X_\lambda)_{\lambda\in\mathbb{R}}$ be a nonincreasing family of sets of $\mathcal{M}$, that is $X_\lambda \subset X_\mu$ if $\lambda \geq \mu$. Then, for almost every λ in $\mathbb{R}$,*

$$X_\lambda = \bigcap_{\mu<\lambda} X_\mu, \qquad \textit{almost everywhere.} \tag{8.5}$$

Proof. Consider an integrable and strictly positive continuous function $h \in L^1(\mathbb{R}^N)$ (for instance, the gaussian.) Set $m(X) = \int_X h(\mathbf{x})d\mathbf{x}$. We notice that $m(X) = 0$ if and only if measure$(X) = 0$. 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 (8.5). □

Corollary 8.17. *Let $(X_\lambda)_{\lambda \in \mathbb{R}}$ be a family of measurable subsets of S_N such that $X_\lambda \subset X_\mu$ for $\lambda \geq \mu$, $X_\lambda = \emptyset$ for $\lambda \geq \lambda_0$, $X_\lambda = S_N$ for $\lambda \leq \mu_0$. Then the function u defined on S_N by the superposition principle*

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

is bounded and satisfies for almost every λ, $X_\lambda = \mathcal{X}_\lambda u$ almost everywhere.

Proof. It is easily checked that $\mu_0 \leq u \leq \lambda_0$. We have

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

Now, if $\mathbf{x} \in X_\lambda$, we have $\sup\{\mu \mid \mathbf{x} \in X_\mu\} \geq \lambda$ which implies $\mathbf{x} \in \mathcal{X}_\lambda u$. Thus, $X_\lambda \subset \mathcal{X}_\lambda u$. Conversely, let λ be chosen so that $X_\lambda = \cap_{\mu<\lambda} X_\mu$ almost everywhere. This is by Lemma 8.16 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. □

Exercise 8.7. By using Corollary 8.17 show that if two measurable functions u and v are such that $\mathcal{X}_\lambda u = \mathcal{X}_\lambda v$ almost everywhere for almost every λ, then u and v are almost everywhere equal. ■

Proposition 8.18. *Let $\mathcal{T} : \mathcal{L} \to \mathcal{M}$ be a standard monotone set operator and T its stack filter. If $u \in \mathcal{F}$ then for almost every level $\lambda \in \mathbb{R}$,*

$$\mathcal{X}_\lambda(Tu) = \mathcal{T}(\mathcal{X}_\lambda(u)) \quad \textit{almost everywhere.}$$

Proof. Since Tu is obtained from the sets $\mathcal{T}(\mathcal{X}_\lambda u)$ by superposition principle, this is an immediate consequence of Corollary 8.17. □

8.4.2 Chessboard dilemma and fattening effect

With any standard monotone contrast invariant function operator T we can associate a stack filter $\mathcal{T}$, and by the above proposition the commutation with thresholds is true for almost every level. Yet for some levels the commutation with thresholds may not occur! As follows from the proofs of Proposition 8.18 and Lemma 8.16, the levels λ for which commutation does not occur are those

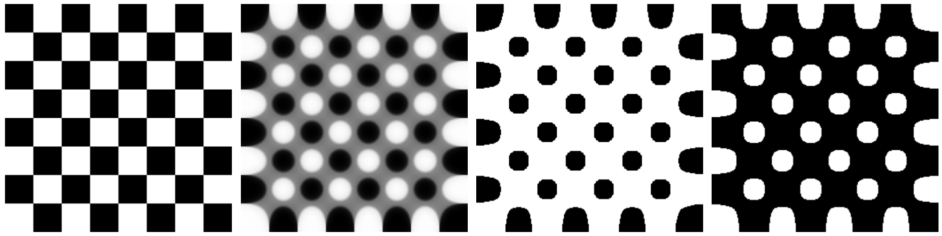

Figure 8.1: The chessboard dilemma. Left: a chessboard image. The next image is obtained by a self-dual version of the median filter. Notice the expansion of the median grey level, 127.5, who was invisible in the original image and grows in the second one. This effect is called "fattening effect". The third and fourth image show the evolution of the level sets at levels 127 and 128 respectively. This experiment illustrates a dilemma as to whether we consider the chessboard as black squares on white background, or conversely. There is fundamental perceptual instability here, that no theory can eliminate.

such that measure($\{\mathbf{x} \mid \mathcal{X}_\lambda u = \lambda\} > 0$. We call such sets flat parts of the function u.

As will be illustrated in Figure 8.4.2, Tu can have flat parts even if u had none. If Tu has a flat part at level λ, by monotonicity for every $\varepsilon > 0$ the sets $\mathcal{T}(\mathcal{X}_{\lambda-\varepsilon}u)$ and $\mathcal{T}(\mathcal{X}_{\lambda+\varepsilon}u)$ differ by a measure larger than the measure of the flat part. Thus, the set $\mathcal{X}_\lambda u$ becomes somewhat ambiguous for the operator $\mathcal{T}$.

Figure 8.4.2 proves that this ambiguity is perceptually sound. In this figure a self-dual version T of the median filter has been applied iteratively to a function u whose grid values are equal to 255 on the white squares and to 0 on the black square. The function is then interpolated by standard bilinear interpolation. The iso-level set $I_{127.5}u := \{\mathbf{x} \mid u(\mathbf{x}) = 127.5\}$ consists of the line segments separating the squares and has therefore zero measure. As we know, the median filter tends to smooth, to round off the level lines of the image. Yet we have with a chessboard a fundamental ambiguity : are these iso-level lines surrounding the black squares, or are they surrounding the white squares? In other terms, do we see in a chessboard a set of white squares on black background, or conversely?

Since our operator is self-dual it doesn't favor any of the considered interpretations: it rounds off simultaneously the lines surrounding the black squares and the level lines surrounding the white squares (second image of Figure 8.4.2). This results in the "fattening" of the level lines separating white and black, which have the mid-level 128. Hence the appearance in the second image of a grey zone separating the smoothed out black and white squares. If we take a level set $\mathcal{X}_\varepsilon Tu$ of this image with $\varepsilon < 0$ (third image), the fattened set joins the level set and we observe black squares on white background. Symmetrically if $\varepsilon > 0$ the level set shows white squares on black background.

8.5 Exercises

Exercise 8.8. It is useful to have a test for $\mathcal{B}$ to determine whether or not the operator $\mathcal{T}$ can be expected to be upper semicontinuous on $\mathcal{L}$. Prove that the translation-invariant monotone operator in Theorem 8.2 defined by a given set $\mathcal{B}$ is upper semicontinuous on $\mathcal{L}$ if and only if the following condition holds: If $\bigcap_{n\in\mathbb{N}} \mathcal{T}X_n \neq \emptyset$, then there is a $B \in \mathcal{B}$ such that $\mathbf{x} + B \subset \bigcap_{n\in\mathbb{N}} X_n$, where $\mathbf{x} \in \bigcap_{n\in\mathbb{N}} \mathcal{T}X_n$ and $(X_n)_{n\in\mathbb{N}}$ is any nonincreasing sequence in $\mathcal{L}$. ■

Exercise 8.9. Suppose that $\mathcal{B} \subset \mathcal{L}$ contains exactly one set. Show that $\mathcal{T}$ is u.s.c. Generalize this to the case where $\mathcal{B}$ contains a finite number of sets. ■

Exercise 8.10. Use Theorem 8.6 and Proposition 8.4 to show that the extrema killer T_a can be represented as a sup-inf function operator with the structuring elements

$$\mathcal{B}_a = \{B \mid B \text{ is compact, connected, meas}\,(B) = a, \text{ and } 0 \in B\} \cup \{\infty\}. \blacksquare$$

Check that $\mathcal{B}_a$ is standard.

Exercise 8.11. Let $\mathcal{B} = \{\{\mathbf{x}\} \mid \mathbf{x} \in D(0,1)\}$, $D(0,1) = \{\mathbf{x} \mid |\mathbf{x}| \leq 1\}$ and consider the associated set operator $\mathcal{T}$ and the associated function operator T, *defined on all measurable sets and functions of* $\mathbb{R}^N$ by formulas (8.1) and (8.2).

1) Check that $Tu(\mathbf{x}) = \sup_{\mathbf{y}\in\mathbf{x}+D} u(\mathbf{y})$.

2) Let $(q_n)_{n\in\mathbb{N}}$ be a countable dense set in $\mathbb{R}^N$ and consider u defined by $u(\mathbf{x}) = 1-1/n$ if $\mathbf{x} = q_n$ and $u(\mathbf{x}) = 0$ otherwise. Show that $\mathcal{T}\mathcal{X}_1 u \neq \mathcal{X}_1 Tu$. The operator T in this exercise is one of the classic image operators called a *dilation*. Check that T commutes with thresholds when its domain of definition is restricted to $\mathcal{F}$ and the domain of $\mathcal{T}$ to $\mathcal{L}$. This example shows that this restriction is useful to get a simple theory. ■

Exercise 8.12. Show the following property used in the proof of Lemma : if h is a positive continuous integrable function on $\mathbb{R}^N$ and if we set $m(X) = \int_X h(\mathbf{x})d\mathbf{x}$, then for every measurable set X, $m(X) = 0$ if and only if measure$(X) = 0$. ■

8.6 Comments and references

The formalism presented in this chapter is due to Matheron [202] in the case of set operators and to Serra [253] and Maragos [192] in the case of function operators. Serra's formalism is actually more general than the one presented here; it will be developed in Chapter **??**, which is about "nonflat" morphology. Our presentation relating the sup-inf form of the operator directly to contrast invariance and establishing the full equivalence between sup-inf operators and contrast-invariant monotone operators is original. The fact, proven in Section 8.18 that commutation with thresholds occurs without further assumption was proven in [132].

The mysterious "set of structuring elements" has received a great deal of attention in the literature. Here are a few references: on finding the right set of structuring elements [245, 264]; on simplifying them [252]; on decomposing them into simpler ones as one does with linear filters [225, 295, 296]; on reducing the number [237].

Chapter 9

Erosions and Dilations

We are going to study in detail two of the simplest operators of mathematical morphology, the erosions and dilations. In fact, there will be essentially four operators: two set operators and the two related function operators. These operators will depend on a scale parameter t. We will also study the underlying PDEs $\partial u/\partial t = c|Du|$, where $c = 1$ for dilations and $c = -1$ for erosions.

9.1 Set and function erosions and dilations

We saw in chapter 8 that every contrast-invariant monotone function operator has a sup-inf and an inf-sup representation in terms of some set of structuring elements. This is the point of view we take here, and furthermore, we assume that the set of structuring elements $\mathcal{B}$ has the simplest possible form, namely, $\mathcal{B} = \{B\}$. We actually introduce a parameter t scaling the size of B and therefore consider the two operators of the next definition.

Definition 9.1. *For $u \in \mathcal{F}$, define $D_{tB}u = D_t u$ by*

$$D_t u(\mathbf{x}) = \sup_{\mathbf{y} \in tB} u(\mathbf{x} - \mathbf{y}), \tag{9.1}$$

the "dilation of u by tB. In the same way, define $E_{tB}u = E_t u$, the "erosion of u by $-tB$", by

$$E_t u(\mathbf{x}) = \inf_{\mathbf{y} \in -tB} u(\mathbf{x} - \mathbf{y}). \tag{9.2}$$

These function operators have associated set operators.

Definition 9.2. *Let B be a non empty subset of $\mathbb{R}^N$ and let $t \geq 0$ be a scale parameter. The set operators $\mathcal{D}_{tB}$ and $\mathcal{E}_{tB}$ are defined on subsets $X \in \mathcal{M}(\mathbb{R}^N)$ by*

$$\mathcal{D}_{tB}X = \mathcal{D}_t X = X + tB = \{\mathbf{x} \mid \exists b \in B, \mathbf{x} - tb \in X\}, \tag{9.3}$$

$$\mathcal{E}_{tB} = \mathcal{E}_t X = \{\mathbf{x} \mid \mathbf{x} + tB \subset X\}, \tag{9.4}$$

and extended to $\mathcal{M}(S_N)$ by the standard extension (Definition 7.1.) $\mathcal{D}_t X$ is called the dilation of X by B at scale t. $\mathcal{E}_t X$ is called the erosion of X by B at scale t.

Exercise 9.1. (Duality formulas.) Show that $E_{tB}u = -D_{-tB}(-u)$ and $\mathcal{E}_{tB}X = (\mathcal{D}_{-tB}X^c)^c$. ∎

Exercise 9.2. Show that if B is bounded, dilations and erosions are standard monotone operators. Compute their associated set of structuring elements (Proposition 8.2) and check that it is standard. ∎

Theorem 9.3. *The function erosion by tB is the stack filter of the set erosion by tB ; the function dilation by tB is the stack filter of the set dilation by $t\overline{B}$ and the commutation with thresholds holds. In other terms for $u \in \mathcal{F}$ and all λ in $\mathbb{R}$, and calling $\overline{\mathcal{D}}_t$ the dilation by $t\overline{B}$,*

$$D_t u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{D}_t \mathcal{X}_\lambda u\}, \quad \overline{\mathcal{D}}_t \mathcal{X}_\lambda u = \mathcal{X}_\lambda D_t u; \tag{9.5}$$

$$E_t u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{E}_t \mathcal{X}_\lambda u\}, \quad \mathcal{E}_t \mathcal{X}_\lambda u = \mathcal{X}_\lambda E_t u. \tag{9.6}$$

Proof. We prove the statement for the dilations, the case of the erosions being just simpler. Consider some $X \in \mathcal{L}$ and $u(\mathbf{x}) \le 0$ a function vanishing on X only. By the definition 7.18 of the level set extension $\tilde{\mathcal{D}_t}$ of D_t, $\tilde{\mathcal{D}}_t(X) = \mathcal{X}_0 D_t(u)$. Thus, using (9.3),

$$\mathbf{x} \in \tilde{\mathcal{D}}_t(X) \Leftrightarrow (D_t u)(\mathbf{x}) = 0 \Leftrightarrow \sup_{\mathbf{y} \in -tB} u(\mathbf{x} - \mathbf{y}) = 0 \Leftrightarrow$$

$$\exists \mathbf{y} \in t\overline{B},\ \mathbf{x} - \mathbf{y} \in X \Leftrightarrow \mathbf{x} \in X + t\overline{B} \Leftrightarrow \mathbf{x} \in \overline{\mathcal{D}}_t(X).$$

□

The operators $\mathcal{D}_t$ and $\mathcal{E}_t$ are in a certain sense the inverse of each other. This is clearly the case, for example, if $B = \{\mathbf{x}_0\}$. Then $\mathcal{D}_t$ is just the translation by $t\mathbf{x}_0$, and $\mathcal{E}_t = \mathcal{D}_t^{-1}$ is the translation by $-t\mathbf{x}_0$. If B is the open ball centered at zero with radius one, then $\mathcal{D}_t X$ is the set of all points whose distance from X is less than t, or the t-neighborhood of X. When B is symmetric with respect to zero, the operator $\mathcal{D}_t\mathcal{E}_t$ is called an *opening at scale t* and $\mathcal{E}_t\mathcal{D}_t$ is called a *closing at scale t*. These names have a topological origin. If B is the open ball centered at zero with radius one, then the opening at scale t of a set X is the union of all balls with radius t contained in X. The interior of X is the union of all open balls contained in X; it is also the largest open set contained in X. If we call the interior map $\mathcal{T}^\circ X = X^\circ$ the opening, then an opening at scale t appears as a quantified opening (see Exercise 9.6). The topological statement "the closure of the complement of X is the complement of the interior of X" has its counterpart for openings and closings at scale t, as shown in exercise 9.6. The actions of erosions and dilations are illustrated in Figures 9.2, 9.2, and 9.2; actions of openings and closings are illustrated in Figures 9.2, 9.2, 9.2, 9.3, and 9.3.

9.2 Multiscale aspects

We say that the family of dilations $\{D_t \mid t > 0\}$ associated with a structuring element B is *recursive* if $D_t D_s = D_{t+s}$ for all $s, t > 0$, and similarly for the family $\{E_t \mid t > 0\}$. (A recursive family is also called a *semigroup*.) Being recursive is a very desirable property for any family of scaled operators used

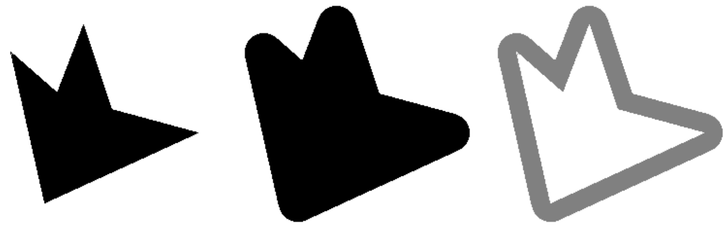

Figure 9.1: Dilation of a set. Left to right: A set; its dilation by a ball of radius 20; the difference set.

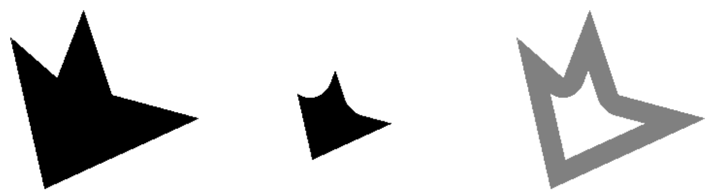

Figure 9.2: Erosion of a set. Left to right: A set; its erosion by a ball of radius 20; the difference set.

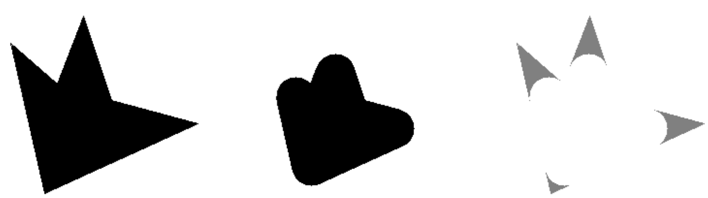

Figure 9.3: Opening of a set as curvature threshold from above. Left to right: A set X; its opening by a ball of radius 20; the difference set. This opening transforms X into the union of all balls of radius 20 contained in it. The resulting operation can be understood as a threshold from above of the curvature of the set boundary.

for image analysis. Having $D_t = (D_{t/n})^n$ is useful for practical computations. $\{D_t \mid t > 0\}$ and $\{E_t \mid t > 0\}$ will be recursive if and only if B is convex, but before proving this result we need the condition for B to be convex given in the next lemma. The proof of the next statement is an easy exercise.

Lemma 9.4. *B is convex if and only if $(s+t)B = sB + tB$ for all $s, t \geq 0$.*

Proposition 9.5. *The dilations $\mathcal{D}_t$ and the erosions $\mathcal{E}_t$ are recursive if and only the structuring element B is convex.*

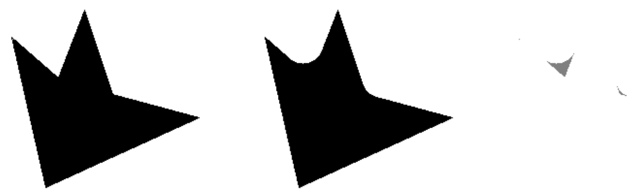

Figure 9.4: Closing of a set as a curvature threshold from below. Left to right: A set X; its closing by a ball of radius 20; the difference set. The closing of X is just the opening of X^c. It can be viewed as a threshold from below of the curvature of the set boundary.

Proof. By the stack filter construction and the level set extension, we see that the proof of the equivalence can be performed on set dilations. Taking for simplicity B closed, we have

$$\mathcal{D}_t\mathcal{D}_s X = (X + sB) + tB = X + sB + tB$$

and

$$\mathcal{D}_{s+t}X = X + (s+t)B.$$

If $(t+s)B = tB + sB$, then clearly $\mathcal{D}_t\mathcal{D}_s X = \mathcal{D}_{s+t}X$. Conversely, if $\mathcal{D}_t\mathcal{D}_s X = \mathcal{D}_{s+t}X$, then by taking $X = \{0\}$ we see that $(t+s)B = tB + sB$. One can deduce the corresponding equivalence for erosions from the duality formula (exercise 9.1.) $\square$

9.3 The PDEs associated with erosions and dilations

As indicated in the introduction to the chapter, scaled dilations and erosions are associated with the equations $\partial u/\partial t = \pm|Du|$. To explain this connection, we begin with a bounded convex set B that contains the origin, and we define the gauge $\|\cdot\|_B$ on $\mathbb{R}^N$ associated with B by $\|\mathbf{x}\|_B = \sup_{\mathbf{y}\in B}(\mathbf{x}\cdot\mathbf{y})$. If B is a ball centered at the origin with radius one, then $\|\cdot\|_B$ is the usual Euclidean norm, which we write simply as $|\cdot|$.

Proposition 9.6. *[Hopf–Lax formula [98, 179]]. Assume that B is a bounded convex set in $\mathbb{R}^N$ that contains the origin. Given $u_0 : \mathbb{R}^N \to \mathbb{R}$, define $u : \mathbb{R}^+ \times \mathbb{R}^N \to \mathbb{R}$ by $u(t,\mathbf{x}) = D_t u_0(\mathbf{x})$. Then u satisfies the equation*

$$\frac{\partial u}{\partial t} = \|Du\|_{-B}$$

at each point $(t,\mathbf{x})$ where u has continuous derivatives in t and $\mathbf{x}$. The same result hold when D_t is replaced by E_t and the equation is replaced with $\partial u/\partial t = -\|Du\|_{-B}$.

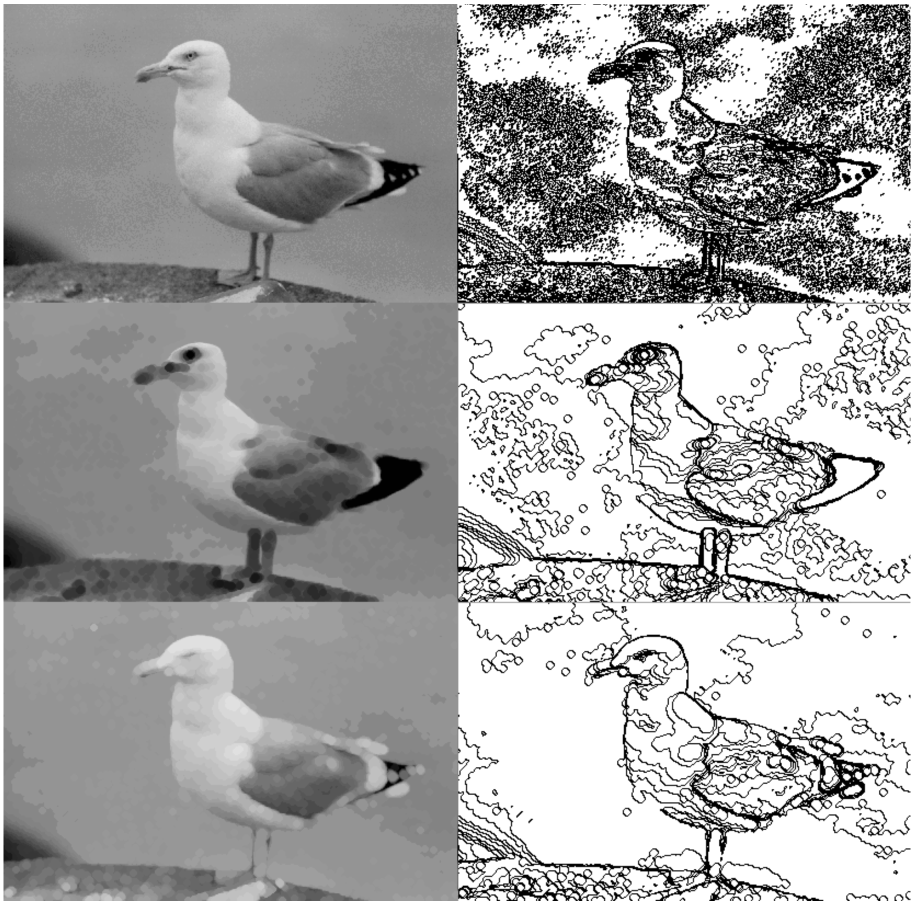

Figure 9.5: Erosion and dilation of a natural image. First row: a sea bird image and its level lines for all levels multiple of 12. Second row: an erosion with radius 4 has been applied. On the right, the resulting level lines where the circular shape of the structuring element (a disk with radius 4) 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 the eye of the bird, are expanded. Third row: the effect of a dilation with radius 4 and the resulting level lines. We see how local minima are removed (for example, 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.

Proof. We begin by proving the result for D_t at $t = 0$. Thus assume that u_0 is C^1 at $\mathbf{x}$. Then

$$u_0(\mathbf{x} - \mathbf{y}) - u_0(\mathbf{x}) = -Du_0(\mathbf{x}) \cdot \mathbf{y} + o(|\mathbf{y}|),$$

and we have by applying D_h,

$$u(h, \mathbf{x}) - u(0, \mathbf{x}) = \sup_{\mathbf{y} \in hB} (-Du_0(\mathbf{x}) \cdot \mathbf{y} + o(|\mathbf{y}|)).$$

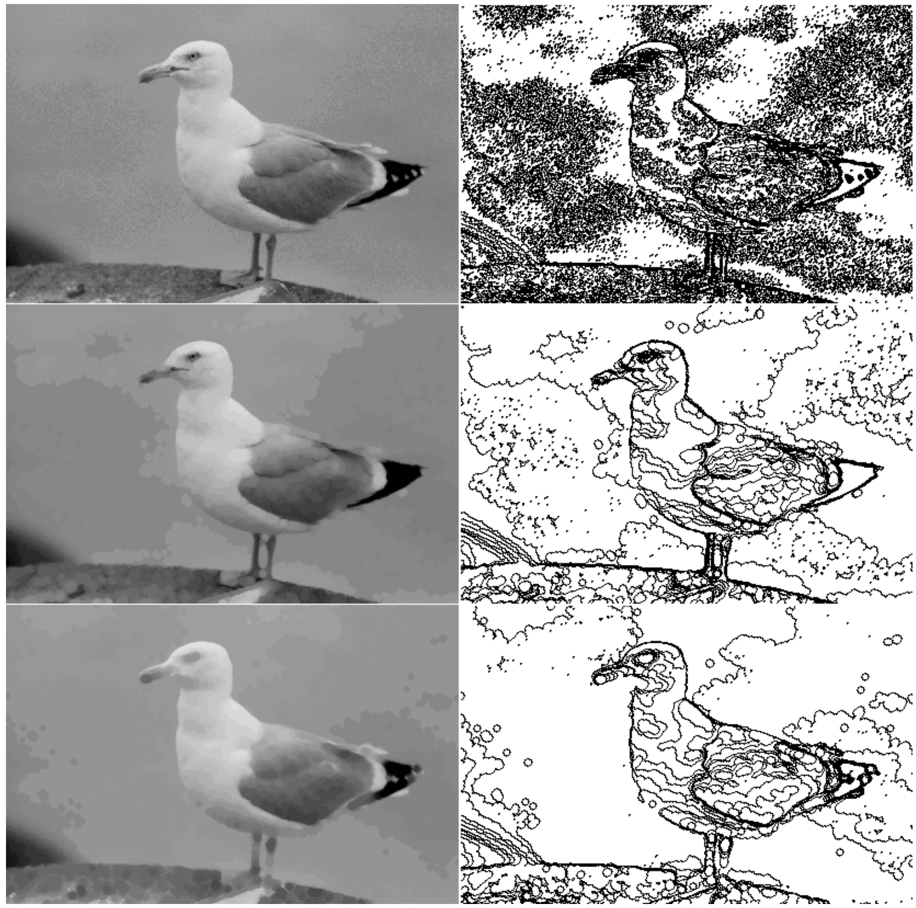

Figure 9.6: Openings and closings of a natural image. First row: the original image and its level lines for all levels multiple of 12. Second row: an opening with radius 4 has been applied. Third row: a closing with radius 4 has been applied. We can recognize the circular shape of the structuring element in the level lines displayed on the right.

Since B is bounded, the term $o(|\mathbf{y}|)$ is $o(|h|)$ uniformly for $\mathbf{y} \in hB$, and we get

$$u(h,\mathbf{x}) - u(0,\mathbf{x}) = h \sup_{\mathbf{z}\in B}((-Du_0(\mathbf{x}) \cdot \mathbf{z}) + o(|h|).$$

We can divide both sides by h and pass to the limit as $|h| \to 0$ to obtain

$$\frac{\partial u}{\partial t}(0,\mathbf{x}) = \|Du_0(\mathbf{x})\|_{-B},$$

which is the result for $t = 0$. For an arbitrary $t > 0$, we have $D_{t+h} = D_t D_h = D_h D_t$, and we can write

$$u(t+h,\mathbf{x}) - u(t,\mathbf{x}) = D_h u(t,\cdot)(\mathbf{x}) - u(t,\mathbf{x}).$$

By repeating the argument made for $t = 0$ with u_0 replaced with $u(t,\cdot)$, we arrive at the general result. The proof for E_t is similar. $\square$

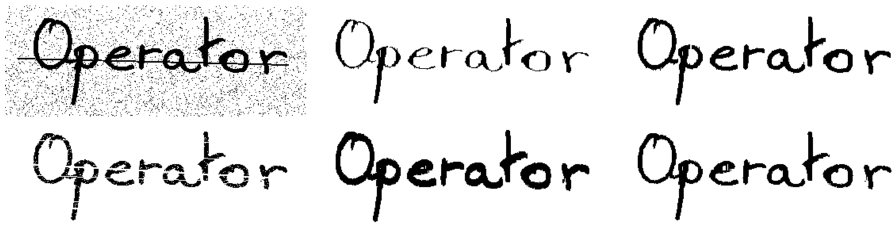

Figure 9.7: Denoising based on openings and closings. First row: scanned picture of the word "operator" with black dots and a black line added; a dilation with a 2×5 rectangle; an erosion with the same structuring element applied to the middle image. The resulting operator is a closing. Small black structures are removed by such a process. Second row: the word "operator" with a white line and white dots inside the letters; erosion with a rectangle 2×5; a dilation with the same structuring element applied to the middle image. The resulting operator is an opening. This time, small white structures are removed.

Exercise 9.3. Prove the above result for E_t.

■

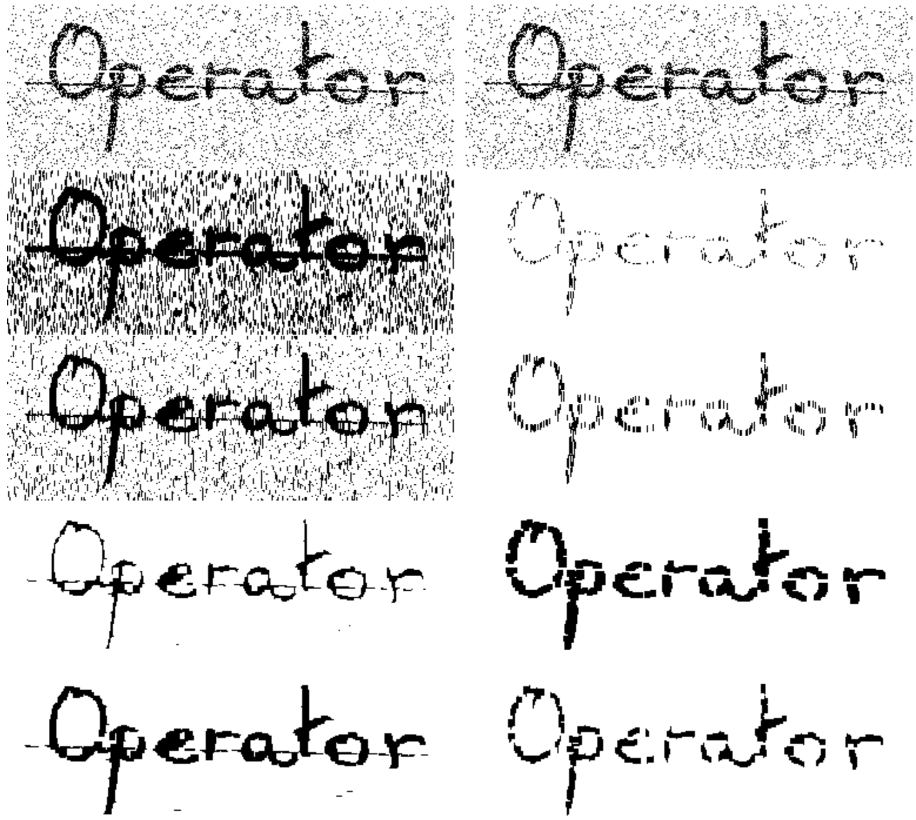

Figure 9.8: Examples of denoising based on opening or closing, as in Figure 9.7. Perturbations made with both black and white lines or dots have been added to the "operator" image. First column, top to bottom: original perturbed image; erosion with a 1×3 rectangle; then dilation with the same structuring element. (In other words, opening with this rectangle.) Then a dilation is applied with a rectangle 3×1, and finally an erosion with the same rectangle. 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. These operators do not commute. See Figure **??**, where an application of the median filter is more successful.

9.4 Exercises

Exercise 9.4. A straightforward adaptation on a grid $\mathbb{Z} \times \mathbb{Z}$ of the formulas $u(t,x) := \sup_{y \in B(x,t)} u_0(y)$ for dilation and $u(t,x) := \inf_{y \in B(x,t)} u_0(y)$ for erosion leads to the zero-order schemes

$$u^{n+1}(i,j) = \sup_{(k,l) \in B((i,j),t) \cap \mathbb{Z}^2} u^n(k,l)$$

and

$$u^{n+1}(i,j) = \inf_{(k,l) \in B((i,j),t) \cap \mathbb{Z}^2} u^n(k,l), \quad u^0(i,j) = u_0(i,j).$$

Unfortunately, the zero-order schemes are strongly grid dependent. They do not make any difference between two balls which contain the same discrete pixels. In particular, such schemes only permit discrete motions of the shape boundaries. Thus, they are efficient only when t is large. Section 9.3 suggests that we can implement erosion and dilations on a finite image grid by more clever numerical schemes. One can try to discretize the associated PDE's $\partial u/\partial t = \pm|Du|$ by the Rouy-Tourin scheme:

$$u_{ij}^{n+1} = u_{ij}^n + \Delta t \left(\max(0, u_{i+1,j}^n - u_{ij}^n, u_{i-1,j}^n - u_{ij}^n)^2 + \max(0, u_{i,j+1}^n - u_{ij}^n, u_{i,j-1}^n - u_{ij}^n)^2\right)^{\frac{1}{2}}$$

for dilation and

$$u_{ij}^{n+1} = u_{ij}^n - \Delta t \left(\max(0, u_{ij}^n - u_{i+1,j}^n, u_{ij}^n - u_{i-1,j}^n)^2 + \max(0, u_{ij}^n - u_{i,j+1}^n, u_{ij}^n - u_{i,j-1}^n)^2\right)^{\frac{1}{2}}$$

for erosion. In both cases if $t = n\Delta t$ then $u^n(i,j)$ is a discrete version of $u(t,(i,j))$.

1) Explain why the schemes are consistent with their underlying partial differential equation. Check that with this clever scheme local maxima of u^n do not go up by dilation and local minima do not go down by erosion. Show that for example the following scheme would be a catastrophe at extrema (you'll have to try it anyway):

$$u_{ij}^{n+1} = u_{ij}^n + \Delta t \left(\max(|u_{i+1,j}^n - u_{ij}^n|, |u_{i-1,j}^n - u_{ij}^n|)^2 + \max(|u_{i,j+1}^n - u_{ij}^n|, |u_{i,j-1}^n - u_{ij}^n|)^2\right)^{\frac{1}{2}}.$$

2) Implement the schemes and compare their performance with the discrete zero order schemes for several shapes and images.

3) Compute on some well-chosen images the "top hat transforms" $u - O_t u$ and $F_t u - u$. The first transform aims at extracting all structures from an image which are thinner than t and have brightness above the average. The second transform does the same job for dark structures. These transforms can be successfully applied on aerial images for extracting roads or rivers, and in many biological applications. ■

Exercise 9.5. Show that $E_t(u) = -D_t(-u)$ if B is symmetric with respect to zero. ■

Exercise 9.6.

(i) Let $B = \{\mathbf{x} \mid |\mathbf{x}| < 1\}$. Show that $\mathcal{D}_t\mathcal{E}_t X$ is the union of all open balls with radius t contained in X.

(ii) Let B be any structuring element that is symmetric with respect to zero. Write $X^c = \mathbb{R}^N \setminus X$. Show that $\mathcal{D}_t X^c = (\mathcal{E}_t X)^c$. Use this to show that $\mathcal{E}_t\mathcal{D}_t X^c = (\mathcal{D}_t\mathcal{E}_t X)^c$. ■

Exercise 9.7. Prove that the dilation and erosion set operators associated with B are standard monotone if and only if B is bounded. If B is bounded and isotropic, prove that the associated erosion and dilation function operators are local *SMTCII* operators. ■

9.5 Comments and references

Erosions and dilations. Matheron introduced dilations and erosions as useful tools for set and shape analysis in his fundamental book [202]. A full account of the properties of dilations, erosions, openings, and closings, both as set operators and function operators, can be found in Serra's books [253, 255]. We also suggest the introductory paper by Haralick, Sternberg, and Zhuang [136] and an earlier paper by Nakagawa and Rosenfeld [214]. An axiomatic algebraic approach to erosions, dilations, openings, and closings has been developed by Heijmans and Ronse [139, 239]. We did not develop this algebraic point of view here. The obvious relations between the dilations and erosions of a set and the distance function have been exploited numerically in [144], [169], and [259]. 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 for computing the skeleton this way is proposed in [170].

The PDEs. The connection between the PDEs $\partial u/\partial t = \pm|Du|$ and multiscale dilations and erosions comes from the work of Lax, where it is used to give stable and efficient numerical schemes for solving the equations [179]. Rouy and Tourin [240] have shown that the distance function to a shape is a viscosity solution of $1 - |Du| = 0$ with the null boundary condition (Dirichlet condition) on the boundary of the shape. To define efficient numerical schemes for computing the distance function, they actually implement the evolution equation $\partial u/\partial t = 1 - |Du|$ starting from zero and with the null boundary condition on the boundary of the shape. The fact that the multiscale dilations and erosions can be computed using the PDEs $\partial u/\partial t = \pm|Du|$ has been rediscovered or revived, thirty years after Lax's work, by several authors: Alvarez et. al. [11], van den Boomgaard and Smeulders [273], Maragos [193, 194]. See also [272] for a numerical review. For an implementation using curve evolution, see [248]. Curiously, the link between erosions, dilations, and their PDEs seems to have remained unknown or unexploited until 1992. The erosion and dilation PDEs can be used for shape thinning, which is a popular way to compute the skeleton. Pasquignon developed an erosion PDE with adaptive stopping time that allows one to compute directly a skeleton that does not look like barbed wire [226].

Chapter 10

Median Filters and Mathematical Morphology

This entire chapter is devoted to median filters. They are among the most characteristic and numerically efficient contrast-invariant monotone operators. The denoising effects of median filters are illustrated in Figures 10.1 and 10.2; the smoothing effect of a median filter is illustrated in Figure 10.3. They also are extremely useful in 3D-image or movie denoising.

As usual, there will be two associated operators, a set operator and a function operator. All of the median operators (or filters) will be defined in terms of a nonnegative measurable weight function $k : \mathbb{R}^N \to [0, +\infty)$ that is normalized:

$$\int_{\mathbb{R}^N} k(\mathbf{y})\,\mathrm{d}\mathbf{y} = 1.$$

The k-measure of a measurable subset $B \subset \mathbb{R}^N$ is denoted by $|B|_k$ and defined by

$$|B|_k = \int_B k(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int_{\mathbb{R}^N} k(\mathbf{y})\mathbf{1}_B(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

Clearly, $0 \le |B|_k \le 1$. The simplest example for k is given by the function $k = c_N^{-1}(r)\mathbf{1}_{B(0,r)}$, where $B(0, r)$ denotes the ball of radius r centered at the origin and $c_N(r)$ is the Lebesgue measure of $B(0, r)$. Another classical example to think of is the Gaussian.

10.1 Set and function medians

We first define the set operators, whose form is simpler. We define them on $\mathcal{M}(\mathbb{R}^N)$, the set of measurable subsets of $\mathbb{R}^N$ and then apply the standard extension to $\mathcal{M}(S_N)$ given in Definition 7.1.

Definition 10.1. *Let $X \in \mathcal{M}(\mathbb{R}^N)$ and let k be a weight function. The median set of X weighted by k is defined by*

$$\mathcal{M}\mathrm{ed}_k X = \{\mathbf{x} \mid |X - \mathbf{x}|_k \ge \tfrac{1}{2}\} \tag{10.1}$$

and its standard extension to $\mathcal{M}(S_N)$ by

$$\mathcal{M}\mathrm{ed}_k X = \{\mathbf{x} \mid |X - \mathbf{x}|_k \geq \tfrac{1}{2}\} \cup (X \cap \{\infty\}). \tag{10.2}$$

The extension amounts to add ∞ to $\mathcal{M}\mathrm{ed}_k X$ if ∞ belongs to X. Note that we have already encountered the median operator in Section 4.1. Koenderink and van Doorn defined the dynamic shape of X at scale t to be the set of $\mathbf{x}$ such that $G_t * \mathbf{1}_X(\mathbf{x}) \geq 1/2$. The dynamic shape is, in our terms, a Gaussian-weighted median filter.

To gain some intuition about median filters, we suggest considering the weight k defined on $\mathbb{R}^2$ by $k = (1/\pi r^2)\mathbf{1}_{B(0,r)}$. Then $\mathbf{x} \in \mathbb{R}^2$ belongs to $\mathcal{M}\mathrm{ed}_k X$ if and only if the Lebesgue measure of $X \cap B(\mathbf{x}, r)$ is greater than or equal to half the measure of $B(0,r)$. Thus, $\mathbf{x} \in \mathcal{M}\mathrm{ed}_k X$ if points of X are in the majority around $\mathbf{x}$.

Lemma 10.2. *$\mathcal{M}\mathrm{ed}_k$ is a standard monotone operator on $\mathcal{M}$.*

Proof. Obviously $\mathcal{M}\mathrm{ed}_k(\emptyset) = \emptyset$ and $\mathcal{M}\mathrm{ed}_k(S_N) = S_N$. By definition, $\infty \in \mathcal{M}\mathrm{ed}_k X \Leftrightarrow \infty \in X$. If X is bounded, it is a direct application of Lebesgue theorem that

$$|X - \mathbf{x}|_k = \int k(\mathbf{y})\mathbf{1}_{X-\mathbf{x}}(\mathbf{y})d\mathbf{y} \to 0 \text{ as } \mathbf{x} \to \infty.$$

Thus $|X - \mathbf{x}|_k < \frac{1}{2}$ for $\mathbf{x}$ large enough and $\mathcal{M}\mathrm{ed}_k X$ is therefore bounded. In the same way, if X^c is bounded $|X - \mathbf{x}|_k \to 1$ as $\mathbf{x} \to \infty$ and therefore $(\mathcal{M}\mathrm{ed}_X)^c$ is bounded. □

Lemma 10.3. *We can represent $\mathcal{M}\mathrm{ed}_k$ by*

$$\mathcal{M}\mathrm{ed}_k X = \{\mathbf{x} \mid \mathbf{x} + B \subset X, \textit{ for some } B \in \mathcal{B}\}, \tag{10.3}$$

where $\mathcal{B} = \{B \mid |B|_k \geq \frac{1}{2}\}$ or $\mathcal{B} = \{B \mid |B|_k = \frac{1}{2}\}$.

Proof. By Lemma 10.2, $\mathcal{M}\mathrm{ed}_k$ is standard monotone and it is obviously translation invariant. So we can apply Theorem 8.2. The canonical set of structuring elements of $\mathcal{M}\mathrm{ed}_k$ is

$$\mathcal{B} = \{B \mid 0 \in \mathcal{M}\mathrm{ed}_k B\} = \{B \mid |B|_k \geq \frac{1}{2}\}.$$

The second set $\mathcal{B}$ mentioned in the lemma, which we call now for convenience $\mathcal{B}'$, is a subset of $\mathcal{B}$ such that for every $B \in \mathcal{B}$, there is some $B' \in \mathcal{B}'$ such that $B' \subset B$. Thus by Proposition 8.4, $\mathcal{M}\mathrm{ed}_k$ can be defined from $\mathcal{B}'$. □

The next lemma will help defining the function operator Med_k associated with the set operator $\mathcal{M}\mathrm{ed}_k$.

Lemma 10.4. *The set operator $\mathcal{M}\mathrm{ed}_k$ is monotone, translation invariant and upper semicontinuous on $\mathcal{M}$.*

Proof. The first two properties are straightforward. Consider a nonincreasing sequence $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$ and let us show that

$$\mathcal{M}\mathrm{ed}_k \bigcap_{n\in\mathbb{N}} X_n = \bigcap_{n\in\mathbb{N}} \mathcal{M}\mathrm{ed}_k X_n.$$

Since $\mathcal{M}\mathrm{ed}_k$ is monotone, it is always true that $\mathcal{M}\mathrm{ed}_k \bigcap_{n\in\mathbb{N}} X_n \subset \bigcap_{n\in\mathbb{N}} \mathcal{M}\mathrm{ed}_k X_n$. To prove the other inclusion, assume that $\mathbf{x} \in \bigcap_{n\in\mathbb{N}} \mathcal{M}\mathrm{ed}_k X_n$. If $\mathbf{x} \in \mathbb{R}^N$, by the definition of $\mathcal{M}\mathrm{ed}_k$, $|X_n - \mathbf{x}| \geq 1/2$ for all $n \in \mathbb{N}$. Since $X_n - \mathbf{x} \downarrow \bigcap_{n\in\mathbb{N}}(X_n - \mathbf{x})$, we deduce from Lebesgue Theorem that $|X_n - \mathbf{x}|_k \downarrow |\bigcap_{n\in\mathbb{N}}(X_n - \mathbf{x})|_k$. This means that $|\bigcap_{n\in\mathbb{N}}(X_n - \mathbf{x})|_k \geq 1/2$, and hence that $\mathbf{x} \in \mathcal{M}\mathrm{ed}_k \bigcap_{n\in\mathbb{N}}(X_n - \mathbf{x})$. If $\mathbf{x} = \infty$, it belongs to $\mathcal{M}\mathrm{ed}_k X_n$ for all n and therefore to X_n for all n. Thus, it belongs to $\bigcap_{n\in\mathbb{N}} X_n$ and therefore to $\mathcal{M}\mathrm{ed}_k(\bigcap_{n\in\mathbb{N}} X_n)$. □

Definition 10.5 (and proposition). *Define the function operator* Med_k *from* $\mathcal{M}\mathrm{ed}_k$ *as a stack filter,*

$$\mathrm{Med}_k u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda u\}.$$

Then Med_k *is standard monotone, contrast invariant and translation invariant from* $\mathcal{F}$ *to* $\mathcal{F}$*.* Med_k *and* $\mathcal{M}\mathrm{ed}_k$ *commute with thresholds,*

$$\mathcal{X}_\lambda \mathrm{Med}_k u = \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda u. \tag{10.4}$$

If k is radial, Med_k *therefore is SMTCII.*

Proof. By Lemma 10.4, $\mathcal{M}\mathrm{ed}_k$ is upper semicontinuous and by Lemma 10.2 it is standard monotone. It also is translation invariant. So we can apply Theorem 7.16, which yields all announced properties for Med_k. □

We get a sup-inf formula for the median as a direct application of Theorem 8.6.

Proposition 10.6. *The median operator* Med_k *has the sup-inf representation*

$$\mathrm{Med}_k u(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}), \tag{10.5}$$

where $\mathcal{B} = \{B \mid B \in \mathcal{M}, |B|_k = 1/2\}$.

A median value is a kind of average, but with quite different results, as is illustrated in Exercise 10.4.

10.2 Self-dual median filters

The median operator Med_k, as defined, is not invariant under "reverse contrast," that is, it does not satisfy $-\mathrm{Med}_k u = \mathrm{Med}_k(-u)$ for all $u \in \mathcal{F}$. This is clear from the example in the next exercise. Self-duality is a conservative requirement which is true for all linear filters. It means that the white and black balance is respected by the operator. We have seen that dilations favor whites and erosions favor black colors: These operators are not self-dual.

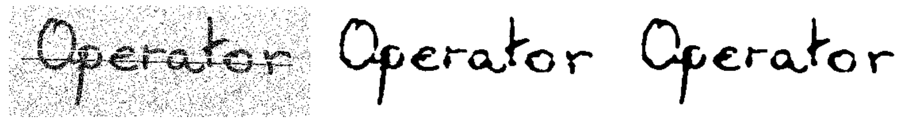

Figure 10.1: Example of denoising with a median filter. Left to right: scanned picture of the word "operator" with perturbations and noise made with black or white lines and dots; the image after one application of a median filter with a circular neighborhood of radius 2; the image after a second application of the same filter. Compare with the denoising using openings and closings (Figure 9.8).

Exercise 10.1. Consider the one-dimensional median filter with $k = \frac{1}{2}\mathbf{1}_{[-2,-1]\cup[1,2]}$. Let $u(x) = -1$ if $x \leq -1$, $u(x) = 1$ if $x \geq 1$, $u(x) = x$ elsewhere. Check that $\mathrm{Med}_k u(0) \neq -\mathrm{Med}(-u)(0)$. ■

As we did with erosions and dilations, one can define a dual version of the median Med_k^- by

$$\mathrm{Med}_k^- u = -\mathrm{Med}_k(-u), \text{ so that} \tag{10.6}$$

$$\mathrm{Med}_k^- u(\mathbf{x}) = \inf_{|B|_k \geq \frac{1}{2}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}). \tag{10.7}$$

A quite general condition on k is sufficient to guarantee that Med_k and Med_k^- agree on continuous functions.

Definition 10.7. *We say that k is not separable if $|B|_k \geq 1/2$ and $|B'|_k \geq 1/2$ imply that $\overline{B} \cap \overline{B'} \neq \emptyset$.*

Proposition 10.8.

(*i*) *For every measurable function u, $\mathrm{Med}_k u \geq \mathrm{Med}_k^- u$.*

(*ii*) *Assume that k is not separable. Then for every $u \in \mathcal{F}$, $\mathrm{Med}_k u = \mathrm{Med}_k^- u$ and Med_k is self-dual.*

Proof. Both operators are translation invariant, so without loss of generality we may assume that $\mathbf{x} = 0$. To prove (i), let $\lambda = \mathrm{Med}_k u(0) = \sup_{|B|_k \geq 1/2} \inf_{y\in B} u(\mathbf{y})$. Take $\varepsilon > 0$ and consider the level set $\mathcal{X}_{\lambda+\varepsilon} u$. Then $\inf_{\mathbf{y}\in\mathcal{X}_{\lambda+\varepsilon}} u(\mathbf{y}) \geq \lambda+\varepsilon$. Thus $|\mathcal{X}_{\lambda+\varepsilon} u|_k < 1/2$, since $\inf_{\mathbf{y}\in B} \leq \lambda$ for any set B such that $|B| \geq 1/2$. Thence $|(\mathcal{X}_{\lambda+\varepsilon} u)^c|_k \geq 1/2$. By the definition of level sets, $\sup_{\mathbf{y}\in(\mathcal{X}_{\lambda+\varepsilon} u)^c} u(\mathbf{y}) \leq \lambda + \varepsilon$. These two last relations imply that

$$\inf_{|B|_k \geq \frac{1}{2}} \sup_{\mathbf{y}\in B} u(\mathbf{y}) \leq \lambda + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this proves (i).

The assumption that k is not separable implies that for all B and B' having k-measure greater than or equal to $1/2$, we have $\inf_{\mathbf{y}\in\overline{B}} u(\mathbf{y}) \leq \sup_{\mathbf{y}\in\overline{B'}} u(\mathbf{y})$. Since $u \in \mathcal{F}$ is continuous, $\inf_{\mathbf{y}\in B} u(\mathbf{y}) \leq \sup_{\mathbf{y}\in B'} u(\mathbf{y})$. Since B and B' were arbitrary except for the conditions $|B|_k \geq 1/2$ and $|B'|_k \geq 1/2$, the last inequality implies that

$$\sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y}\in B} u(\mathbf{y}) \leq \inf_{|B'|_k \geq \frac{1}{2}} \sup_{\mathbf{y}\in B'} u(\mathbf{y}).$$

Figure 10.2: Denoising based on a median filter. Left: an image altered on 40% of its pixels with salt and pepper noise. Right: the same image after three iterations of a median filter with a 3×3 square mask.

From this last inequality and (i), we conclude that $\mathrm{Med}_k u = \mathrm{Med}_k^- u$. □

10.3 Discrete median filters and the "usual" median value

We define a discrete median filter by considering, instead of a function, a uniform discrete measure $k = \sum_{i=1,\dots,N} \delta_{\mathbf{x}_i}$, where $\delta_{\mathbf{x}_i}$ denotes the Dirac mass at $\mathbf{x}_i$. We could normalize k, but this is not necessary, as will become clear. Translates of the points $\mathbf{x}_i$ create the discrete neighborhood that is used to compute the median value of a function u at a point $\mathbf{x}$. We denote the set of subsets of $\{1,\dots,N\}$ by $\mathcal{P}(N)$ and the number of elements in $P \in \mathcal{P}(N)$ by $\mathrm{card}(P)$. Since $\mathrm{card}(P) = |P|_k$, we will suppress the k-notation is favor of the more transparent "$\mathrm{card}(P)$," but one should remember that the k-measure is still there. An immediate generalization of the definition of the median filters to the case where k is such a discrete measure yields

$$\mathrm{Med}u(\mathbf{x}) = \sup_{\substack{P \in \mathcal{P}(N) \\ \mathrm{card}(P) \geq N/2}} \inf_{i \in P} u(\mathbf{x} - \mathbf{x}_i),$$

$$\mathrm{Med}^- u(\mathbf{x}) = \inf_{\substack{P \in \mathcal{P}(N) \\ \mathrm{card}(P) \geq N/2}} \sup_{i \in P} u(\mathbf{x} - \mathbf{x}_i).$$

When k was continuous, we could replace "$|B|_k \geq 1/2$" with "$|B|_k = 1/2$," but this is not directly possible in the discrete case, since $N/2$ is not an integer if N is odd. To fix this, we define the function M by $M(N) = N/2$ if N is even and $M(N) = (N/2) + (1/2)$ if N is odd. Now we have

$$\begin{aligned}
\mathrm{Med}u(\mathbf{x}) &= \sup_{\substack{P\in\mathcal{P}(N)\\ \mathrm{card}(P)=M(N)}} \inf_{i\in P} u(\mathbf{x}-\mathbf{x}_i),\\
\mathrm{Med}^- u(\mathbf{x}) &= \inf_{\substack{P\in\mathcal{P}(N)\\ \mathrm{card}(P)=M(N)}} \sup_{i\in P} u(\mathbf{x}-\mathbf{x}_i).
\end{aligned}$$

The fact that we can replace "card$(P) \geq N/2$" with "card$(P) = M(N)$" has been argued elsewhere for the continuous case; for the discrete case, it is a matter of simple combinatorics. Given any $\mathbf{x}$, let $y_i = u(\mathbf{x}-\mathbf{x}_i)$. After a suitable permutation of the i's, we can order the y_i as follows: $y_1 \leq, \cdots, \leq y_M \leq, \cdots \leq y_N$. Then for N even,

$$\begin{aligned}
\{\inf_{i\in P} y_i \mid \mathrm{card}(P) \geq N/2\} &= \{\inf_{i\in P} y_i \mid \mathrm{card}(P) = M\} = \{y_1, \ldots, y_{M+1}\},\\
\{\sup_{i\in P} y_i \mid \mathrm{card}(P) \geq N/2\} &= \{\sup_{i\in P} y_i \mid \mathrm{card}(P) = M\} = \{y_M, \ldots, y_N\},
\end{aligned}$$

and $\mathrm{Med}u(\mathbf{x}) = y_{M+1} \geq y_M = \mathrm{Med}^- u(\mathbf{x})$. If N is odd, we have

$$\begin{aligned}
\{\inf_{i\in P} y_i \mid \mathrm{card}(P) \geq N/2\} &= \{\inf_{i\in P} y_i \mid \mathrm{card}(P) = M\} = \{y_1, \ldots, y_M\},\\
\{\sup_{i\in P} y_i \mid \mathrm{card}(P) \geq N/2\} &= \{\sup_{i\in P} y_i \mid \mathrm{card}(P) = M\} = \{y_M, \ldots, y_N\},
\end{aligned}$$

and $\mathrm{Med}u(\mathbf{x}) = \mathrm{Med}^- u(\mathbf{x}) = y_M$. This shows that $\mathrm{Med} = \mathrm{Med}^-$ if and only if N is odd. What we see here is the discrete version of Proposition 10.8. When N is odd, the measure is not separable, since two sets P and P' with $\mathrm{card}(P) \geq N/2$ and $\mathrm{card}(P') \geq N/2$ always have a nonempty intersection. In general, a median filter with an odd number of pixels is preferred, since $\mathrm{Med} = \mathrm{Med}^-$ in this case.

This discussion shows that the definition of the discrete median filter Med corresponds to the usual statistical definition of the median of a set of data: If the given data consists of the numbers $y_1 \leq y_2 \leq \cdots \leq y_N$ and $N = 2n+1$, them by definition, the median is y_{n+1}. In case $N = 2n$, the median is $(y_n + y_{n+1})/2$. In both cases, half of the terms are greater than or equal to the median and half of the terms are less than or equal to the median. The usual median minimizes the functional $\sum_{i=1}^{N} |y_i - y|$. Exercise 10.9 shows how Med and Med^- relate to this functional.

The discrete median filters can also be defined in terms of a nonuniform measure k that places different weights on the points $\mathbf{x}_i$. To see what this does, assume that the weights are integers k_i, so $|\{x_i\}|_k = k_i$. Then k has total mass $\sum_{i=1}^{N} k_i = K$, and the condition $\mathrm{card}(P) \geq N/2$ is replaced with $|P|_k \geq K/2$. As before, let $y_i = u(\mathbf{x}-\mathbf{x}_i)$ and display the data set as $y_1 \leq y_2 \leq \cdots \leq y_N$. Then $\mathrm{Med}_k u(\mathbf{x}) = y_j$, where j is the largest index such that $k_j + \cdots + k_N \geq N/2$. To see this, transform the original ordered sequence into the expanded ordered sequence

$$\underbrace{y_1 = \cdots = y_1}_{k_1 \text{ terms}} \leq \cdots \leq \underbrace{y_i = \cdots = y_i}_{k_i \text{ terms}} \leq \cdots \leq \underbrace{y_N = \cdots = y_N}_{k_N \text{ terms}}. \tag{10.8}$$

Then by the definition of j, $y_j \in \{\inf_{i\in P} y_i \mid |P|_k \geq K/2\}$, but y_i for $i > j$ is not in this set. Thus, $\mathrm{Med}_k u(\mathbf{x}) = y_j$. Conversely, if $\mathrm{Med}_k u(\mathbf{x}) = y_j$, then y_j

is the largest member of the set $\{\inf_{i\in P} y_i \mid |P|_k \geq K/2\}$. This implies that $k_j + \cdots + k_N \geq N/2$, but that $k_i + \cdots + k_N < K/2$ for $i > j$.

If K is odd, what we have just done implies that $\mathrm{Med}_k^- = \mathrm{Med}_k$, and that $\mathrm{Med}_k u(\mathbf{x})$ is equal to the ordinary median of the ordered set (10.8). Exercise 10.8 completes this part of the theory.

Finally, we wish to show that the discrete median filter Med can be a cyclic operator on discrete images. As a simple example, consider the chessboard image, where $u(i,j) = 255$ if $i+j$ is even and $u(i,j) = 0$ otherwise. When we apply the median filter that takes the median of the four values surrounding a pixel and the pixel value, it is clear that the filter "reverses" the chessboard pattern. Indeed, any white pixel (value 255) is surrounded by four black pixels (value zero), so the median filter transforms the white pixel into a black pixel. In the same way, a black pixel is transformed into a white pixel and this can go for ever.

10.4 Exercises

Exercise 10.2. Check that $\mathcal{M}\mathrm{ed}_k$ as defined in Definition 10.1 is monotone and translation invariant. ■

Exercise 10.3. Koenderink and van Doorn defined the dynamic shape of X at scale t to be the set of $\mathbf{x}$ such that $G_t * \mathbf{1}_X(\mathbf{x}) \geq 1/2$. Check that this is a Gaussian-weighted median filter. ■

Exercise 10.4. Consider the weighted median filter defined on S_1 with $k = (1/2)\mathbf{1}_{[-1,1]}$. Compute $\mathrm{Med}_k u$ for $u(x) = \frac{1}{1+x^2}$. Compare the result with the local average $M_1 u(x) = \frac{1}{2}\int_{-1}^{1} u(x+y)dy$. What happens on intervals where u is monotone? ■

Exercise 10.5. Saying that k is not separable is a fairly weak assumption. It corresponds roughly to saying that the support of k cannot be split into two disjoint connected components each having k-measure 1/2. Show that if k is continuous and if its support is connected, then it is not separable. ■

Exercise 10.6. Prove the following inequalities for any measurable function :

$$\sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \sup_{|B|_k > \frac{1}{2}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \inf_{|B|_k \geq \frac{1}{2}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}),$$

$$\sup_{|B|_k \geq \frac{1}{2}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \inf_{|B|_k > \frac{1}{2}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \geq \inf_{|B|_k \geq \frac{1}{2}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}).$$

■

Exercise 10.7. Median filter on measurable sets and functions. The aim of the exercise is to study the properties of the median filter extended to the set $\mathcal{M}$ of all measurable sets of S_N and all bounded measurable functions ($u \in L^\infty(S_N)$). The definition of $\mathcal{M}\mathrm{ed}_k$ on $\mathcal{M}$ is identical to the current definition.

1) Using the result of Exercise 7.20, show that one can define Med_k from $\mathcal{M}\mathrm{ed}_k$ as a stack filter and that it is monotone, translation and contrast invariant. In addition, $\mathcal{M}\mathrm{ed}_k$ and Med_k still satisfy the commutation with thresholds, $\mathcal{X}_\lambda \mathrm{Med}_k u = \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda u$.

2) Prove that $\mathcal{M}\mathrm{ed}_k$ maps measurable sets into closed sets. Deduce that if u is a measurable function, then $\mathrm{Med}_k u$ is upper semicontinuous and $\mathrm{Med}_k^- u$ is lower semicontinuous.

3) Assume that k is not separable. Check that the proof of Proposition 10.8 still applies to the more general Med_k and Med_k^-, applied to all measurable functions. Deduce that if k is not separable, then $\mathcal{M}\mathrm{ed}_k u$ is continuous whenever u is a measurable function. ■

Exercise 10.8. Let us consider a discrete nonuniform weight distribution k. Check that $\mathrm{Med}_k^- u \le \mathrm{Med}_k u$. Prove that $\mathrm{Med}_k^- u = \mathrm{Med}_k u$ if and only if there is no subset of the numbers $k_1, \dots, k_N$ whose sum is $K/2$. In particular, if K is odd, then $\mathrm{Med}_k^- u = \mathrm{Med}_k u$. ∎

Exercise 10.9. Variational interpretations of the median and the average values.
Let $\mathrm{arginf}_m g(m)$ denote the value of m, if it exists, at which g attains its infimum. Consider N real numbers $\{\mathbf{x}_i \mid i = 1, 2, \dots, N\}$ and denote by $\mathrm{Med}((x_i)_i)$ and $\mathrm{Med}^-((x_i)_i)$ their usual lower and upper median values (we already know that both are equal if N is odd but can be different if N is even).

(i) Show that

$$\frac{1}{N}\sum_{i=1}^{N} x_i = \mathrm{arginf}_m \sum_{i=1}^{N}(x_i - m)^2.$$

(ii) Show that

$$\mathrm{Med}^-((x_i)_i) \le \mathrm{arginf}_m \sum_{i=1}^{N} |x_i - m| \le \mathrm{Med}((x_i)_i).$$

(iii) Let $k = \mathbf{1}_B$, where B is set with Lebesgue measure equal to one. Let $\mathrm{Med}_B u$ denote the *median value of u in B*, defined by $\mathrm{Med}_B u = \mathrm{Med}_k u(0)$. Consider a bounded measurable function u defined on B. Show that

$$\int_B u(\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathrm{arginf}_m \int_B (u(\mathbf{x}) - m)^2\, \mathrm{d}\mathbf{x}$$

and that

$$\mathrm{Med}_B^- u \le \mathrm{arginf}_m \int_B |u(\mathbf{x}) - m|\, \mathrm{d}\mathbf{x} = \frac{\mathrm{Med}_B^- u + \mathrm{Med}_B u}{2} \le \mathrm{Med}_B u.$$

(iv) Deduce from the above that the mean value is the best constant approximation in the L^2 norm and that the median is the best constant approximation in the L^1 norm. ∎

10.5 Comments and references

The remarkable denoising properties and numerical efficiency of median filters for the removal of all kinds of impulse noise in digital images, movies, and video signals are well known and acclaimed [86, 156, 215, 230, 235]. The last reference cited as well as the next three all propose simple and efficient implementations of the median filter [26, 84, 145]. An introduction to weighted median filters can be found in [47, 290], and information about some generalizations (conditional median filters, for example) can be found in [24, 180, 263]. The min, max, and median filters are particular instances of rank order filters; see [80] for a general presentation of these filters. There are few studies on iterated median filters. The use of iterated median filters as a scale space is, however, proposed in [32]. The extension of median filtering to multichannel (color) images is problematic, although there have been some interesting attempts [65, 236].

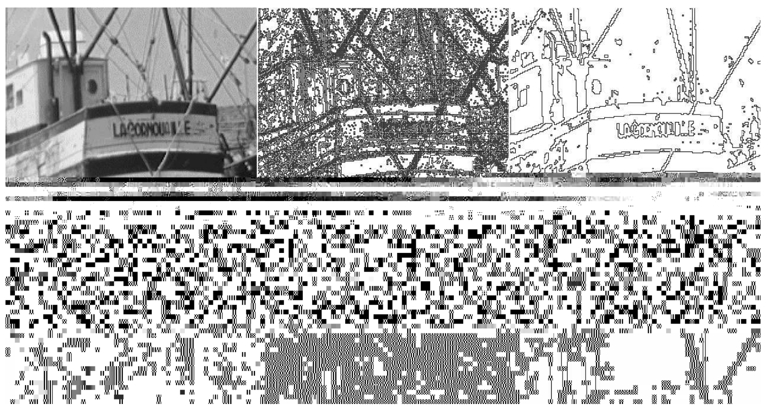

Figure 10.3: Smoothing effect of a median filter on level lines. Above, left to right: original image; all of its level lines (boundaries of level sets) with levels multiple of 12; level lines at level 100. Below,left to right: result of two iterations of a median filter with a disk with radius 2; corresponding level lines (levels multiple of 12); level lines at level 100.

Part III

Local Asymptotic Analysis of Operators

Chapter 11

Curves and Curvatures

This chapter contains the fundamentals of differential geometry that are used in the book. Our main aim is to define the orientation and curvatures of a curve or a surface as the main contrast invariant differential operators we shall deal with in image and curve smoothing.

11.1 Tangent, normal, and curvature

We summarize in this section the concepts and results about smooth curves that are needed in this chapter and elsewhere in the book. The curves we considered will always be plane curves.

Definition 11.1. *We call simple arc or Jordan arc the image* Γ *of a continuous one-to-one function* $\mathbf{x} : [0,1] \to \mathbb{R}^2$, $\mathbf{x}(t) = (x(t), y(t))$. *We say that* Γ *is a simple closed curve or Jordan curve if the mapping restricted to* $(0,1)$ *is one-to-one and if* $\mathbf{x}(0) = \mathbf{x}(1)$. *If* $\mathbf{x}$ *is continuously differentiable on* $[0,1]$, *we define the arc length of the segment of the curve between* $\mathbf{x}(t_0)$ *and* $\mathbf{x}(t)$ *by*

$$L(\mathbf{x}, t_0, t) = \int_{t_0}^{t} |\mathbf{x}'(\tau)| \,\mathrm{d}\tau = \int_{t_0}^{t} \sqrt{\mathbf{x}'(\tau) \cdot \mathbf{x}'(\tau)} \,\mathrm{d}\tau. \tag{11.1}$$

In particular, set

$$L(t) = L(\mathbf{x}, 0, t) = \int_{0}^{t} |\mathbf{x}'(\tau)| \,\mathrm{d}\tau = \int_{0}^{t} \sqrt{\mathbf{x}'(\tau) \cdot \mathbf{x}'(\tau)} \,\mathrm{d}\tau.$$

The curves we deal with will always be smooth. Now, we want the definition of "smoothness" to describe an intrinsic property of Γ rather than a property of some parameterization $\mathbf{x}(s)$ of Γ. If a function $\mathbf{x}$ representing Γ is C^1, then the function L in equation (11.1) has a derivative with respect to s,

$$L'(t) = |\mathbf{x}'(t)|$$

that is continuous. Nevertheless, the curve itself may not conform to our idea of being smooth, which at a minimum requires a tangent at every point $\mathbf{y} \in \Gamma$. For example, the motion of a point on the boundary of a unit disk as it rolls along the x-axis is described by $\mathbf{x}(t) = (t - \sin t, 1 - \cos t)$, which is a C^∞ function.

Nevertheless, the curve has cusps at all multiples of 2π. The problem is that $\mathbf{x}'(2k\pi) = 0$. It is this sort of phenomenon that motivates the definition of smoothness for curves.

Definition 11.2. *We say that a curve Γ admits an arc-length parameterization $s \in \mathbb{R} \mapsto \mathbf{x}(s)$ if the function $\mathbf{x}$ is C^1 and $L'(s) = |\mathbf{x}'(s)| = 1$ for all s. We say that Γ is C^m, $m \in \mathbb{N}$, $m \geq 1$, if the arc-length parameterization $\mathbf{x}$ is a C^m function.*

Exercise 11.1. The aim of the exercise is to give a formula transforming a C^1 parameterization $t \in [0,1] \to \mathbf{x}(t)$ such that $|\mathbf{x}'(t)| \neq 0$ for all t into an arc-length parameterization. Notice that $L : [0,1] \to [0, L(1)]$ is increasing. Set, for $s \in [0, L(1)]$, $\tilde{\mathbf{x}}(s) = \mathbf{x}(L^{-1}(s))$ and check that $\tilde{\mathbf{x}}$ is an arc-length parameterization of the curve defined by $\mathbf{x}$. ■

An arc-length parameterization is also called a *Euclidean parameterization*. If a Jordan curve has an arc-length parameterization $\mathbf{x}$, then the domain of definition of $\mathbf{x}$ on the real line must be an interval $[a, b]$, where $b - a$ is the length of Γ, which we denote by $l(\Gamma)$. In this case, we will always take $[0, l(\Gamma)]$ as the domain of definition of $\mathbf{x}$. We identify $[0, l(\Gamma)]$ algebraically with the circle group by adding elements of $[0, l(\Gamma)]$ modulo $l(\Gamma)$.

Definition 11.3. *Assume that Γ is C^2 and let $s \mapsto \mathbf{x}(s)$ be an arc-length parameterization. The tangent vector $\boldsymbol{\tau}$ is defined as $\boldsymbol{\tau} = \mathrm{d}\mathbf{x}/\mathrm{d}s$. The curvature vector of the curve Γ is defined by $\boldsymbol{\kappa} = \mathrm{d}^2\mathbf{x}/\mathrm{d}s^2$. The normal vector $\mathbf{n}$ is defined by $\mathbf{n} = \boldsymbol{\tau}^{\perp}$, where $(x, y)^{\perp} = (-y, x)$.*

One can easily describe all Euclidean parameterizations of a Jordan curve.

Proposition 11.4. *Suppose that Γ is a C^1 Jordan curve with arc-length parameterization $\mathbf{x} : [0, l(\Gamma)] \to \Gamma$. Then any other arc-length parameterization $\mathbf{y} : [0, l(\Gamma)] \to \Gamma$ is of the form $\mathbf{y}(s) = \mathbf{x}(s + \sigma)$ or $\mathbf{y}(s) = \mathbf{x}(-s + \sigma)$ for some $\sigma \in [0, l(\Gamma)]$.*

Proof. Denote by C the interval $[0, l(\Gamma)]$, defined as an additive subgroup of $\mathbb{R}$ modulo $l(\Gamma)$. Let $\mathbf{x},\ \mathbf{y} : C \mapsto \Gamma$ be two length preserving parameterizations of Γ. Then $f = \mathbf{x} \circ \mathbf{y}^{-1}$ is a length preserving bijection of C. Using the parameterization of C, this implies $f(s) = \pm s + \sigma$ for some $\sigma \in [0, l(\Gamma)]$ and the proof is easily concluded. (See exercise 11.7 for some more details.) □

Proposition 11.5. *Let Γ be a C^2 Jordan curve, and let $\mathbf{x}$ and $\mathbf{y}$ by any two arc-length parameterizations of Γ.*

(*i*) *If $\mathbf{x}(s) = \mathbf{y}(t)$, then $\mathbf{x}'(s) = \pm \mathbf{y}'(t)$.*

(*ii*) *The vector $\boldsymbol{\kappa}$ is independent of the choice of arc-length parameterizations and it is orthogonal to $\boldsymbol{\tau} = \mathbf{x}'$.*

Proof. By Proposition 11.4, $\mathbf{y}(s) = \mathbf{x}(\pm s + \sigma)$ and (*i*) follows by differentiation. This is also geometrically obvious: $\mathbf{x}'(s)$ and $\mathbf{y}'(t)$ are unit vectors tangent to Γ

at the same point. Thus, they either point in the same direction or they point in opposite directions.

Using any of the above representations and differentiating twice shows that $\mathbf{x}'' = \mathbf{y}''$. Since $\mathbf{x}' \cdot \mathbf{x}' = 1$, differentiating this expression shows that $\mathbf{x}'' \cdot \mathbf{x}' = 0$. Thus, $\mathbf{x}''$ and $\mathbf{x}'$ are orthogonal and $\mathbf{x}''$ and $\mathbf{x}'^{\perp}$ are collinear. □

It will be convenient to have a flexible notation for the curvature in the different contexts we will use it. This is the object of the next definition.

Definition 11.6 (and notation). *Given a C^2 curve Γ, which is parameterized by length as $s \mapsto \mathbf{x}(s)$ and $\mathbf{x} = \mathbf{x}(s)$ a point of Γ, we denote in three equivalent ways the curvature of Γ at $\mathbf{x} = \mathbf{x}(s)$,*

$$\boldsymbol{\kappa}(\Gamma)(\mathbf{x}) = \boldsymbol{\kappa}(\mathbf{x}) = \boldsymbol{\kappa}(\mathbf{x}(s)) = \boldsymbol{\kappa}(s) = \mathbf{x}''(s).$$

In the first notation, $\boldsymbol{\kappa}$ is the curvature of the curve Γ at a point $\mathbf{x}$ implicitly supposed to belong Γ. In the second notation, Γ is omitted. In the third notation a particular parameterization of Γ, $\mathbf{x}(s)$, is being used. In the fourth one, $\mathbf{x}$ is omitted.

The above notations create no ambiguity or contradiction, since by Proposition 11.5 the curvature is independent of the Euclidean parameterization. Of course, a smooth Jordan curve is locally a graph. More specifically:

Proposition 11.7. *A C^1 Jordan arc Γ can be represented around each one of its points $\mathbf{x}_0$ as the graph of a C^1 scalar function $y = f(x)$ such that $\mathbf{x}_0 = (0, f(0)) = (0, 0)$, $f'(0) = 0$, and*

$$\boldsymbol{\kappa}(\mathbf{x}_0) = (0, f''(0)). \tag{11.2}$$

Conversely, the graph of any C^1 function f is a C^1 Jordan arc. If f is C^2 the curvature of its associated Jordan curve satisfies (11.2) *at each point where $f'(0) = 0$.*

Proof. Assume we are given a C^1 Jordan arc Γ and an arc-length parameterization $\mathbf{c}$ in a neighborhood of $\mathbf{x}_0 = \mathbf{c}(s_0) \in \Gamma$. We assume, without loss of generality, that $s_0 = 0$. Then we can establish a local coordinate system with origin $\mathbf{x}_0$ and based on the two unit vectors $\mathbf{c}'(0)$ and $\mathbf{c}'(0)^{\perp}$ where the x-axis is positive in the direction of $\mathbf{c}'(0)$. If we write $\mathbf{c}(s) = (x(s), y(s))$ in this coordinate system, then

$$x(s) = \mathbf{c}(s) \cdot \mathbf{c}'(0),$$
$$y(s) = \mathbf{c}(s) \cdot \mathbf{c}'(0)^{\perp}.$$

Since $\mathrm{d}x/\mathrm{d}s(s) = \mathbf{c}'(s) \cdot \mathbf{c}'(0)$, $\mathrm{d}x/\mathrm{d}s(0) = 1$. Then the inverse function theorem implies the existence of a C^1 function g and a $\delta > 0$ such that $s = g(x)$ for $|x| < \delta$. This means that, for $|x| < \delta$, Γ is represented locally by the graph of the C^1 function f, where $f(x) = y(g(x)) = \mathbf{c}(g(x)) \cdot \mathbf{c}'(0)^{\perp}$. To be slightly more precise, denote the graph of f for $|x| < \delta$ by Γ_f. Since g is one-to-one, Γ_f is a homeomorphic image of the open interval $(-\delta, \delta)$ and $\Gamma_f \subset \Gamma$. If Γ is C^2, then

f is C^2 and $f''(0) = \mathbf{c}''(0) \cdot \mathbf{c}'(0)^\perp$. Thus, on the local coordinate system, the coordinates of $c''(0) = \boldsymbol{\kappa}(x_0)$ are $(0, f''(0))$.

Conversely, given a C^1 function f, we can consider the graph Γ_f of f in a neighborhood of the origin. Then Γ_f is represented by $\mathbf{c}$, where $\mathbf{c}(x) = (x, f(x))$. We may assume that $f(0) = 0$ and $f'(0) = 0$ (by a translation and rotation if necessary). The arc-length along Γ is measured by

$$s(x) = \int_0^x \sqrt{1 + [f'(t)]^2}\, \mathrm{d}t,$$

and $\mathrm{d}s/\mathrm{d}x = \sqrt{1 + [f'(x)]^2}$. This time there is a C^1 function h such that $h(s) = x$ and $h'(s) = (1 + [f'(h(s))]^2)^{-1/2}$. Then Γ is represented by $\tilde{\mathbf{c}}(s) = (h(s), f(h(s)))$. Short computations show that $|\tilde{\mathbf{c}}'(s)| = 1$. If in addition f is C^2, then Γ is C^2 and it is an easy check that $\tilde{\mathbf{c}}''(0) \cdot \tilde{\mathbf{c}}'(0)^\perp = f''(0)$. □

Exercise 11.2. Make the above "short computations" and the "easy check". ■

11.2 The structure of the set of level lines

We saw in Chapter 5 how an image can be represented by its level sets. The next step, with a view toward shape analysis, is the representation of an image in terms of its level lines. We rely heavily on the implicit function theorem to develop this representation. We begin with a two-dimensional version. The statement here is just a slight variation on the implicit function theorem quoted in section I.4.

Theorem 11.8. *Let $u \in \mathcal{F}$ be a C^1 function such that $Du(\mathbf{x}_0) \neq 0$ at some $\mathbf{x}_0 = (x_0, y_0)$. Let $\boldsymbol{i}$ denote the unit vector in the direction (u_x, u_y), let $\boldsymbol{j}$ denote the unit vector in the orthogonal direction $(-u_y, u_x)$, and write $\mathbf{x} = \mathbf{x}_0 + x\boldsymbol{i} + y\boldsymbol{j}$. Then there is a disk $D(\mathbf{x}_0, r)$ and a unique C^1 function φ, $\varphi : [-r, r] \to \mathbb{R}$, such that if $\mathbf{x} \in D(\mathbf{x}_0, r)$, then*

$$u(x, y) = 0 \quad \Longleftrightarrow \quad x = \varphi(y).$$

The following corollary is a global version of this local result.

Corollary 11.9. *Assume that $u \in \mathcal{F}$ is C^1 and let $u^{-1}(\lambda) = \{\mathbf{x} \mid u(\mathbf{x}) = \lambda\}$ for $\lambda \in \mathbb{R}$. If $\lambda \neq u(\infty)$ and $Du(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in u^{-1}(\lambda)$, then $u^{-1}(\lambda)$ is a finite union of disjoint Jordan curves.*

Proof. From Theorem 11.8 we know that for each point $\mathbf{x} \in u^{-1}(\lambda)$ there is an open disk $D(\mathbf{x}, r(\mathbf{x}))$ such that $\overline{D}(\mathbf{x}, r(\mathbf{x})) \cap u^{-1}(\lambda)$ is a C^1 Jordan arc $\mathbf{x}(s)$ and we can take the endpoints of the arc on $\partial D(\mathbf{x}, r(\mathbf{x}))$. Since $\lambda \neq u(\infty)$, $u^{-1}(\lambda)$ is compact. Thus there is a finite number of points $\mathbf{x}_i$, $i = 1, \ldots, m$, such that $u^{-1}(\lambda) \subset \bigcup_{i=1}^m D(\mathbf{x}_i, r(\mathbf{x}_i))$. This implies that $u^{-1}(\lambda)$ is a finite union of Jordan arcs which we can parameterize by length. The rest of the proof is very intuitive and is left to the reader. I consists of iteratively gluing the Jordan arcs until they close up into one or several Jordan curves. □

The next theorem is one of the few results that we are going to quote rather than prove, as we have done with the implicit function theorem.

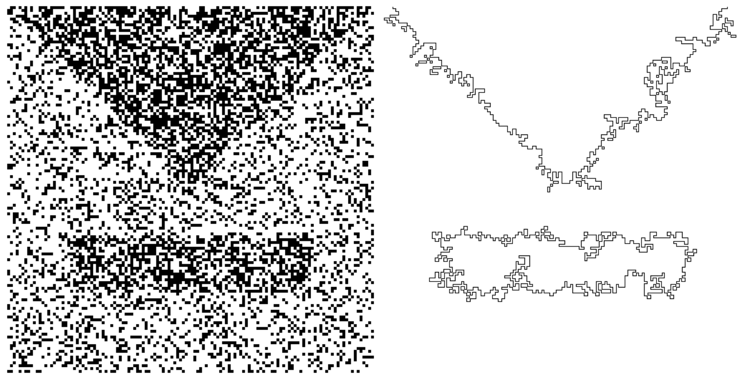

Figure 11.1: Level lines as representatives of the shapes present in an image. Left: noisy binary image with two apparent shapes; right: the two longest level lines.

Theorem 11.10 (Sard's theorem). *Let $u \in \mathcal{F} \cap C^1$. Then for almost every λ in the range of u, the set $u^{-1}(\lambda)$ is nonsingular, which means that for all $\mathbf{x} \in u^{-1}(\lambda)$, $Du(\mathbf{x}) \neq 0$.*

As a direct consequence of Sard's Theorem and Corollary 11.9, we obtain:

Corollary 11.11. *Let $u \in \mathcal{F} \cap C^1$. Then for almost every λ in the range of u, the set $u^{-1}(\lambda)$ is the union of a finite set of disjoint simple closed C^1 curves.*

The sole purpose of the next proposition is to convince the reader that the level lines of a function provide a faithful representation of the function.

Proposition 11.12. *Let $u \in \mathcal{F} \cap C^1$. Then u can be reconstructed from the family of all of its level lines at nonsingular levels, along with their levels.*

Proof. Let G be the closure of the union of the ranges of all level lines of u at nonsingular levels. If $\mathbf{x} \in G$, then there are points $\mathbf{x}_n$ belonging to level lines of some levels λ_n such that $\mathbf{x}_n \to \mathbf{x}$. As a consequence, $\lambda_n = u(\mathbf{x}_n) \to u(\mathbf{x})$. So we get back the value of $u(\mathbf{x})$.
Let now $\mathbf{x}$ belong to the open set G^c. Let us first prove that $Du(\mathbf{x}) = 0$. Assume by contradiction that $Du(\mathbf{x}) \neq 0$. By using the first order Taylor expansion of u around $\mathbf{x}$, one sees that for all $r > 0$ the connected range $u(B(\mathbf{x}, r))$ must contain some interval $(u(\mathbf{x}) - \alpha(r), u(\mathbf{x}) + \alpha(r))$ with $\alpha(r) \to 0$ as $r \to 0$. By Sard's theorem some of the values in this interval are nonsingular. Thus we can find nonsingular levels $\lambda_n \to u(\mathbf{x})$ and points $\mathbf{x}_n \to \mathbf{x}$ such that $u(\mathbf{x}_n) = \lambda_n$. This implies that $\mathbf{x} \in G$ and yields a contradiction.
Thus $Du(\mathbf{x}) = 0$ on G^c and u is therefore constant on each connected component A of G^c. The value of u is then uniquely determined by the value of u on the boundary of A. This value is known, since ∂A is contained in G. $\square$

Figure 11.2: 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 that are multiples of 12 are displayed. Notice that we do not need a previous smoothing to visualize the shape structures in an image: It is sufficient to quantize the displayed levels.

11.3 Curvature of the level lines

The intrinsic local coordinates

We continue to work in $\mathbb{R}^2$. Consider a real-valued function u that is twice continuously differentiable in a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^2$. To simplify the notation, we will often write Du rather than $Du(\mathbf{x}_0)$, and so on.

Definition 11.13. *If the gradient $Du = (u_x, u_y) \neq 0$, then we establish a local coordinate system by letting $\boldsymbol{i} = Du/|Du|$ and $\boldsymbol{j} = Du^{\perp}/|Du|$, where $Du^{\perp} = (-u_y, u_x)$. Thus, for a point $\mathbf{x}$ near $\mathbf{x}_0$, we write $\mathbf{x} = \mathbf{x}_0 + x\boldsymbol{i} + y\boldsymbol{j}$ and the local coordinates of $\mathbf{x}$ are (x, y). (See Figure 11.3.) Without risk of ambiguity we shall write $u(x, y)$ for $u(\mathbf{x}) = u(\mathbf{x}_0 + x\boldsymbol{i} + y\boldsymbol{j})$.*

Since u is C^2, we can use Taylor's formula to express u in this coordinate system in a neighborhood of $\mathbf{x}_0$.

$$u(\mathbf{x}) = u(x, y) = u(\mathbf{x}_0) + px + ax^2 + by^2 + cxy + O(|\mathbf{x}|^3), \tag{11.3}$$

where $p = u_x(0,0) = |Du(\mathbf{x}_0)| > 0$ and

$$\begin{aligned} a &= \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(0,0) = \frac{1}{2}D^2u\Big(\frac{Du}{|Du|}, \frac{Du}{|Du|}\Big)(\mathbf{x}_0), \\ b &= \frac{1}{2}\frac{\partial^2 u}{\partial y^2}(0,0) = \frac{1}{2}D^2u\Big(\frac{Du^{\perp}}{|Du|}, \frac{Du^{\perp}}{|Du|}\Big)(\mathbf{x}_0), \\ c &= \frac{\partial^2 u}{\partial x \partial y}(0,0) = \quad D^2u\Big(\frac{Du^{\perp}}{|Du|}, \frac{Du}{|Du|}\Big)(\mathbf{x}_0). \end{aligned} \tag{11.4}$$

Exercise 11.3. Check the three above formulas. ■

The implicit function theorem 11.8 ensures that in a neighborhood of $\mathbf{x}_0$ the set $\{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$ is a C^2 graph whose equation can be written in the local coordinates $x = \varphi(y)$, where φ is a C^2 function in an interval I containing $y = 0$. In this interval, we have $u(\varphi(y), y) = u(\mathbf{x}_0)$. Differentiating this shows that $u_x\varphi' + u_y = 0$ for $y \in I$. Since $|Du(\mathbf{x}_0)| = u_x(0,0)$ and $u_y(0,0) = 0$

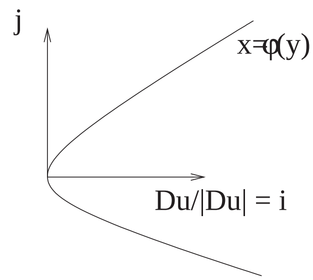

Figure 11.3: Intrinsic coordinates. Note that $\varphi''(0) > 0$, so $b < 0$.

in our coordinate system, we obtain $\varphi'(0) = 0$. A second differentiation of $u_x\varphi' + u_y = 0$ yields

$$(u_{xx}\varphi' + u_{xy})\varphi' + u_x\varphi'' + u_{yx}\varphi' + u_{yy} = 0.$$

Since $\varphi'(0) = 0$, we obtain $\varphi''(0) = -u_{yy}(0,0)/u_x(0,0)$. Using the notation of (11.4), one obtains

$$\varphi(y) = -\frac{2b}{p}y^2 + o(y^2). \tag{11.5}$$

Equation (11.5) is the representation of the level line $\{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$ in the intrinsic coordinates at $\mathbf{x}_0$. Let us set $|2b/p| = 1/R$. If the curve is a circle, R is the radius of this circle. More generally R is called *radius of the osculatory circle* to the curve. See exercise 11.11.

We are now going to do another simple computation to determine the curvature vector of the Jordan arc $\mathbf{c}$ defined by $\mathbf{c}(y) = \mathbf{x}_0 + \varphi(y)\boldsymbol{i} + y\boldsymbol{j}$ near $y = 0$. Recall that we denote the curvature of a curve $\mathbf{c}$ by $\boldsymbol{\kappa}(\mathbf{c})$ and the value of this function at a point $\mathbf{c}(y)$ by $\boldsymbol{\kappa}(\mathbf{c})(y)$.) Since in the local coordinates $\mathbf{c}'(y) = (\varphi'(y), 1)$ and $\mathbf{c}''(y) = (\varphi''(y), 0)$, at $y = 0$, we have $\mathbf{c}'(0) = (0, 1)$ and $\mathbf{c}''(0) = (\varphi''(0), 0)$, so that $\mathbf{c}''(0) \cdot \mathbf{c}'(0) = 0$. Using this and the expression of the curvature in local graph coordinates (11.2) yields $\boldsymbol{\kappa}(\mathbf{c})(0) = (\varphi''(0), 0)$. We now use (11.5) and (11.4) to write the last expression as

$$\boldsymbol{\kappa}(\mathbf{c})(0) = -\frac{1}{|Du|}D^2u\Big(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|}\Big)\frac{Du}{|Du|}(\mathbf{x}_0) \tag{11.6}$$

This tells us that the vectors $\boldsymbol{\kappa}(\mathbf{c})(0)$ and $Du(0)$ are collinear. Equation (11.6) also leads to the following definition and lemma introducing a scalar curvature.

Definition 11.14. *Let u be a real-valued function that is C^2 in a neighborhood of a point $\mathbf{x} \in \mathbb{R}^2$ and assume that $Du(\mathbf{x}) \neq 0$. The curvature of u at $\mathbf{x}$, denoted by* $\mathrm{curv}(u)(\mathbf{x})$*, is the real number defined in the local coordinates at $\mathbf{x}$ by*

$$\frac{1}{|Du|^3}D^2u(Du^\perp, Du^\perp)(\mathbf{x}) = \frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}}(0,0). \tag{11.7}$$

Exercise 11.4. Check the above identity. ∎

Lemma 11.15. *Assume that $u : \mathbb{R}^2 \to \mathbb{R}$ is C^2 in a neighborhood of a point $\mathbf{x}_0$ and assume that $Du(\mathbf{x}_0) \neq 0$. Let $N = N(\mathbf{x}_0)$ be a neighborhood of $\mathbf{x}_0$ in which the iso-level set of u, $\{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$, is a simple C^2 arc, which we still denote by $\mathbf{x} = \mathbf{x}(s)$. Then at every point $\mathbf{x}$ of this arc,*

$$\boldsymbol{\kappa}(\mathbf{x}) = -\mathrm{curv}(u)(\mathbf{x})\frac{Du}{|Du|}(\mathbf{x}). \tag{11.8}$$

Proof. This is an immediate consequence of (11.6) and (11.7). We need only remark that, given the hypotheses of the lemma, there is a neighborhood N of $\mathbf{x}_0$ such that $Du(\mathbf{x}) \neq 0$ for $\mathbf{x} \in N$ and such that $\{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$ is a simple C^2 arc for $\mathbf{x} \in N$. Then the argument we made to derive (11.6) holds for any point $\mathbf{x} \in N \cap \{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$. □

The next exercise proposes as a sanity check a verification that the curvature thus defined is contrast invariant and rotation invariant.

Exercise 11.5. Use equation (11.7) to show that

$$\mathrm{curv}(u) = \frac{\partial}{\partial x}\Big(\frac{Du}{|Du|}\Big) + \frac{\partial}{\partial y}\Big(\frac{Du}{|Du|}\Big) = \mathrm{div}\Big(\frac{Du}{|Du|}\Big). \tag{11.9}$$

Use this last relation to show that $\mathrm{curv}(g(u)) = \mathrm{curv}(u)$ if g is any C^2 function $g : \mathbb{R} \to \mathbb{R}$ such that $g'(x) > 0$ for all $x \in \mathbb{R}$. What happens if $g'(x) < 0$ for all $x \in \mathbb{R}$? Show that $\mathrm{curv}(U) = \mathrm{curv}(u)$, where $U(s,t) = u(x,y)$ and $x = s\cos\theta - t\sin\theta$, $y = s\sin\theta + t\cos\theta$. ■

Before leaving this section, we wish to emphasize geometric aspects of the functions we have introduced. Perhaps the most important fact is that the curvature of a C^2 Jordan arc Γ is an intrinsic property of Γ; it does not depend on the parameterization. If $\mathbf{x}$ is a point on Γ, then the curvature vector $\boldsymbol{\kappa}(\Gamma)(\mathbf{x})$ points toward the center of the osculating circle. Furthermore, $1/|\boldsymbol{\kappa}(\Gamma)(\mathbf{x})|$ is the radius of this circle, so when $|\boldsymbol{\kappa}(\Gamma)(\mathbf{x})|$ is large, the osculating circle is small, and the curve is "turning a sharp corner."

If $Du(\mathbf{x}) \neq 0$, then the vector $Du(\mathbf{x})$ points in the direction of greatest increase, or steepest ascent, of u at $\mathbf{x}$: Following the gradient leads uphill. The function $\mathrm{curv}(u)$ does not have such a clear geometric interpretation, and it is perhaps best thought of in terms of equation (11.8): $\mathrm{curv}(u)(\mathbf{x})$ is the coefficient of $-Du(\mathbf{x})/|Du(\mathbf{x})|$ that yields the curvature vector $\boldsymbol{\kappa}(\mathbf{x})$ of the level curve through the point $\mathbf{x}$. We cannot over emphasize the importance of the two operators curv and Curv for the theories that follow. In addition to (11.8), a further relation between these operators is shown in Proposition 12.8, and it is this result that connects function smoothing with curve smoothing.

11.4 The principal curvatures of a level surface

We saw in Exercise 11.5 that $\mathrm{curv}(u)$ was contrast invariant. This idea will be generalized to $\mathbb{R}^N$ by introducing other differential operators that are contrast invariant. These operators will be functions of the *principal curvatures* of the level surfaces of u. For $\mathbf{z} \in \mathbb{R}^N$, $\mathbf{z}^\perp$ will denote the hyperplane $\{\mathbf{y} \mid \mathbf{z} \cdot \mathbf{y} = 0\}$ that is orthogonal to $\mathbf{z}$. (There should be no confusion with this notation and the same notation for $\mathbf{z} \in \mathbb{R}^2$. In $\mathbb{R}^2$, $\mathbf{z}^\perp$ is a vector orthogonal to $\mathbf{z}$, and the corresponding "hyperplane" is the line $\{t\mathbf{z}^\perp \mid t \in \mathbb{R}\}$.)

Proposition 11.16. *Assume that $u : \mathbb{R}^N \to \mathbb{R}$ is C^2 in a neighborhood of a point $\mathbf{x}_0$ and assume that $Du(\mathbf{x}_0) \neq 0$. Let $g : \mathbb{R} \to \mathbb{R}$ be a C^2 contrast change such that $g'(s) > 0$ for all $s \in \mathbb{R}$. Then $Dg(u(\mathbf{x}_0)) = g'(u(\mathbf{x}_0))Du(\mathbf{x}_0)$, and $\tilde{D}^2 g(u(\mathbf{x}_0)) = g'(u(\mathbf{x}_0))\tilde{D}^2 u(\mathbf{x}_0)$, where $\tilde{D}^2 u(\mathbf{x}_0)$ denotes the restriction of the quadratic form $D^2 u(\mathbf{x}_0)$ to the hyperplane $Du(\mathbf{x}_0)^\perp$. This means, in particular, that $(1/|Du(\mathbf{x}_0)|)\tilde{D}^2 u(\mathbf{x}_0)$ is invariant under such a contrast change.*

Proof. To simplify the notation, we will suppress the argument $\mathbf{x}_0$; thus, we write Du for $Du(\mathbf{x}_0)$, and so on. We use the notation $\mathbf{y} \otimes \mathbf{y}$, $\mathbf{y} \in \mathbb{R}^N$, to denote the linear mapping $\mathbf{y} \otimes \mathbf{y} : \mathbb{R}^N \to \mathbb{R}^N$ defined by $(\mathbf{y} \otimes \mathbf{y})(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{y})\mathbf{y}$. The range of $\mathbf{y} \otimes \mathbf{y}$ is the one-dimensional space $\mathbb{R}\mathbf{y}$.

An application of the chain rule shows that $Dg(u) = g'(u)Du$. This implies that $Du^\perp = Dg(u)^\perp$. (Recall that $g'(s) > 0$ for all $s \in \mathbb{R}$.) A second differentiation shows that

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

If $\mathbf{y} \in Du^\perp$, then $(Du \otimes Du)(\mathbf{y}) = 0$ and $D^2 g(u)(\mathbf{y}, \mathbf{y}) = g'(u) D^2 u(\mathbf{y}, \mathbf{y})$. This means that $D^2 g(u) = g'(u) D^2 u$ on $Du^\perp = Dg(u)^\perp$, which proves the result. □

Exercise 11.6. Taking euclidian coordinates, give the matrix of $\mathbf{y} \otimes \mathbf{y}$. Check the above differentiations. ■

We are now going to define locally the level surface of a smooth function u, and for this we quote one more version of the implicit function theorem, in arbitrary dimension N.

Theorem 11.17 (Implicit function theorem). *Assume that $u : \mathbb{R}^N \to \mathbb{R}$ is C^m in the neighborhood of $\mathbf{x}_0$ and assume that $Du(\mathbf{x}_0) \neq 0$. Write $\mathbf{x} = \mathbf{x}_0 + \mathbf{y} + z\boldsymbol{i}$, where $\boldsymbol{i} = Du(\mathbf{x}_0)/|Du(\mathbf{x}_0)|$ and $\mathbf{y} \in Du(\mathbf{x}_0)^\perp$. Then there exists a ball $B(\mathbf{x}_0, \rho)$ and a unique real-valued C^m function φ defined on $B(\mathbf{x}_0, \rho) \cap \{\mathbf{x} \mid \mathbf{x} = \mathbf{x}_0 + \mathbf{y}, \boldsymbol{i} \cdot \mathbf{y} = 0\}$ such that for every $\mathbf{x} \in B(\mathbf{x}_0, \rho)$*

$$u(\mathbf{x}) = u(\mathbf{x}_0) \Longleftrightarrow \varphi(\mathbf{y}) = z.$$

In other words, the equation $\varphi(\mathbf{y}) = z$ describes the set $\{\mathbf{x} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$ near $\mathbf{x}_0$ as the graph of a C^m function φ. Thus, locally we have a surface passing through $\mathbf{x}_0$ that we call the *level surface of u around* $\mathbf{x}_0$.

We are going to use Proposition 11.16 and Theorem 11.17, first, to give a simple intrinsic representation for the level surface of a function u around a point $\mathbf{x}_0$ and, second, to relate the eigenvalues of the quadratic form introduced in Proposition 11.16 to the curvatures of lines drawn on the level surface of u.

Proposition 11.18. *Assume that $u : \mathbb{R}^N \to \mathbb{R}$ is C^2 in a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^N$ and that $p = Du(\mathbf{x}_0) \neq 0$. Denote the eigenvalues of the restriction of the quadratic form $D^2 u(\mathbf{x}_0)$ to the hyperplane $Du(\mathbf{x}_0)^\perp$ by $\mu_1, \dots, \mu_{N-1}$. Let $\boldsymbol{i}_N = Du(\mathbf{x}_0)/|Du(\mathbf{x}_0)|$ and select $\boldsymbol{i}_1, \dots, \boldsymbol{i}_{N-1}$ so they form an orthonormal basis of eigenvectors of the restriction of $D^2 u(\mathbf{x}_0)$ to $Du(\mathbf{x}_0)^\perp$. Write $\mathbf{x} = \mathbf{x}_0 + \mathbf{z}$, where $\mathbf{z} = x_1\boldsymbol{i}_1 + \cdots + x_N\boldsymbol{i}_N = \mathbf{y} + x_N\boldsymbol{i}_N$. Then if $|\mathbf{z}|$ is sufficiently small, the* *function $\varphi(\mathbf{y}) = x_N$ that solves the equation $u(\mathbf{y}, \varphi(\mathbf{y})) = u(\mathbf{x}_0)$ can be expressed locally as*

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

Proof. Assume, without loss of generality, that $\mathbf{x}_0 = 0$ and that $u(0) = 0$. Using the notation of Theorem 11.17, for $\mathbf{x} \in B(0, \rho)$, $u(\mathbf{y}, x_N) = 0$ if and only if $\varphi(\mathbf{y}) = x_N$, and φ is C^2 in $B(0, \rho)$. Furthermore, by differentiating the expression $u(\mathbf{y}, \varphi(\mathbf{y})) = 0$, we see that $u_{x_i} + u_{x_N}\varphi_{x_i} = 0$, $i = 1, \dots, N-1$ for $|\mathbf{x}| < \rho$. In particular, $u_{x_i}(0) + u_{x_N}(0)\varphi_{x_i}(0) = 0$. In the local coordinate system we have chosen, $|Du(0)| = |u_{x_N}(0)|$, and since $Du(0) \neq 0$, we conclude that $u_{x_i}(0) = 0$ for $i = 1, \dots, N-1$ and hence that $\varphi_{x_i}(0) = 0$ for $i = 1, \dots, N-1$. This means that the local expansion of φ has the form

$$\varphi(\mathbf{y}) = \frac{1}{2} D^2\varphi(0)(\mathbf{y}, \mathbf{y}) + o(|\mathbf{y}|^2).$$

Now differentiate the relation $u_{x_i} + u_{x_N}\varphi_{x_i} = 0$ again to obtain

$$u_{x_i x_j} + u_{x_i x_N}\varphi_{x_j} + (u_{x_N x_j} + u_{x_N x_N}\varphi_{x_j})\varphi_{x_i} + u_{x_N}\varphi_{x_i x_j} = 0.$$

Since we have just shown that $\varphi_{x_i}(0) = 0$ for $i = 1, \dots, N-1$, we see from this last expression that $\tilde{D}^2u(0) + p\tilde{D}^2\varphi(0) = 0$, where $p = u_{x_N}(0)$ and $\tilde{D}^2u(0)$, $\tilde{D}^2\varphi(0)$ are the restrictions of the quadratic forms $D^2u(0)$ and $D^2\varphi(0)$ to the hyperplane $Du(0)^\perp$. Thus we have

$$x_N = \frac{-1}{2p} D^2u(0)(\mathbf{y}, \mathbf{y}) + o(|\mathbf{y}|^2). \tag{11.10}$$

Recall that $\mathbf{y} \in Du(0)^\perp$ and that $\mathbf{y} = x_1\boldsymbol{i}_1 + \dots + x_{N-1}\boldsymbol{i}_{N-1}$, where the $\boldsymbol{i}_i$ are an orthonormal basis of eigenvectors of $D^2(0)$ restricted to $Du(0)^\perp$. Thus,

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

which is what we wished to prove. □

This formula reads

$$x_2 = \frac{-1}{2p}\mu_1 x_1^2 + o(|x_1|^2)$$

if $N = 2$, which is just equation (11.5) with different notation. Thus, $\mu_1 = |Du|\mathrm{curv}(u)$, confirming that $\mu_1 = \partial^2 u/\partial x_1^2$. We are now going to use our two-dimensional analysis to give a further interpretation of the eigenvalues μ_i for $N > 2$. We begin by considering the curve $\Gamma_{\boldsymbol{\nu}}$ defined by the two equations $\mathbf{x} = \mathbf{x}_0 + t\boldsymbol{\nu} + x_N\boldsymbol{i}_N$ and $\varphi(t\boldsymbol{\nu}) = x_N$, where $\boldsymbol{\nu}$ is a unit vector in $Du(\mathbf{x}_0)^\perp$. Their solution in the local coordinates is $\varphi(t\boldsymbol{\nu}) = x_N$, whenever $t \in \mathbb{R}$ is small. Thus, Γ_ν is a curve passing by $\mathbf{x}_0$, drawn on the level surface of u and projecting into a straight line of $Du^\perp$. By (11.10) its equation is

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

and its normal at $\mathbf{x}_0$ is $\frac{Du(\mathbf{x}_0)}{|Du(\mathbf{x}_0)|}$. Thus the curvature vector of Γ_ν at $\mathbf{x}_0$ is

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

By defining $\kappa_{\boldsymbol{\nu}} = |Du(\mathbf{x}_0)|^{-1} D^2u(\mathbf{x}_0)(\boldsymbol{\nu},\boldsymbol{\nu})$, we have

$$\boldsymbol{\kappa}(\Gamma_{\boldsymbol{\nu}})(\mathbf{x}_0) = -\kappa_{\boldsymbol{\nu}} \frac{Du(\mathbf{x}_0)}{|Du(\mathbf{x}_0)|},$$

which has the same form as equation (11.8). So the modulus of $\kappa_{\boldsymbol{\nu}}$ is equal to the modulus of the curvature of Γ_ν at $\mathbf{x}_0$. This leads us to call principal curvatures of the level surface of u at $\mathbf{x}_0$ the numbers $\kappa_{\boldsymbol{\nu}}$ obtained by letting $\boldsymbol{\nu} = \boldsymbol{i}_j$, $j = 1, \dots, N-1$, where the unit vectors $\boldsymbol{i}_j$ are an orthonormal system of eigenvectors of $D^2u(\mathbf{x}_0)$ restricted to $Du(\mathbf{x}_0)^\perp$.

Definition 11.19. *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be C^2 at $\mathbf{x}_0$, with $Du(\mathbf{x}_0) \neq 0$. The principal curvatures of u at $\mathbf{x}_0$ are the real numbers*

$$\kappa_j = \frac{\mu_j}{|Du(\mathbf{x}_0)|},$$

where μ_j are the eigenvalues of $D^2u(\mathbf{x}_0)$ restricted to $Du(\mathbf{x}_0)^\perp$.

It follows from Proposition 11.16 that the principal curvatures are invariant under a C^2 contrast change g such that $g'(s) > 0$ for all $s \in \mathbb{R}$.

Definition 11.20. *The mean curvature of a C^2 function $u : \mathbb{R}^N \to \mathbb{R}$ at $\mathbf{x}_0 \in \mathbb{R}^N$ is the sum of the principal curvatures at $\mathbf{x}_0$. It is denoted by* $\mathrm{curv}(u)(\mathbf{x}_0)$.

Note that this definition agrees with Definition 11.2 when $N = 2$. The next result provides another representation for $\mathrm{curv}(u)$.

Proposition 11.21. *The mean curvature of u is given by*

$$\mathrm{curv}(u) = \mathrm{div}\Big(\frac{Du}{|Du|}\Big).$$

Proof. Represent the matrix D^2u in the coordinate system $\boldsymbol{i}_j$, $j = 1, \dots, N-1$, and $\boldsymbol{i}_N = Du(\mathbf{x}_0)/|Du(\mathbf{x}_0)|$, where the $\boldsymbol{i}_j$, $j = 1, \dots, N-1$, form a complete set of eigenvectors of the linear mapping $D^2u(\mathbf{x}_0)$ restricted to $Du^\perp(\mathbf{x}_0)$. Then in this coordinate system, $D^2u(\mathbf{x}_0)$ has the following form (illustrated for $N = 5$):

$$D^2u(\mathbf{x}_0) = \begin{bmatrix} u_{11} & 0 & 0 & 0 & u_{15} \\ 0 & u_{22} & 0 & 0 & u_{25} \\ 0 & 0 & u_{33} & 0 & u_{35} \\ 0 & 0 & 0 & u_{44} & u_{45} \\ u_{51} & u_{52} & u_{53} & u_{54} & u_{55} \end{bmatrix},$$

where $u_{jk} = u_{x_j x_k}(\mathbf{x}_0)$, and $u_{jj} = \kappa_j$ is the eigenvalue associated with $\boldsymbol{i}_j$. Thus, by definition, we see that

$$\mathrm{curv}(u) = \frac{\Delta u}{|Du|} - \frac{1}{|Du|} D^2u\Big(\frac{Du}{|Du|}, \frac{Du}{|Du|}\Big).$$

We also have

$$\begin{aligned}\operatorname{div}\Big(\frac{Du}{|Du|}\Big) &= \sum_{j=1}^{N}\frac{\partial}{\partial x_j}\Big(\frac{u_{x_j}}{|Du|}\Big)\\ &= \frac{1}{|Du|}\sum_{j=1}^{N}u_{x_jx_j} - \frac{1}{|Du|^3}\sum_{j,k=1}^{N}u_{x_jx_k}u_{x_j}u_{x_k}\\ &= \frac{\Delta u}{|Du|} - \frac{1}{|Du|}D^2u\Big(\frac{Du}{|Du|},\frac{Du}{|Du|}\Big). \qquad \square\end{aligned}$$

$\square$

With this representation, it is clear that the mean curvature has the same invariance properties as the curvature of a C^2 function defined on $\mathbb{R}^2$. (See Exercise 11.5.)

11.5 Exercises

Exercise 11.7. Let Γ by a Jordan arc parameterized by $\mathbf{x} : [0,1] \to \Gamma$ and by $\mathbf{y} : [0,1] \to \Gamma$. Show that $\mathbf{x} = \mathbf{y}\circ f$ or $\mathbf{x} = \mathbf{y}\circ(1-f)$, where f is a continuous, strictly increasing function that maps $[0,1]$ onto $[0,1]$. Hint: $\mathbf{x}$ and $\mathbf{y}$ are one-to-one, and since $[0,1]$ is compact, they are homeomorphisms. Thus, $\mathbf{y}^{-1}(\mathbf{x}) = f$ is a one-to-one continuous mapping of $[0,1]$ onto itself. As an application, give a proof of Proposition 11.4. ■

Exercise 11.8. State and prove an adaptation of Propositions 11.4 and 11.5 to a Jordan arc. ■

The curvature vector has been defined in terms of the arc length. Curves, however, are often naturally defined in terms of other parameters. The next two exercises develop the differential relations between an arc-length parameterization and another parameterization.

Exercise 11.9. Assume that Γ is a C^2 Jordan arc or curve. Let $s \mapsto \mathbf{x}(s)$ be an arc-length parameterization and let $t \mapsto \mathbf{y}(t)$ be any other parameterization with the property that $\mathbf{y}'(t) \neq 0$. Since $\mathbf{x}$ and $\mathbf{y}$ are one-to-one, we can consider the function $\mathbf{y}^{-1}(\mathbf{x}) = \varphi$. Then $\mathbf{x}(s) = \mathbf{y}(\varphi(s))$, where $\varphi(s) = t$. The inverse function φ^{-1} is given by

$$s = \varphi^{-1}(t) = \int_{t_0}^{t}\sqrt{\mathbf{y}'(r)\cdot\mathbf{y}'(r)}\,\mathrm{d}r,$$

so we know immediately that φ^{-1} is absolutely continuous with continuous derivative equal to $\sqrt{\mathbf{y}'(t)\cdot\mathbf{y}'(t)}$. Thus, we also know that $\varphi'(s) = |\mathbf{y}'(\varphi(s))|^{-1}$. Note that we made a choice above by taking $\sqrt{\mathbf{y}'(r)\cdot\mathbf{y}'(r)}$ to be positive. This is equivalent to assuming that $\mathbf{x}'(s)$ and $\mathbf{y}'(\varphi(s))$ point in the same direction or that $\varphi'(s) > 0$.

(i) Show that $\boldsymbol{\kappa}(s) = \mathbf{x}''(s) = \mathbf{y}''(\varphi(s))[\varphi'(s)]^2 + \mathbf{y}'(\varphi(s))\varphi''(s)$ and deduce that

$$\varphi''(s) = -\frac{\mathbf{y}''(\varphi(s))\varphi'(s)\cdot\mathbf{y}'(\varphi(s))}{|\mathbf{y}'(\varphi(s))|^3} = -\frac{\mathbf{y}''(\varphi(s))\cdot\mathbf{y}'(\varphi(s))}{|\mathbf{y}'(\varphi(s))|^4}.$$

(ii) Use the results of (i) to show that

$$\boldsymbol{\kappa}(s) = \mathbf{x}''(s) = \frac{1}{|\mathbf{y}'(t)|^2}\left[\mathbf{y}''(t) - \left(\mathbf{y}''(t)\cdot\frac{\mathbf{y}'(t)}{|\mathbf{y}'(t)|}\right)\frac{\mathbf{y}'(t)}{|\mathbf{y}'(t)|}\right], \tag{11.11}$$

where $\varphi(s) = t$. Show that we get the same expression for the right-hand side of (11.11) with the assumption that $\varphi'(s) < 0$. This shows that the curvature vector $\boldsymbol{\kappa}$ does not depend on the choice of parameter.

(iii) Consider the scalar function $\kappa(\mathbf{y})$ defined by $\kappa(\mathbf{y})(s) = \boldsymbol{\kappa}(s)\cdot\mathbf{x}'(s)^{\perp}$. Use equation (11.11) to show that

$$\kappa(\mathbf{y})(t) = \frac{\mathbf{y}''(t)\cdot[\mathbf{y}'(t)]^{\perp}}{|\mathbf{y}'(t)|^3}$$

Note that $\kappa(\mathbf{y})$ is determined up to a sign that depends on the sign of $\varphi'(s)$; however, $|\kappa(\mathbf{y})| = |\boldsymbol{\kappa}|$ is uniquely determined. ■

Exercise 11.10. Assume that Γ is a Jordan arc or curve that is represented by a C^1 function $t \mapsto \mathbf{x}(t)$ with the property that $\mathbf{x}'(t) \neq 0$. Prove that Γ is C^1.

Exercise 11.11.

(i) Consider the arc-length parameterization of the circle with radius r centered at the origin given by $\mathbf{x}(s) = (r\cos(s/r), r\sin(s/r))$. Show that the length of the curvature vector is $1/r$.

(ii) Compute the scalar curvature of the graph of $y = (a/2)x^2$ at $x = 0$. ■

Exercise 11.12. Complete the proof of Corollary 11.9.

Exercise 11.13. The kinds of techniques used in this exercise are important for work in later chapters. The exercise demonstrates that it is possible to bracket a C^2 function locally with two functions that are radial and either increasing or decreasing. We say that a function f is radial and increasing if there exists an increasing function $g : \mathbb{R}^+ \to \mathbb{R}$ such that $f(\mathbf{x}) = g(|\mathbf{x}_c - \mathbf{x}|^2)$, $\mathbf{x}_c \in \mathbb{R}^2$. We say that f is radial and decreasing if g is decreasing. Let $u : \mathbb{R}^2 \to \mathbb{R}$ be C^2 and assume that $Du(\mathbf{x}_0) \neq 0$. We wish to show that for every $\varepsilon > 0$ there exist two C^2 radial functions f_ε^- and f_ε^+ (increasing or decreasing, depending on the situation) that satisfy the following four conditions:

$$f_\varepsilon^-(\mathbf{x}_0) = u(\mathbf{x}_0) = f_\varepsilon^+(\mathbf{x}_0), \tag{11.12}$$

$$Df_\varepsilon^-(\mathbf{x}_0) = Du(\mathbf{x}_0) = Df_\varepsilon^+(\mathbf{x}_0), \tag{11.13}$$

$$\operatorname{curv}(f_\varepsilon^-)(\mathbf{x}_0) + \frac{2\varepsilon}{p} = \operatorname{curv}(u)(\mathbf{x}_0) = \operatorname{curv} f_\varepsilon^+(\mathbf{x}_0) - \frac{2\varepsilon}{p}, \tag{11.14}$$

$$f_\varepsilon^-(\mathbf{x}) + o(|\mathbf{x}_0 - \mathbf{x}|^2) \le u(\mathbf{x}) \le f_\varepsilon^+(\mathbf{x}) + o(|\mathbf{x}_0 - \mathbf{x}|^2). \tag{11.15}$$

1. Without loss of generality, take $\mathbf{x}_0 = (0,0)$, $u(0,0) = 0$, and $Du(\mathbf{x}_0) = (p,0)$, $p > 0$. Then we have the Taylor expansion

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

where a, b, and c are given in (11.4). Show that for every $\varepsilon > 0$,

$$px + \Big(-\frac{c^2}{\varepsilon} + a\Big)x^2 + (b-\varepsilon)y^2 + o(x^2+y^2) \le u(x,y) \le px + \Big(\frac{c^2}{\varepsilon} + a\Big)x^2 + (b+\varepsilon)y^2 + o(x^2+y^2).$$

2. Let f be a radial function defined by $f(x,y) = g((x-x_c)^2 + y^2)$, where $g : \mathbb{R}^+ \to \mathbb{R}$ is C^2 and either increasing or decreasing. Show by expanding f at $(0,0)$ that

$$f(x,y) = g(x_c^2) - 2x_c g'(x_c^2)x + (2x_c^2 g''(x_c^2) + g'(x_c^2))x^2 + g'(x_c^2)y^2 + o(x^2+y^2).$$

3. The idea is to construct f_ε^+ and f_ε^- by matching the coefficients of the expansion of f with the coefficients of the functions $px + (\pm(c^2/\varepsilon) + a)x^2 + (b \pm \varepsilon)y^2$. There are three cases to consider: $b < 0$, $b = 0$, and $b > 0$. Show that in each case it is possible to find values of x_c and functions g so the functions f_ε^+ and f_ε^- satisfy the four condition. Note that both x_c and g depend on ε. Discuss the geometry for each case. ■

Exercise 11.14. By computing explicitly the terms $\partial g(u)/\partial x_i$, verify that $Dg(u) = g'(u)Du$. Similarly, verify that $D^2(g(u)) = g''(u)Du \otimes Du + g'(u)D^2u$ by computing the second-order terms $\partial^2 g(u)/\partial x_i \partial x_j$. ■

11.6 Comments and references

Calculus and differential geometry. The differential calculus of curves and surfaces used in this chapter can be found in many books, and no doubt most readers are familiar with this material. Nevertheless, a few references to specific results may be useful. As a general reference on calculus, and as a specific reference for the implicit function theorem, we suggest the text by Courant and John [77]. (The implicit function theorem can be found on page 221 of volume II.) Elementary results about classical differential geometry can be found in [262]. A statement and proof of Sard's theorem can be found in [177].

Level lines. An introduction to the use of level lines in computer vision can be found in [58]. A complete discussion of the definition of level lines for BV functions can be found in [16]. One can decompose an image into into its level lines at quantized levels and conversely reconstruct the image from this topographic map. A fast algorithm, the Fast Level Set Transform (FLST) performing these algorithms is described in [185]. Its principle is very simple: a) perform the bilinear interpolation, b) rule out all singular levels where saddle point occur c) quantize the other levels, in which the level lines are finite unions of parametric Jordan curves. The image is then parsed into a set of parametric Jordan curves. This set is easily ordered in a tree structure, since two Jordan level curves do not meet. Thus either one surrounds the other one or conversely. The level lines tree is a shape parser for the image, many level lines surrounding perceptual shapes or parts of perceptual shapes.

Curvature. It is a well-known mathematical technique to define a set implicitly as the zero set of its distance function. In case the set is a curve, one can compute its curvature at a point $\mathbf{x}$ by computing the curvature $\mathrm{curv}(u)(\mathbf{x})$, where u is a signed distance function of the curve. This yields an intrinsic formula for the curvature that is not dependent on a parameterization of the curve. The same technique has been applied in recent years as a useful numerical tool. This started with Barles report on flame propagation [34] and was extended by Sethian [256] and by Osher and Sethian [224] in a series of papers on the numerical simulation of the motion of a surface by its mean curvature.

Chapter 12

The Main Curvature Equations

The purpose of this chapter is to introduce the curvature motion PDE's for Jordan curves and images. Our main task is to establish a formal link between curve evolution and image evolution. This link will be established through the PDE formulation. The basic differential geometry used in this chapter was thoroughly developed in Chapter 11, which must therefore be read first.

12.1 The definition of a shape and how it is recognized

Relevant information in images has been reduced to the image level sets in Chapter 5. By Corollary 11.9, if the image is C^1, the boundary of its level sets is a finite set Jordan curves at almost every level. Thus, shape analysis can be led back to the study of these curves which we shall call "elementary shapes".

Definition 12.1. *We call elementary shape any C^1 planar Jordan curve.*

The many experiments where we display level lines of digital images make clear enough why a smoothing is necessary to restore their structure. These experiments also show that we can in no way assimilate these level lines with our common notion of shape as the silhouette of a physical object in full view. Indeed, in images of a natural environment, most observed objects are partially hidden (occluded) by other objects and often deformed by perspective. When we observe a level line we cannot be sure that it belongs to a single object; it may be composed of pieces of the boundaries of several objects that are occluding each other. Shape recognition technology has therefore focused on local methods, that is, methods that work even if a shape is not in full view or if the visible part is distorted. As a consequence, image analysis adopts the following principle: *Shape recognition must be based on local features of the shape's boundary, in this case local features of the Jordan curve, and not on its global features. If the boundary has some degree of smoothness, then these local features are based on the derivatives of the curve, namely the tangent vector, the curvature, and so on.*

Before beginning the technical aspects of this version of shape recognition, we note that most local recognition methods involve the "salient" points of a shape, which are the points where the curvature is zero (inflection points) and points where the curvature has a maximum or minimum (the "corners" of the shape). These methods reduce a shape to a finite code that consists of the coordinates of a set of characteristic points, which are mainly corners and inflection points. Recognition then amounts to comparing these sets of numbers.

12.2 Multiscale features and scale space

The methods we have just outlined—in fact, all non global computational shape recognition methods—make the following two basic assumptions, neither of which is true in practice for the rough shape data:

- The shape is a smooth Jordan curve.
- The boundary has a finite number of inflexion points and points where the curvature has a local maximum or local minimum and this number can be made as small as desired by smoothing.

The fact that these conditions can be obtained by properly smoothing a C^1 Jordan curve was proven in 1986-87 by Gage and Hamilton [120] and Grayson [128]. They showed that it is possible to transform a C^1 Jordan curve into a C^∞ Jordan curve by using the so-called *intrinsic heat equation*. The more precise statement follows soon.

Before proceeding, we wish to inject a comment about notation. For convenience, and unless it would cause ambiguity, we will not make a distinction between a Jordan curve Γ as a subset of the plane and a function $s \mapsto \mathbf{x}(s)$ such that $\Gamma = \{\mathbf{x}(s)\}$. As we have already done, we will speak of the Jordan curve $\mathbf{x}$. Since we will be speaking of families of Jordan curves dependent on a parameter $t > 0$, we will most often denote these families by $\mathbf{x}(t, s)$, where the second variable is a parameterization of the Jordan curve. Thus, $\mathbf{x}(t, s)$ has three meanings: a family of Jordan curves, a family of functions that represent these curves, and a particular point on one of these curves. The notation s will be usually reserved to an arc-length parameter. Finally, everything we do in this chapter is local—it takes place in some neighborhood of a given point. This means we are generally speaking of Jordan arcs rather than Jordan curves.

Definition 12.2. *Let $\mathbf{x}(t, s)$, $t > 0$, be a family of C^2 Jordan curves and assume that for each t, s is an arc length parameterization of $\mathbf{x}(t, s)$. We say that $\mathbf{x}(t, s)$ satisfies the intrinsic heat equation if*

$$\frac{\partial \mathbf{x}}{\partial t}(t, s) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t, s) = \boldsymbol{\kappa}(\mathbf{x})(t, s). \tag{12.1}$$

Theorem 12.3 (Grayson). *Let $\mathbf{x}_0$ be a C^1 Jordan curve. By using the intrinsic heat equation, it is possible to evolve $\mathbf{x}_0$ into a family of Jordan curves $\mathbf{x}(t, s)$ such that $\mathbf{x}(0, s) = \mathbf{x}_0(s)$ and such that for every $t > 0$, $\mathbf{x}(t, s)$ is C^∞ (actually analytical) and satisfies the equation* (12.1). *Furthermore, for every $t > 0$, $\mathbf{x}(t, s)$ has only a finite number of inflection points and curvature extrema, and the number of these points does not increase with t. For every initial curve,* *there is a scale t_0 such that the curve $\mathbf{x}(t,s)$ is convex for $t \geq t_0$ and there is a scale t_1 such that the curve $\mathbf{x}(t,s)$ is a single point for $t \geq t_1$.*

It is time to say what we mean by "curve scale space", or "shape scale space." We will refer to any process that smooths a Jordan curve and that depends on a real parameter t. Thus a shape scale space associates with an initial Jordan curve $\mathbf{x}(0,s) = \mathbf{x}_0(s)$ a family of smooth curves $\mathbf{x}(t,s)$. For example, the intrinsic heat equation eliminates spurious details of the initial shape and retains simpler, more reliable versions of the shape, and these smoothed shapes have finite codes. Suppose that we wish to compare two original versions of a shape $\mathbf{x}_0$ and $\mathbf{x}_1$ that have been captured under different conditions of noise and distortions. Comparing these two shapes is simply impossible. If, however, they are smoothed to the shapes $\mathbf{x}_0(t,\cdot)$ and $\mathbf{x}_1(t,\cdot)$, then it is possible to compare the codes of $\mathbf{x}_0(t,\cdot)$ and $\mathbf{x}_1(t,\cdot)$. A scale space is causal in the terminology of vision theory if it does not introduce new features. Grayson's theorem therefore defines a causal scale space.

12.3 From image motion to curve motion

The intrinsic heat equation is only one example from a large family of nonlinear equations that move curves with a curvature-dependent speed, that is, $\partial\mathbf{x}/\partial t$ is a function of the curvature of the curve $\mathbf{x}$. The only requirement for our purposes is that the speed is a nondecreasing function of the magnitude of the curvature $|\boldsymbol{\kappa}(\mathbf{x})|$.

Definition 12.4. *We say that a C^2 function $u : \mathbb{R}^+ \times \mathbb{R}^2 \to \mathbb{R}$ satisfies a curvature equation if for some real-valued function $g(\kappa,t)$, which is nondecreasing in κ and satisfies $g(0,t) = 0$,*

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = g(\mathrm{curv}(u)(t,\mathbf{x}),t)|Du|(t,\mathbf{x}). \tag{12.2}$$

Definition 12.5. *Let $\mathbf{x}(t,s)$ be a family of C^2 Jordan curves such that for every $t > 0$, $s \mapsto \mathbf{x}(t,s)$ is an arc-length parameterization. We say that the functions $\mathbf{x}(t,s)$ satisfy a curvature equation if for some real-valued function $g(\kappa,t)$ nondecreasing in κ with $g(0,t) = 0$, they satisfy*

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = g(|\boldsymbol{\kappa}(\mathbf{x})(t,s)|,t)\mathbf{n}(t,s), \tag{12.3}$$

where $\mathbf{n}$ is a unit vector in the direction of $\boldsymbol{\kappa}(\mathbf{x})$.

In the preceding definition, the equation makes sense if $\boldsymbol{\kappa}(\mathbf{x}) = 0$ since then the second member is zero. As we shall see, these equations are the only candidates to be curve or image scale spaces, and one of the main objectives of this book is to identify which forms for g are particularly relevant for image analysis. The above definitions are quite restrictive because they require the curves or images to be C^2. A more generally applicable definition of solutions for these equations will be given in Chapter 19 with the introduction of viscosity solutions. Our immediate objective is to establish the link between the motion of an image and the motion of its level lines. This will establish the relation between equations (12.2) and (12.3).

12.3.1 A link between image and curve evolution

Lemma 12.6. (Definition of the "normal flow"). *Suppose that* $(t, \mathbf{x}) \mapsto u(t, \mathbf{x})$ *is* C^2 *in a neighborhood* $T \times U$ *of the point* $(t_0, \mathbf{x}_0) \in \mathbb{R} \times \mathbb{R}^2$*, and assume that* $Du(t_0, \mathbf{x}_0) \neq 0$*. Then there exists an open interval* J *centered at* t_0*, an open disk* V *centered at* $\mathbf{x}_0$*, and a unique* C^1 *function* $\mathbf{x} : J \times V \to \mathbb{R}^2$ *that satisfy the following properties:*

(i) $u(t, \mathbf{x}(t, \mathbf{y})) = u(t_0, \mathbf{y})$ *and* $\mathbf{x}(t_0, \mathbf{y}) = \mathbf{y}$ *for all* $(t, \mathbf{y}) \in J \times V$*.*

(ii) The vectors $(\partial \mathbf{x}/\partial t)(t, \mathbf{y})$ *and* $Du(t, \mathbf{x}(t, \mathbf{y}))$ *are collinear.*

In addition, the function $\mathbf{x}$ *satisfies the following differential equation:*

$$\frac{\partial \mathbf{x}}{\partial t}(t, \mathbf{y}) = -\Big(\frac{Du}{|Du|^2}\frac{\partial u}{\partial t}\Big)(t, \mathbf{x}(t, \mathbf{y})). \tag{12.4}$$

The trajectory $t \mapsto \mathbf{x}(t, \mathbf{y})$ *is called the* normal flow starting from $(t_0, \mathbf{y})$.

Proof. Differentiating the relation $u(t, \mathbf{x}(t)) = 0$ with respect to t yields $\frac{\partial u}{\partial t} + Du.\frac{\partial \mathbf{x}}{\partial t} = 0$. By multiplying this equation by the vector Du we see that $\frac{\partial \mathbf{x}}{\partial t}$ is collinear to Du if and only if (12.4) holds. Now, this relation defines $\mathbf{x}(t)$ as the solution of an ordinary differential equation, with initial condition $\mathbf{x}(0) = \mathbf{y}$. Since u is C^2, the second member of (12.4) 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 an open interval J such that the O.D.E. (12.4) has a unique solution $\mathbf{x}(t, \mathbf{y})$ for all $\mathbf{y}$ in a neighborhood of $\mathbf{x}_0$ and $t \in J$. □

Proposition 12.7. *Assume that the function* $(t, \mathbf{x}) \mapsto u(t, \mathbf{x})$ *is* C^2 *in a neighborhood of* $(t_0, \mathbf{x}_0)$ *and that* $Du(t_0, \mathbf{x}_0) \neq 0$*. Then* u *satisfies the curvature motion equation*

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \mathrm{curv}(u)(t, \mathbf{x})|Du|(t, \mathbf{x}) \tag{12.5}$$

in a neighborhood of $(t_0, \mathbf{x}_0)$ *if and only if the normal parameterization of the level lines of* u *passing in this neighborhood satisfies the intrinsic heat equation* (12.1),

$$\frac{\partial \mathbf{x}}{\partial t}(t, \mathbf{y}) = \boldsymbol{\kappa}(t, \mathbf{x}(t, \mathbf{y})), \tag{12.6}$$

where $\boldsymbol{\kappa}(t, \mathbf{x}(t, \mathbf{y}))$ *denotes the curvature vector of the level line of* $u(t)$ *passing by* $\mathbf{x}(t, \mathbf{y})$*.*

Proof. Assume first that $\mathbf{x}(t, \mathbf{y})$ satisfies (12.6). Applying (11.8) for all t in a neighborhood of t_0 to each image $u(t) : \mathbf{x} \to u(t, \mathbf{x})$ yields

$$\boldsymbol{\kappa}(t, \mathbf{x}(t, \mathbf{y})) = -\mathrm{curv}(u)\frac{Du}{|Du|}(t, \mathbf{x}(t, \mathbf{y})).$$

Substituting (12.6) in this last relation we obtain

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

and by the normal flow equation (12.4),

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

Multiplying this equation by $Du(t,\mathbf{x}(t,\mathbf{y}))$ yields the curvature motion equation (12.5).

The converse statement follows exactly the same lines backwards. □

Exercise 12.1. Write the proof of the converse statement of Proposition 12.7. ■

The preceding proof is immediately adaptable to all curvature equations :

Proposition 12.8. *Assume that the function $(t,\mathbf{x}) \mapsto u(t,\mathbf{x})$ is C^2 in a neighborhood of $(t_0,\mathbf{x}_0)$ and that $Du(t_0,\mathbf{x}_0) \neq 0$. Let $g : \mathbb{R}\times\mathbb{R}^+ \to \mathbb{R}$ be continuous and nondecreasing with respect to κ and such that $g(-\kappa,t) = -g(\kappa,t)$. Then u satisfies the curvature motion equation*

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = g(\mathrm{curv}(u)(t,\mathbf{x}),t)|Du|(t,\mathbf{x}) \tag{12.7}$$

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

$$\frac{\partial \mathbf{x}}{\partial t}(t,\mathbf{y}) = g(|\boldsymbol{\kappa}(t,\mathbf{x}(t,\mathbf{y}))|)\frac{\boldsymbol{\kappa}(t,\mathbf{x}(t,\mathbf{y}))}{|\boldsymbol{\kappa}(t,\mathbf{x}(t,\mathbf{y}))|}. \tag{12.8}$$

12.3.2 Introduction to the affine curve and function equations

There are two curvature equations that are affine invariant and are therefore particularly well suited for use in shape recognition. In their definition, for $x \in \mathbb{R}$, $x^{1/3}$ stands for $\mathrm{sign}(x)|x|^{1/3}$.

Definition 12.9. *The image evolution equation*

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = (\mathrm{curv}(u)(t,\mathbf{x}))^{1/3}|Du(t,\mathbf{x})| \tag{12.9}$$

is called affine morphological scale space (AMSS). The curve evolution equation

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = |\boldsymbol{\kappa}(\mathbf{x}(t,s))|^{1/3}\mathbf{n}(t,s) \quad \Big(= \frac{\boldsymbol{\kappa}(\mathbf{x}(t,s))}{|\boldsymbol{\kappa}(\mathbf{x}(t,s))|^{2/3}}\Big) \tag{12.10}$$

is called affine scale space (ASS).

It is clear that AMSS and ASS are equivalent in the sense of Proposition 12.8. As one would expect from the names of these equations, they both have some sort of affine invariance. This is the subject of the next definition, Exercises 12.3 and 12.4 and the next section.

Definition 12.10. *We say that a curvature equation (E) (image evolution equation) is affine invariant, if for every linear map A with positive determinant, there is a positive constant $c = c(A)$ such that $(t,\mathbf{x}) \mapsto u(t,\mathbf{x})$ is a solution of (E) if and only if $(ct, A\mathbf{x}) \mapsto u(ct, A\mathbf{x})$ is a solution of (E).*

12.3.3 The affine scale space as an intrinsic heat equation

Suppose that for each scale t, $\sigma \mapsto \mathbf{x}(t,\sigma)$ is a Jordan arc (or curve) parameterized by σ, which is not in general an arc length. As in Chapter 11, we will denote the curvature of $\mathbf{x}$ by $\boldsymbol{\kappa}$. We wish to demonstrate a formal equivalence between the affine scale space,

$$\frac{\partial \mathbf{x}}{\partial t} = |\boldsymbol{\kappa}|^{1/3}\mathbf{n}(\mathbf{x}), \tag{12.11}$$

and an "intrinsic heat equation"

$$\frac{\partial \mathbf{x}}{\partial t} = \frac{\partial^2 \mathbf{x}}{\partial \sigma^2}, \tag{12.12}$$

where σ is a special parameterization called *affine length*. We define an *affine length parameter* of a Jordan curve (or arc) to be any parameterization $\sigma \mapsto \mathbf{x}(\sigma)$ such that

$$[\mathbf{x}_\sigma, \mathbf{x}_{\sigma\sigma}] = 1, \tag{12.13}$$

where $[\mathbf{x}, \mathbf{y}] = \mathbf{x}^\perp \cdot \mathbf{y}$. If s is an arc-length parameterization, then we have (Definition 11.3)

$$\boldsymbol{\tau} = \mathbf{x}_s \quad \mathbf{n} = |\boldsymbol{\kappa}|^{-1}\mathbf{x}_{ss} \quad \Big(= \frac{\boldsymbol{\kappa}(\mathbf{x})}{|\boldsymbol{\kappa}(\mathbf{x})|}\Big). \tag{12.14}$$

We also have

$$\mathbf{x}_\sigma = \mathbf{x}_s \frac{\partial s}{\partial \sigma} \quad \text{and} \quad \mathbf{x}_{\sigma\sigma} = \mathbf{x}_{ss}\Big(\frac{\partial s}{\partial \sigma}\Big)^2 + \mathbf{x}_s \frac{\partial^2 s}{\partial \sigma^2}. \tag{12.15}$$

Thus,

$$[\mathbf{x}_\sigma, \mathbf{x}_{\sigma\sigma}] = [\mathbf{x}_s, \mathbf{x}_{ss}]\Big(\frac{\partial s}{\partial \sigma}\Big)^3,$$

and if (12.13) holds, then

$$[\mathbf{x}_s, \mathbf{x}_{ss}]\Big(\frac{\partial s}{\partial \sigma}\Big)^3 = 1.$$

Since by (12.14) $[\mathbf{x}_s, \mathbf{x}_{ss}] = \mathrm{sign}([\mathbf{x}_s, \mathbf{x}_{ss}])|\boldsymbol{\kappa}|$, we conclude that

$$\frac{\partial s}{\partial \sigma} = (\mathrm{sign}([\mathbf{x}_s, \mathbf{x}_{ss}])|\boldsymbol{\kappa}|)^{-1/3}. \tag{12.16}$$

Substituting this result in the expression for $\mathbf{x}_{\sigma\sigma}$ shown in (12.15) and writing $\mathbf{x}_s = \boldsymbol{\tau}$, we see that

$$\mathbf{x}_{\sigma\sigma} = |\boldsymbol{\kappa}|^{1/3}\mathbf{n} + \Big(\frac{\partial^2 s}{\partial \sigma^2}\Big)\boldsymbol{\tau}.$$

This tells us that equation (12.12) is equivalent to the following equation:

$$\frac{\partial \mathbf{x}}{\partial t} = |\boldsymbol{\kappa}|^{1/3}\mathbf{n} + \Big(\frac{\partial^2 s}{\partial \sigma^2}\Big)\boldsymbol{\tau}. \tag{12.17}$$

Now it turns out that the graphs of the functions $\mathbf{x}$ that you get from one time to another do not depend on the term involving $\boldsymbol{\tau}$; you could drop this term and get the same graphs. More precisely, Epstein and Gage [94] have shown that the tangential component of an equation like (12.17) does not matter as far as the geometric evolution of the curve is concerned. In fact, the tangential term just moves points along the curve itself, and the total curve evolution is determined by the normal term. As a consequence, equation (12.11) is equivalent to equation (12.12) in any neighborhood that avoids an inflection point, that is, in any neighborhood where $\mathbf{n}(\mathbf{x}) \neq 0$. At an inflection point, $\boldsymbol{\kappa} = 0$, and the two equations give the same result.

12.4 Curvature motion in N dimensions

We consider an evolution $(t, \mathbf{x}) \mapsto u(t, \mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^N$ and $u(0, \cdot) = u_0$ is an initial N–dimensional image. Let $\kappa_i(u)(t, \mathbf{x})$, $i = 1, \ldots, N-1$, denote the i^{th} principal curvature at the point $(t, \mathbf{x})$. By definition 11.20 the mean curvature is $\operatorname{curv}(u) = \sum_{i=1}^{N-1} \kappa_i$. We will now define three curvature motion flow equations in N dimensions.

Mean curvature motion. This equation is a direct translation of equation (12.5) in N dimensions:

$$\frac{\partial u}{\partial t} = |Du| \operatorname{curv}(u).$$

This says that the motion of a level hypersurface of u in the normal direction is proportional to its mean curvature.

Gaussian curvature motion for convex functions. We say that a function is convex if all of its principal curvatures have the same sign. An example of such a function is the signed distance function to a regular convex shape. The equation is

$$\frac{\partial u}{\partial t} = |Du| \prod_{i=1}^{N-1} \kappa_i.$$

The motion of a level hypersurface is proportional to the product of its principal curvatures, which is the Gaussian curvature. As we will see in Chapter 22, this must be modified before it can be applied to a nonconvex function.

Affine-invariant curvature motion. The equation is

$$\frac{\partial u}{\partial t} = |Du| \Big| \prod_{i=1}^{N-1} \kappa_i \Big|^{1/(N+1)} H\Big(\sum_{i=1}^{N-1} \operatorname{sign}(\kappa_i) \Big),$$

where $H(N-1) = 1$, $H(-N+1) = -1$, and $H(n) = 0$ otherwise. The motion is similar to Gaussian curvature motion, but the affine invariance requires that the Gaussian curvature be raised to the power $1/(N+1)$. There is no motion at a point where the principal curvatures have mixed signs. This means that only concave or convex parts of level surfaces get move by such an equation.

12.5 Exercises

Exercise 12.2. Check that all of the curvature equations (12.2) are contrast invariant. That is, assuming that h is a real-valued C^2 increasing function defined on $\mathbb{R}$ and u is C^2, show that the function v defined by $v(t,\mathbf{x}) = h(u(t,\mathbf{x}))$ satisfies one of these equations if and only if u satisfies the same equation. ∎

Exercise 12.3. Assume that $(t,\mathbf{x}) \mapsto u(t,\mathbf{x})$ is a C^2 function and that A is a 2×2 matrix with positive determinant, which we denote by $|A|$. Define the function v by $v(t,\mathbf{x}) = u(ct, A\mathbf{x})$, where $c = |A|^{-2/3}$.

(i) Prove that for each point $\mathbf{x}$ such that $Du(\mathbf{x}) \neq 0$ one has the relation

$$\operatorname{curv}(v)(\mathbf{x})|Dv(\mathbf{x})|^3 = |A|^2 \operatorname{curv}(u)(A\mathbf{x})|Du(A\mathbf{x})|^3.$$

(ii) Use (i) to deduce that the AMSS equation (12.9) is affine invariant, that is, $(t,\mathbf{x}) \mapsto u(t,\mathbf{x})$ is a solution of AMSS if and only $(t,\mathbf{x}) \mapsto v(t,\mathbf{x})$ does. ∎

Exercise 12.4. This exercise is to show that the affine scale space (equation (12.10)) is affine invariant. It relies directly on results from Exercise 11.9. Let $\sigma \mapsto \mathbf{c}(\sigma)$ be a C^2 curve, and assume that $|\mathbf{c}'(\sigma)| > 0$. Then we know from Exercise 11.9 that

$$\boldsymbol{\kappa}(\mathbf{c})(\sigma) = \frac{1}{|\mathbf{c}'(\sigma)|^2}\left[\mathbf{c}''(\sigma) - \left(\mathbf{c}''(\sigma) \cdot \frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|}\right)\frac{\mathbf{c}'(\sigma)}{|\mathbf{c}'(\sigma)|}\right]. \tag{12.18}$$

Now assume that we have a family of C^2 Jordan arcs $(t,\sigma) \mapsto \mathbf{c}(t,\sigma)$. By projecting both sides of the intrinsic heat equation onto the unit vector $\mathbf{c}'^{\perp}/|\mathbf{c}'|$ and by using (12.18), we have the following equation:

$$\frac{\partial \mathbf{c}}{\partial t} \cdot \frac{\mathbf{c}'^{\perp}}{|\mathbf{c}'|} = \frac{\mathbf{c}'' \cdot \mathbf{c}'^{\perp}}{|\mathbf{c}'|^3} \tag{12.19}$$

We say that $\mathbf{c}$ satisfies a parametric curvature equation if it satisfies equation (12.19). In the same spirit, we say that $\mathbf{c}$ satisfies a parametric affine equation if for some constant $\gamma > 0$

$$\frac{\partial \mathbf{c}}{\partial t} \cdot \mathbf{c}'^{\perp} = \gamma(\mathbf{c}'' \cdot \mathbf{c}'^{\perp})^{1/3}. \tag{12.20}$$

(i) Suppose that $\sigma = s$, an arc-length parameterization of $\mathbf{c}$. Show that equation (12.19) can be written as

$$\frac{\partial \mathbf{c}}{\partial t} = \boldsymbol{\kappa}(\mathbf{c}) + \lambda\tau,$$

where λ is a real-valued function and τ is the unit tangent vector $\partial\mathbf{c}/\partial s$. (See the remark following equation (12.17).)

(ii) Let A be a 2×2 matrix with positive determinant, and define the curve $\mathbf{y}$ by $\mathbf{y}(t,\sigma) = A\mathbf{c}(t,\sigma)$. We wish to show that if $\mathbf{c}$ satisfies a parametric affine motion, then so does $\mathbf{y}$. As a first step, show that $A\mathbf{x} \cdot (A\mathbf{y})^{\perp} = |A|\mathbf{x} \cdot \mathbf{y}$ and hence that $A(\mathbf{x}^{\perp}) \cdot (A\mathbf{x})^{\perp} = |A||\mathbf{x}|^2$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

(iii) Show that if $\mathbf{c}$ satisfies equation (12.20), then $\mathbf{y}$ satisfies

$$\frac{\partial \mathbf{y}}{\partial t} \cdot \mathbf{y}'^{\perp} = \gamma|A|^{2/3}(\mathbf{y}'' \cdot \mathbf{y}'^{\perp})^{1/3}. \blacksquare$$

12.6 Comments and references

Our definition of shape. The Italian mathematician Renato Caccioppoli proposed a theory of sets whose boundaries have finite length (finite Hausdorff measure). From his theory, it can be deduced that the boundary of a Caccioppoli set is composed of a countable number of Jordan curves, up to a set with zero length. This decomposition can even be made unambiguous. In other words, the set of Jordan curves associated with a given Caccioppoli set is unique and gives enough information to reconstruct the set [15]. This result justifies our focus on Jordan curves as the representatives of shapes.

The role of curvature in shape analysis. After Attneave's founding paper [27], let us mention the thesis by G. J. Agin [4] as being one of the first references dealing with the use of curvature for the representation and recognition of objects in computer vision. The now-classic paper by Asada and Brady [25] entitled "The curvature primal sketch" introduced the notion of computing a "multiscale curvature" as a tool for object recognition. (The title is an allusion to David Marr's famous "raw primal sketch," which is a set of geometric primitives extracted from and representing an image.) The Asada–Brady paper led to a long series of increasingly sophisticated attempts to represent shape from curvature [92, 93] and to compute curvature correctly [208]. The shape recognition programme we sketched in the beginning of this chapter was anticipated in a visionary paper by Attneave [27] and has been very recently fully developed in the works of José Luis Lisani, Pablo Musé, Frédéric Sur, Yann Gousseau and Frédéric Cao [211], [212], [54], [55].

Curve shortening. The mathematical study of the intrinsic heat equation (or curvature motion in two dimensions) was done is a series of brilliant papers in differential geometry between 1983 and 1987. We repeat a few of the titles, which indicate the progress: There was Gage [118] and Gage [119]: "Curve shortening makes convex curves circular." Then there was Gage and Hamilton [120]: "The heat equation shrinking convex plane curves." In this paper the authors showed that a plane convex curve became asymptotically close to a shrinking circle. In 1987 there was the paper by Epstein and Gage [94], and, in the same year, Grayson removed the convexity condition and finished the job [128]: "The heat equation shrinks embedded plane curves to round points." As the reviewer, U. Pinkall, wrote, "This paper contains the final solution of the long-standing *curve-shortening problem* for plane curves."

The first papers that brought curve shortening (and some variations) to image analysis were by Kimia, Tannenbaum, and Zucker [167] and by Mackworth and Mokhtarian [189]. Curve shortening was introduced as a way to do a multiscale analysis of curves, which were considered as shapes extracted from an image. In the latter paper, curve shortening was proposed as an efficient numerical tool for multiscale shape analysis.

Affine-invariant curve shortening. Affine-invariant geometry seems to have been founded by W. Blaschke. His three-volume work "Vorlesungen über Differentialgeometrie" (1921–1929) contains definitions of affine length and affine curvature. Curves with constant affine curvature are discussed in [190]. The term "affine shortening" and the corresponding curve evolution equation were introduced by Sapiro and Tannenbaum in [249]. Several mathematical properties were developed by the same authors in [250] and [251]. Angenent, Sapiro, and Tannenbaum gave the first existence and uniqueness proof of affine shortening in [21] and prove a theorem comparable to Grayson's theorem : they prove that a shape eventually becomes convex and thereafter evolves towards an ellipse before collapsing.

Mean curvature motion. In his famous paper entitled "Shapes of worn stones," Firey proposed a model for the natural erosion of stones on a beach [108]. He suggested that the rate of erosion of the surface of a stone was proportional to the Gaussian curvature of the surface, so that areas with high Gaussian curvature eroded faster than areas with lower curvature, and he conjectured that the final shape was a sphere. The first attempt at a mathematical definition of the mean curvature motion is found in Brakke [44]. Later in the book, we will discuss the Sethian's clever numerical implementation of the same equation [258]. Almgren, Taylor, Wang proposed a more general formulation of mean curvature motion that is applicable to crystal growth and, in general, to the evolution of anisotropic solids [5].

Chapter 13

Finite Difference Schemes for Curvature Motions

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 and equal $u_{i,j}$. Each one of the squares is called *pixel* (for "picture element").

13.1 Case of Mean curvature motion.

We start with the "Mean curvature motion" equation (M.C.M.) given by

$$\frac{\partial u}{\partial t} = |Du| curv(u) = \frac{u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy}}{u_x^2 + u_y^2}$$

In order to discretize this equation by finite differences we shall introduce an explicit scheme which uses a fixed stencil of 3×3 points to discretize the differential operators. We denote by $\Delta x = \Delta y$ the pixel width. From the PDE viewpoint Δx is considered as an infinitesimal length with respect to the image scale. Thus we shall write formulas containing $o(\Delta x)$. Numerically Δx is equal to 1, and the image scale ranges from 512 to 4096 and more. By the order 1 Taylor formula one can give the following discrete versions of the first derivatives u_x and u_y at a point (i,j) of the grid:

$$(u_x)_{i,j} = \frac{2(u_{i+1,j} - u_{i-1,j}) + u_{i+1,j+1} - u_{i-1,j+1} + u_{i+1,j-1} - u_{i-1,j-1}}{8\Delta x} + O(\Delta x^2);$$

$$(u_y)_{i,j} = \frac{2(u_{i,j+1} - u_{i,j-1}) + u_{i+1,j+1} - u_{i+1,j-1} + u_{i-1,j+1} - u_{i-1,j-1}}{8\Delta x} + O(\Delta x^2);$$

$$|Du_{i,j}| = ((u_x)_{i,j}^2 + (u_y)_{i,j}^2)^{\frac{1}{2}}.$$

Definition 13.1. *A discrete scheme approximating a differential operator is said to be consistent if, when the grid mesh Δx tends to zero, the discrete scheme tends to the differential operator.*

λ4	λ2	λ3
λ1	–4 λ0	λ1
λ3	λ2	λ4

Figure 13.1: A 3×3 stencil

Clearly the above discrete versions of the partial derivatives and of the gradient of u are consistent. When $|Du| \neq 0$, we can denote by ξ the direction orthogonal to the gradient of u. It is easily deduced from Definition 11.14 that

$$|Du|curv(u) = u_{\xi\xi}.$$

Exercise 13.1. Show this formula. ■

Defining θ as the angle between the x direction and the gradient, we have

$$\xi = (-\sin\theta, \cos\theta) = (\frac{-u_y}{\sqrt{u_x^2 + u_y^2}}, \frac{u_x}{\sqrt{u_x^2 + u_y^2}}), \text{ and}$$

$$u_{\xi\xi} = \sin^2(\theta)u_{xx} - 2\sin(\theta)\cos(\theta)u_{xy} + \cos^2(\theta)u_{yy}. \tag{13.1}$$

We would like to write $u_{\xi\xi}$ as a linear combination of the values of u on the fixed 3×3 stencil. Of course, the coefficients of the linear combination will depend on ξ. Since the direction of ξ is defined modulo π, we must assume by symmetry that the coefficients of points symmetrical with respect to the central point of the stencil are equal (see Figure 13.1.)

In order to ensure consistency with the differential operator $u_{\xi\xi}$, we must find $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4$, such that

$$(u_{\xi\xi})_{i,j} = \frac{1}{\Delta x^2}(-4\lambda_0 u_{i,j} + \lambda_1(u_{i+1,j} + u_{i-1,j}) + \lambda_2(u_{i,j+1} + u_{i,j-1})$$

$$+\lambda_3(u_{i-1,j-1} + u_{i+1,j+1}) + \lambda_4(u_{i-1,j+1} + u_{i+1,j-1})) + o((\Delta x)^2). \tag{13.2}$$

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 corresponding relations for the other points of the stencil. By substituting these relations into (13.2) and by using (13.1) one obtains four links between the five coefficients, namely

$$\begin{cases} \lambda_1(\theta) = 2\lambda_0(\theta) - \sin^2\theta \\ \lambda_2(\theta) = 2\lambda_0(\theta) - \cos^2\theta \\ \lambda_3(\theta) = -\lambda_0(\theta) + 0.5(\sin\theta\cos\theta + 1) \\ \lambda_4(\theta) = -\lambda_0(\theta) + 0.5(-\sin\theta\cos\theta + 1) \end{cases} \tag{13.3}$$

Exercise 13.2. Prove these four relations. ■

Thus, one degree of freedom is left for our coefficients : we can for example choose $\lambda_0(\theta)$ as we wish. This choice will be driven by stability and geometric invariance requirements. Denoting by $u^n_{i,j}$ an approximation of $u(i\Delta x, j\Delta x, n\Delta t)$ we can write our explicit scheme as

$$u^{n+1}_{i,j} = u^n_{i,j} + \Delta t(u^n_{\xi\xi})_{i,j} \tag{13.4}$$

Notice that this scheme can be rewritten as $u^{n+1}_{i,j} = \sum_{k,l=-1}^{1} \alpha_{k,l} u^n_{i+k,j+l}$ 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 L^∞ *stability* in this kind of schemes.

Lemma 13.2. *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 $\sum_{k,l=-1}^{1} \alpha_{k,l} = 1$. We say that the scheme is L^∞-*stable *if for all i, j,*

$$\min_{i,j} u(i,j) \leq T(u)_{i,j} \leq \max_{i,j} u(i,j).$$

Then the scheme is L^∞ stable 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 the L^∞ 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

$$(Tu)_{i,j} = \sum_{k \neq k_0, l \neq l_0}^{1} \alpha_{k,l} max + \alpha_{k_0,l_0} min = max + \alpha_{k_0,l_0}(min - max) > max,$$

which means that the L^∞ stability is violated.

□

Following this lemma, in order to guarantee the L^∞ stability in the scheme (13.4) we should 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 the links between these coefficients make it impossible to obtain these relations, except for the particular values of $\theta = (0, \frac{\pi}{4}, \frac{\pi}{2}, ...)$. Indeed, for θ in $[0, \frac{\pi}{4}]$,

$$\lambda_1 \geq \lambda_2 \qquad \text{and} \qquad \lambda_3 \geq \lambda_4$$

But

$$\lambda_2(\theta) \geq 0 \Rightarrow \lambda_0(\theta) \geq \frac{\cos^2(\theta)}{2}$$

$$\lambda_4(\theta) \geq 0 \Rightarrow \lambda_0(\theta) \leq \frac{1 - \sin(\theta)\cos(\theta)}{2}$$

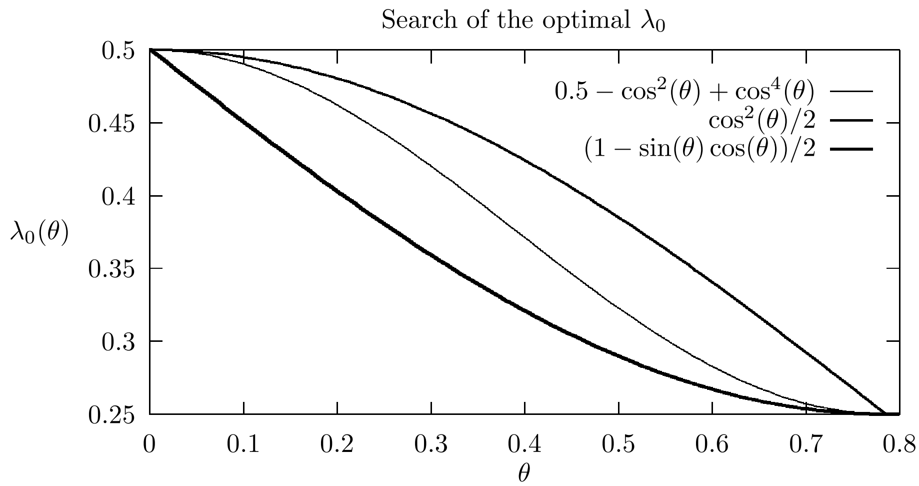

Figure 13.2: The middle curve represents the choice of the function λ_0 of Formula 13.6. The upper function represents the smallest possibility for $\lambda_0(\theta)$ securing $\lambda_2 \geq 0$ for all angles and the lower one represents the largest values of $\lambda_0(\theta)$ securing $\lambda_4(\theta) \geq 0$. Thus, it is not possible to satisfy simultaneously both conditions. The intermediate curve is the simplest trigonometric function which lies between these two bounds.

We cannot find $\lambda_0(\theta)$ satisfying both inequalities, since

$$\frac{\cos^2(\theta)}{2} \geq \frac{1 - \sin(\theta)\cos(\theta)}{2}$$

If we chose $\lambda_0(\theta) \geq \frac{\cos^2(\theta)}{2}$, $\lambda_4(\theta)$ would be significantly below zero. If we took $\lambda_0(\theta) \leq \frac{1-\sin(\theta)\cos(\theta)}{2}$, $\lambda_2(\theta)$ would be significantly below zero. Thus we shall choose λ_0 somewhere between both functions, so that λ_2 and λ_4 become only slightly negative. (see Figure 13.2.)

In addition, we can try to impose on λ_0 the following geometrical requirements

(i). Invariance by rotation of angle $\frac{\pi}{2}$

$$\lambda_0(\theta + \frac{\pi}{2}) = \lambda_0(\theta)$$

(ii). Purely one-dimensional 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 one-dimensional 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.

Two choices for the function $\lambda_0(\theta)$ using as basis the trigonometric polynomials were tested. The first one corresponds to an average of the boundary functions:

$$\lambda_0(\theta) = \frac{\cos^2(\theta) + 1 - \sin(\theta)cos(\theta)}{4} \tag{13.5}$$

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 trigonometric polynomial with least degree satisfying the above conditions and lying between both boundary functions is

$$\lambda_0(\theta)) = 0.5 - \cos^2(\theta)\sin^2(\theta) \tag{13.6}$$

The formulas of the other λ_i's are deduced using (13.3). 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 $|Du| = 0$, the direction of the gradient is unknown. Therefore the diffusion term $u_{\xi\xi}$ is not defined. We chose to replace this term by half the Laplacian. (The Laplacian is equal to the sum of the two second derivatives in orthogonal directions, whereas the diffusion term $u_{\xi\xi}$ is the second derivative in just one). However, other possibilities will be considered in Section 13.6. Summarizing, a consistent, almost L^∞ stable finite difference scheme for the mean curvature motion is (iterations start with u^0 as initial function)

1. If $|Du| \geq T_g$

$$u^{n+1} = u^n + \frac{\Delta t}{\Delta x^2}(-4\lambda_0 u_{i,j} + \lambda_1(u_{i+1,j} + u_{i-1,j}) + \lambda_2(u_{i,j+1} + u_{i,j-1})$$

$$+\eta_3(u_{i-1,j-1} + u_{i+1,j+1}) + \eta_4(u_{i-1,j+1} + u_{i+1,j-1})).$$

2. Otherwise,

$$u^{n+1} = u^n + \frac{1}{2}\frac{\Delta t}{\Delta x^2}(-4\lambda_0 u_{i,j} + u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1})$$

Two parameters have to be fixed in the previous algorithm:

• **The iteration step scale** $s := \frac{\Delta t}{\Delta x^2}$ has to be chosen as large as possible in order to reduce the number of iterations. However, $\frac{1}{2}$ is a natural upper bound for s. Indeed, consider the discrete image defined by $u^0_{i,j} = 0$ for all i, j, except for $i = j = 0$ where $u^0_{0,0} = 1$. Then the second formula yields $u^1_{0,0} = 1 - 2 * s$. If we want L^∞ stability to be ensured we must have $u^1(0,0) \geq 0$, which yields $s \leq 1/2$. Imposing this condition

$$\frac{\Delta t}{\Delta x^2} \leq \frac{1}{2} \tag{13.7}$$

it is an experimental observation that there is 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 13.3: 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.

13.2 FDS for AMSS

We will use the ideas developed in the above section. We rewrite the AMSS equation as

Figure 13.4: 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 13.3 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 13.6, if the initial image is noisy, the difference between both methods can be huge.

Figure 13.5: Iterated median filter with normalized scales 1, 2, 3, 4, 5. The scale normalization permits to compare very different schemes on the same images. Compare with Figure 13.4. The striking similarity of the results anticipates Theorem 14.7, according to which the application of the median filter is equivalent to a mean curvature motion.

$$\frac{\partial u}{\partial t} = (|Du|^3 curv(u))^{\frac{1}{3}} = (u_y^2 u_{xx} - 2u_x u_y u_{xy} + u_x^2 u_{yy})^{\frac{1}{3}} \tag{13.8}$$

We remark that $|Du|^3 curv(u) = |Du|^2 u_{\xi\xi}$ 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 $|Du|^2$. We choose $\lambda_0(\theta)$ given by (13.5) because it corresponds to a trigonometric polynomial of degree two and then multiplying it by $|Du|^2$ the coefficients $\eta_i = |Du|^2\lambda_i$, $i = 0, 1, 2, 3, 4$, are polynomials of degree two with respect to u_x and u_y. Indeed, we obtain for $\theta \in [0, \frac{\pi}{4}]$

$$(|Du|^2 u_{\xi\xi})_{i,j} = \frac{1}{\Delta x^2}(-4\eta_0 u_{i,j} + \eta_1(u_{i+1,j} + u_{i-1,j}) + \eta_2(u_{i,j+1} + u_{i,j-1})$$

$$+\eta_3(u_{i-1,j-1} + u_{i+1,j+1}) + \eta_4(u_{i-1,j+1} + u_{i+1,j-1})) + O(\Delta x^2)$$

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

$$u_{i,j}^{n+1} = u_{i,j}^n + \Delta t(|Du^n|^2 u_{\xi\xi}^n)_{i,j}^{\frac{1}{3}} \tag{13.9}$$

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. Contrast invariance can only be obtained asymptotically by taking a very small time step Δt. The experimental results presented in Figures 13.3 and **??** have been obtained by using these schemes with $\Delta t = 0.1$ in the case of mean curvature motion and $\Delta t = 0.01$ in the case of affine curvature motion. Indeed, while experimental stability is achieved with $\Delta t \leq 0.1$, the experimental affine invariance needs $\Delta t < 0.05$ (see Figure **??**.)

13.3 IL MANQUE UNE EXPERIENCE AMSS SUR L'INVARIANCE AFFINE!

13.4 Numerical normalization of scale.

(or Relation between scale and the number of iterations).

The case of the curvature motion. Setting the distance between pixels Δx to 1, the scale achieved with N iterations is simply $N \times \Delta t$. Now, the scale t associated with the PDE is somewhat arbitrary : It has no geometric meaning. In order to get it, we need a rescaling $T \to t(T)$ which we will call scale normalization.

A good way to perform this scale normalization is to define the correspondence $t(T)$ as the time for which a circle with initial radius T vanishes under curvature motion. Such a circle moves at a speed equal to its curvature, which is the inverse of its radius. Thus have for a disk with radius $R(t)$

$$\frac{dR(t)}{dt} = -\frac{1}{R(t)}$$

which yields

$$\frac{1}{2}(R^2(0) - R^2(t)) = t.$$

Exercise 13.3. Check this relation! ■

The disk disappears when $R(t) = 0$, that is, at scale $T = R^2(0)/2$. This last relation gives a scale normalization: In order to arrive at the normalized scale T (at which any disk with radius less or equal to T vanishes), we have to evolve the PDE at $t = N\Delta t = T^2/2$. This fixes the number of needed iterations as

$$N = T^2/2\Delta t.$$

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 disappearance time is therefore $t = \frac{3}{4}R^{\frac{4}{3}}$. As for the curvature motion, we define the normalized scale T as the one at which a disk with radius T vanishes. In order to achieve this scale T, the needed number of iterations is

$$N = \frac{3}{4\Delta t}T^{\frac{4}{3}}.$$

Exercise 13.4. Check the last two formulas! ■

13.5 Contrast invariance and the level set extension

Both schemes (M.C.M and A.M.S.S) presented above are not numerically contrast invariant. We have seen that a contrast operator cannot create new gray levels (Exercise 7.22.) Now, starting with a binary image u^0 and applying a scheme defined by such a formula as

$$u^{n+1} = u^n + \Delta t(...)$$

does not ensure that u^{n+1} will be also a binary image.

A natural idea to overcome this problem is the following. Starting with a binary image (with values 0 and 1): apply the scheme until the expected scale is achieved, then threshold the obtained image at $\lambda = \frac{1}{2}$. This of course works only for binary images. However, the level set extension (see Section 7.3) gives us the key to extend this to general images.

The contrast invariance can be fully obtained by first applying the finite difference scheme on each level set (considered as a binary image) separately. Then by the superposition principle the evolved image is computed from the evolved level sets. The procedure is the following :

Algorithm starting with an image u_0 and evolving it to $u(t, \mathbf{x})$ by curvature motion

For each $\lambda \in [0, 255]$, in increasing order:

- Let $v_\lambda(\mathbf{x})$ be the characteristic function of $\mathcal{X}_\lambda u_0$. (This function is equal to 1 inside the level set and to 0 outside.)
- Apply to v_λ the MCM or AMSS FDS-scheme until scale t. This yields the images $w_\lambda(t, .)$.
- Set $u(t, \mathbf{x}) = \lambda$ at each point $(t, \mathbf{x})$ where $w_\lambda(t, \mathbf{x}) \geq 0.5$.

13.6 Problems at extrema

For MCM and AMSS we raised the question of performing numerically the equation when $|Du| = 0$. For MCM the right hand part of the equation is simply not defined. For AMSS one can set by continuity as $Du \to 0$, $(|Du^n|^2 u^n_{\xi\xi})^{\frac{1}{3}}_{i,j} = 0$. Now, numerically, this would imply that isolated black or white extrema will not evolve by the equation. We know that this is simply wrong, since small sets collapse by curvature motion.

In short, FDS for MCM and AMSS are not consistent with the equation at extrema. In Figure 13.6, we added to an image a strong “salt and pepper” noise. More than one fourth of the pixels have been given a uniform random value in $[0, 255]$ and most of them have become local extrema. Not only these values do not evolve but they contaminate their neighboring pixels. There are easy ways to avoid this spurious effect :

- One can first zoom by 2 the image by duplicating pixels. This, however, multiplies by 16 the number of computations.
- One can first remove pixels extrema with diameter k since they must anyway disappear by the equation at normalized scale $\frac{k}{2}$.
- One can use the level set method. This multiplies the number of computations by the initial number of gray-levels.

All of these solutions are efficient, as shown in Figure 13.6.

Figure 13.6: Various implementations of curvature motion on a noisy image. Top left : image with 40% pixels replaced by a uniform random value in $[0, 255]$. Top right: application of the finite difference scheme (FDS) at normalized scale 3. On the lines 2 to 4, we see various solutions to the disastrous diffusion of extrema. On the left the image is processed at normalized scale 1 and on the right at normalized scale 3. Second line: FDS applied on the image previously zoomed by a factor 2; third line: 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). Fourth line: FDS applied separately on each level set and application of the threshold superposition principle. The third scheme offers a good speed-quality compromise.

13.7 Conclusion

We have seen that standard finite difference schemes are easy to implement but cannot 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. No full contrast invariance. For instance FDS create new grey levels and blur edges. Also, a spurious diffusion occurs around the image extrema. However this last problem was dealt with efficiently in the previous section. The full contrast invariance has been restored by the level set extension of the numerical schemes.

3. The worst drawback of FDS is the lack of Euclidean or affine invariance which can be only approximately obtained by grid local schemes. A much more clever strategy to achieve full invariance is to evolve all level curves of the image and the reconstruct it. This is the aim of Section 16.4, but we have already seen in Chapter 4 how to evolve curves by curvature.

13.8 Comments and references

Difference schemes for the curvature motion and the AMSS The presented difference scheme follows mainly [130], improved in Alvarez et al. [14]. 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 [2], Kacur and Mikula [162, **?**]. Weickert and the Ütrecht school [216, 278, 1, 284] address many aspects of implementation of nonlinear scale spaces, namely speed, parallelism and robustness. Crandall and Lions [81] also proposed a finite difference scheme for mean curvature motion, valid in any dimension. Sethian's book [257] 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").

Chapter 14

Asymptotic Behavior of SMTCII Operators, Dimension Two

As we know by Theorem 8.15, a function operator on $\mathcal{F}$ is contrast and translation invariant and standard monotone if and only if it has a sup-inf, or equivalently an inf-sup form

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

where $\mathcal{B}$ is a standard subset of $\mathcal{L}$. In case we require such operators to be isotropic and local, it is enough to take for $\mathcal{B}$ any set of sets invariant by rotation and contained in some $B(0, M)$ by Proposition 8.11.

We will see, however, that such operators fall into a few classes when we make them more and more local. To see this, we introduce a scale parameter $0 < h \leq 1$ and define the scaled operators T_h by

$$T_h u(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y}).$$

We will prove that in the limit, as h tends to zero, the action of T_h on smooth functions is not as varied as one might expect given the possible sets of structuring elements. As an example, we will show that if T_h is a scaled median operator, then

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = h^2 C |Du(\mathbf{x})| \mathrm{curv}(u)(\mathbf{x}) + o(h^2),$$

where the constant C depends only on the function k used to define the median operator. Thus, the operator $|Du|\mathrm{curv}(u)$ plays the same role for the weighted median filters, as the Laplacian Δu does for linear operators. In short, we shall get contrast invariant analogues of Theorem 2.2.

14.1 Asymptotic behavior theorem in $\mathbb{R}^2$

A simple real function will describe the asymptotic behavior of any local contrast invariant filter.

Definition 14.1. *Let T be a SMTCII local operator. Consider the real function $H(s)$, $s \in \mathbb{R}$,*

$$H(s) = T[x + sy^2](0), \tag{14.1}$$

where $T[x+sy^2]$ denotes "Tu with $u(x,y) = x+sy^2$." H is called the structure function *of T.*

Notice that $u(x,y) = x + sy^2$ is not in $\mathcal{F}$, so we use here the extension described in the introduction. The function $H(s)$ is well defined by the result of Exercise 8.14.

Proposition 14.2. *The structure function of a local SMTCII operator is nondecreasing, Lipschitz, and satisfies for $h > 0$,*

$$T_h[x + sy^2](0) = hT[x + hsy^2](0) = hH(hs), \tag{14.2}$$

$$T_h[x](0) = hT[x](0) = hH(0) \tag{14.3}$$

Proof. Take T in the inf-sup form with $\mathcal{B} \subset B(0,M)$, $0 \le M < 1$.

Since T is monotone, H is a nondecreasing function. Let $B \in \mathcal{B}$ be one of the structuring elements that define T and write $x + s_1y^2 = x + s_2y^2 + (s_1 - s_2)y^2$. Then

$$\sup_{(x,y)\in B}(x + s_1y^2) \le \sup_{(x,y)\in B}(x + s_2y^2) + |s_2 - s_1|M^2,$$

since B is contained in $D(0,M)$. By taking the infimum over $B \in \mathcal{B}$ of both sides and using the definition of H, we see that

$$H(s_1) - H(s_2) \le |s_1 - s_2|M^2.$$

By interchanging s_1 and s_2 in this last inequality, we deduce the Lipschitz relation

$$|H(s_1) - H(s_2)| \le |s_1 - s_2|M^2. \tag{14.4}$$

□

Theorem 14.3. *Let T be a local SMTCII operator and T_h, $1 \ge h > 0$ its scaled versions. Call H its structure function. Then, for any C^2 function $u : \mathbb{R}^2 \to \mathbb{R}$,*

$$T_hu(\mathbf{x}) - u(\mathbf{x}) = hH(0)|Du(\mathbf{x})| + o(h^2).$$

Proof. By Propositions 8.9, 8.11 and 8.13, we can take T in the inf-sup form and assume, for all B in $\mathcal{B}$, that $B \subset B(0,M)$ and that $\mathcal{B}$ is invariant under rotations. Set $p = |Du(\mathbf{x})|$. By a suitable rotation, and since T is isotropic, we may assume that $Du(\mathbf{x}) = (|Du(\mathbf{x})|, 0)$, and the first-order Taylor expansion of u in a neighborhood of $\mathbf{x}$ can be written as

$$u(\mathbf{x} + \mathbf{y}) = u(\mathbf{x}) + px + O(\mathbf{x}, |\mathbf{y}|^2), \tag{14.5}$$

where $\mathbf{y} = (x,y)$ and $|O(\mathbf{x}, |\mathbf{y}|^2)| \le C|\mathbf{y}|^2$ for $\mathbf{y} \in D(0,M)$. Hence,

$$u(\mathbf{x} + h\mathbf{y}) - u(\mathbf{x}) \le phx + Ch^2|\mathbf{y}|^2 \quad \text{and} \quad phx \le u(\mathbf{x} + h\mathbf{y}) - u(\mathbf{x}) + Ch^2|\mathbf{y}|^2 \tag{14.6}$$

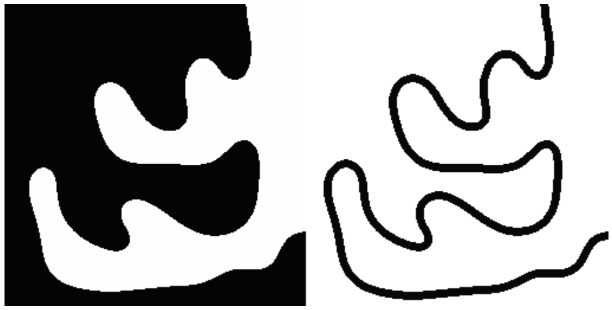

Figure 14.1: The result of smoothing with an erosion is independent of the curvature of the level lines. Left: image of a simple shape. Right: difference of this image and its eroded image. Note that the width of the difference is constant. By Theorem 14.3, all filters such that $H(0) \neq 0$ perform such an erosion, or a dilation.

for all $\mathbf{y} \in D(0, M)$. Since $hB \subset D(0, hM)$, we see from the first inequality of (14.6) that

$$\sup_{\mathbf{y}\in B} u(\mathbf{x}+h\mathbf{y}) - u(\mathbf{x}) \leq \sup_{\mathbf{y}\in B}[phx] + \sup_{\mathbf{y}\in B} Ch^2|\mathbf{y}|^2 = hp \sup_{\mathbf{y}\in B}[x] + CM^2h^2.$$

This implies that

$$T_h u(\mathbf{x}) - u(\mathbf{x}) \leq hp \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B}[x] + CM^2h^2$$

for $0 < h \leq 1$, and since $\inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B}[x] = T[x](0) = H(0)$, we see that

$$T_h u(\mathbf{x}) - u(\mathbf{x}) \leq hpH(0) + CM^2h^2.$$

The same argument applied to the second inequality of (14.6) shows that

$$hpH(0) \leq T_h u(\mathbf{x}) - u(\mathbf{x}) + CM^2h^2,$$

so $|T_h u(\mathbf{x}) - u(\mathbf{x}) - hpH(0)| \leq CM^2h^2$. Since $p = |Du(\mathbf{x})|$, we see that

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = hH(0)|Du(\mathbf{x})| + O(\mathbf{x}, h^2),$$

which proves the result in case $p \neq 0$. $\square$

Interpretation. Theorem 14.3 tells us that the behavior of local contrast invariant operators T_h depends, for small h, completely on the action of T on the test function $u(x, y) = x$. Assume $H(0) = H(0) \neq 0$. When $h \to 0$, T acts like a dilation by a disk $D(0, h)$ if $H(0) > 0$ and like an erosion with $D(0, h)$ if $H(0) < 0$ (see Proposition 9.6). Thus, if $H(0) \neq 0$, there is no need to define T with a complicated set of structuring elements. Asymptotically these operators are either dilations or erosions, and these can be defined with a single structuring element, namely, a disk. Exercise 14.4 gives the more general PDE obtained when T is local but not isotropic.

14.1.1 The asymptotic behavior of T_h when $T[x](0) = 0$

If $H(0) = T[x](0) = 0$, then Theorem 14.3 is true but not very interesting. On the other hand, operators for which $T[x](0) = 0$ are interesting. If we consider $Tu(\mathbf{x})$ to be a kind of average of the values of u in a neighborhood of $\mathbf{x}$, then assuming that $T[x](0) = 0$ makes sense. This means, however, that we must consider the next term in the expansion of $T_h u$; to do so we need to assume that u is C^3. This is the content of the next theorem, which is the main theoretical result of the chapter. The proof is more involved than that of Theorem 14.3, but at the macro level, they are similar. We start with some precise Taylor expansion of u.

Lemma 14.4. *Let $u(\mathbf{y})$ be C^3 around some point $\mathbf{x} \in \mathbb{R}^2$. By using adequate Euclidean coordinates $\mathbf{y} = (x, y)$, we can expand u in a neighborhood of $\mathbf{x}$ as*

$$u(\mathbf{x} + h\mathbf{y}) = u(\mathbf{x}) + h(px + ahx^2 + bhy^2 + chxy) + R(\mathbf{x}, h\mathbf{y}), \tag{14.7}$$

where $|R(\mathbf{x}, h\mathbf{y})| \le Ch^3$ for all $\mathbf{x} \in K$, $\mathbf{y} \in D(0, M)$ and $0 \le h \le 1$.

Proof. Set $p = |Du(\mathbf{x})|$. We define the local coordinate system by taking $\mathbf{x}$ as origin and $Du(\mathbf{x}) = (p, 0)$. Relation (14.7) is nothing but a Taylor expansion where R can be written as

$$R(\mathbf{x}, h\mathbf{y}) = \left(\int_0^1 (1-t)^2 D^3 u(\mathbf{x} + th\mathbf{y})\, \mathrm{d}t \right) h^3 \mathbf{y}^{(3)}.$$

The announced estimate follows because the function $\mathbf{x} \mapsto \|D^3 u(\mathbf{x})\|$ is continuous and thus bounded on the compact set $K + D(0, M)$. $\square$

Theorem 14.5. *Let T be a local SMTCII operator on $\mathcal{F}$ whose structure function H satisfies $H(0) = 0$. Then for every C^3 function u on $\mathbb{R}^2$,*

(i) On every compact set $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$,

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = h|Du(\mathbf{x})| H\Big(\frac{1}{2} h \operatorname{curv}(u)(\mathbf{x})\Big) + O(\mathbf{x}, h^3),$$

where $|O(\mathbf{x}, h^3)| \le C_K h^3$ for some constant C_K that depends only on u and K.

(ii) On every compact set K in $\mathbb{R}^2$,

$$|T_h u(\mathbf{x}) - u(\mathbf{x})| \le C'_K h^2$$

where the constant C'_K depends only on u and K.

Proof. We take T in the inf-sup form and $\mathcal{B}$ bounded by $D(0, M)$ and isotropic. Let us use the Taylor expansion (14.7). For $0 < h \le 1$,

$$u(\mathbf{x} + h\mathbf{y}) = u(\mathbf{x}) + h(px + ahx^2 + bhy^2 + chxy) + R(\mathbf{x}, h\mathbf{y}),$$

and so for any $B \in \mathcal{B}$,

$$\sup_{\mathbf{y}\in B} u(\mathbf{x}+h\mathbf{y}) \le u(\mathbf{x}) + h \sup_{\mathbf{y}\in B}[u_h(x,y)] + \sup_{\mathbf{y}\in B} |R(\mathbf{x},h\mathbf{y})|.$$

Thus,

$$T_h u(\mathbf{x}) \le u(\mathbf{x}) + hT[u_h(x,y)](0) + \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B} |R(\mathbf{x},h\mathbf{y})|, \tag{14.8}$$

where $u_h(x,y) = px + ahx^2 + bhy^2 + chxy$ and $\mathbf{y} = (x,y)$. Now let K be an arbitrary compact set. From Lemma 14.4 we deduce that

$$T_h u(\mathbf{x}) \le u(\mathbf{x}) + hT[u_h(x,y)](0) + Ch^3 \tag{14.9}$$

for all $\mathbf{x} \in K$. The same analysis shows that

$$u(\mathbf{x}) \le T_h u(\mathbf{x}) + hT[u_h(x,y)](0) + Ch^3, \tag{14.10}$$

and we conclude that

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = hT[u_h(x,y)](0) + O(\mathbf{x},h^3) \tag{14.11}$$

for all $\mathbf{x} \in K$ where $|O(\mathbf{x},h^3)| \le C_K h^3$. Relation (14.11) reduces the proof to an analysis of $Tu_h(0)$.

Step 1: Estimating $\boldsymbol{Tu_h(0)}$. If $\mathbf{x} \in K$ and $\mathbf{y} = (x,y) \in B$, then $|\mathbf{y}| \le M$ and

$$px - h(|a|+|b|+|c|)M^2 \le u_h(x,y) \le px + h(|a|+|b|+|c|)M^2.$$

We write this as

$$px - \frac{hM^2}{2}\|D^2u(\mathbf{x})\| \le u_h(x,y) \le px + \frac{hM^2}{2}\|D^2u(\mathbf{x})\|.$$

By assumption $T[x](0) = 0$ (hence $T[px](0) = 0$), so after applying T to the inequalities, we see that

$$|T[u_h(x,y)](0)| \le \frac{hM^2}{2}\|D^2u(\mathbf{x})\|. \tag{14.12}$$

This and equation (14.11) show that

$$|T_h u(\mathbf{x}) - u(\mathbf{x})| \le \frac{h^2M^2}{2}\|D^2u(\mathbf{x})\| + C_K h^3 \tag{14.13}$$

for $\mathbf{x} \in K$ and $0 < h \le 1$. This proves part (ii). Let us now prove (i). We just recall the meaning of p and b, namely $b = (1/2)\mathrm{curv}(u)(\mathbf{x})|Du(\mathbf{x})|$ and $p = |Du(\mathbf{x})|$. Those terms are the only terms appearing in the main announced result (i). So the proof of (i) consists of getting rid of a and c in the asymptotic expansion $(Tu_h)(0)$. This elimination is performed in Steps 2 and 3.

Step 2: First reduction. We now focus on proving (i), and for this we assume that $p = |Du(\mathbf{x})| \ne 0$. Define $C = (|a|+|b|+|c|)M^2$. By Step 1, for every $B \in \mathcal{B}$, we see that

$$\sup_{\mathbf{y}\in B} u_h(x,y) \ge \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B} u_h(x,y) = T[u_h(x,y)](0) \ge -Ch.$$

If $\mathbf{y} = (x, y) \in B$ and $x < -2Ch/p$, then

$$u_h(x, y) = px + ahx^2 + bhy^2 + chxy < -2Ch + h(|a| + |b| + |c|)M^2 = -Ch.$$

Thus, if we let $C' = 2C/p$, then for any $B \in \mathcal{B}$ we have

$$\sup_{\mathbf{y}\in B} u_h(x, y) = \sup_{\mathbf{y}\in B\cap\{(x,y)|x\geq -C'h\}} u_h(x, y)$$

Step 3: Second reduction. Since $T[u_h(x, y)](0) \leq Ch$ (Step 1), it is not necessary to consider sets B for which $\sup_{\mathbf{y}\in B} u_h(x, y) \geq Ch$. If $\sup_{\mathbf{y}\in B} u_h(x, y) \leq Ch$, then for all $(x, y) \in B$

$$px + ahx^2 + bhy^2 + chxy \leq Ch,$$

and hence

$$x \leq \frac{1}{p}(Ch + (|a| + |b| + |c|)M^2 h) \leq \frac{2Ch}{p} = C'h.$$

This means that we can write

$$T[u_h(x, y)](0) = \inf_{B\in\mathcal{B}, B\subset\{(x,y)|x\leq C'h\}} \sup_{\mathbf{y}\in B} u_h(x, y), \tag{14.14}$$

and by the result of Step 2,

$$T[u_h(x, y)](0) = \inf_{B\in\mathcal{B}, B\subset\{(x,y)|x\leq C'h\}} \sup_{\mathbf{y}\in B\cap\{(x,y)|x\geq -C'h\}} u_h(x, y). \tag{14.15}$$

This relation is true if we replace $u_h(x, y)$ with $px + bhy^2$ and leads directly to the inequality

$$\begin{aligned} T[u_h(x, y)](0) &\leq T[px + bhy^2](0) \\ &+ h \inf_{B\in\mathcal{B}, B\subset\{(x,y)|x\leq C'h\}} \sup_{\mathbf{y}\in B\cap\{(x,y)|x\geq -C'h\}} |ax^2 + cxy| \end{aligned}$$

and, by interchanging $u_h(x, y)$ and $px + bhy^2$, to the equation

$$T[u_h(x, y)](0) = T[px + bhy^2](0) + \varepsilon(x, y). \tag{14.16}$$

The error term is

$$|\varepsilon(x, y)| \leq h^3|a|C'^2 + h^2|c|C'M.$$

Step 4: Conclusion. We now return to equation (14.11),

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = hT[u_h(x, y)](0) + O(\mathbf{x}, h^3),$$

and replace $T[u_h(x, y)](0)$ with $T[px + bhy^2](0) + \varepsilon(x, y)$ to obtain

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = hT[px + bhy^2](0) + h\varepsilon(x, y) + O(\mathbf{x}, h^3).$$

By definition $H(s) = T[x + sy^2](0)$, so the last equation can be written as

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = hpH(bh/p) + h\varepsilon(x, y) + O(\mathbf{x}, h^3),$$

or, by replacing p and b with $|Du(\mathbf{x})|$ and $(1/2)\mathrm{curv}(u)(\mathbf{x})|Du(\mathbf{x})|$, as

$$T_h u(\mathbf{x}) - u(\mathbf{x}) = h|Du(\mathbf{x})|H\Big(h\frac{1}{2}\mathrm{curv}(u)(\mathbf{x})\Big) + h\varepsilon(x,y) + O(\mathbf{x},h^3). \quad (14.17)$$

To finish the proof, we must examine the error term ε to establish a uniform bound on compact sets where $Du(\mathbf{x}) \neq 0$. Thus, let K be any compact subset of $\mathbb{R}^2$ such that $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$. For $\mathbf{y} \in D(0,M)$ (hence for $\mathbf{y} \in B \in B \in \mathcal{B}$), we have $|\varepsilon(x,y)| \leq h^3|a|C'^2 + h^2|c||C'|M$. Now, $|a|C'^2 + |c||C'|M$ is a continuous function of $Du(\mathbf{x})$ and $D^2u(\mathbf{x})$ at each point $\mathbf{x}$ where $Du(\mathbf{x}) \neq 0$. Since u is C^3, all of the functions on the right-hand side of this relation are continuous on K. Thus there is a constant C'_K that depends only on u and K such that $|\varepsilon(x,y)| \leq h^2 C'_K$. By combining and renaming the constants C_K and C'_K, this completes the proof of (i). □

Exercise 14.1. Returning to the meaning in the preceding proof of a, b, c, p and C' in term of derivatives of u, check that $|a|C'^2 + |c||C'|M$ is, as announced, a continuous function at each point where $Du(\mathbf{x}) \neq 0$.

14.2 Median filters and curvature motion in $\mathbb{R}^2$

Recall that the median filter, Med_k, defined in Chapter 10 can be written by Proposition 10.6 as

$$\mathrm{Med}_k u(\mathbf{x}) = \sup_{B\in\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}), \quad (14.18)$$

where $\mathcal{B} = \{B \in \mathcal{M} \mid |B|_k = 1/2\}$. The first example we examine is $k = \mathbf{1}_{D(0,1)}/\pi$. This function is not separable in the sense of Definition 10.7. So, by Proposition 10.8, $\mathrm{Med}_k u = \mathrm{Med}_k^- u$ and the median also has the inf-sup form

$$\mathrm{Med}_k u(\mathbf{x}) = \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}). \quad (14.19)$$

From Proposition 8.11 follows that the set of structuring elements $\mathcal{B}' = \{B \in \mathcal{B} \mid B \subset D(0,1)\}$ generates the same median filter. Thus we assume in what follows that $B \subset D(0,1)$. There is one more point that needs to be clarified, and we relegate it to the next exercise.

Exercise 14.2. The scaled median filter $(\mathrm{Med}_k)_h$, $h < 1$, is defined by

$$(\mathrm{Med}_k)_h u(\mathbf{x}) = \inf_{B\in h\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}). \quad (14.20)$$

At first glance, it is not clear that this is a median filter, but, in fact, it is: Show that $(\mathrm{Med}_k)_h = \mathrm{Med}_{k_h}$, where $k_h = \mathbf{1}_{D(0,h)}/\pi h^2$. ■

The actions of median filters and comparisons of these filters with other simple filters are illustrated in Figures 14.1, 14.2, 14.4, 14.5, and 14.6. Everything is now in place to investigate the asymptotic behavior of the scaled median filter Med_{k_h}, which is represented by

$$\mathrm{Med}_{k_h} u(\mathbf{x}) = \inf_{B\in h\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}),$$

where $h\mathcal{B} = \{B \mid |B|_{k_h} = 1/2, B \subset D(0,h)\}$. The main result of this section, Theorem 14.7, gives an infinitesimal interpretation of this filter. We know that the median is an SMTCII operator, and it is local in our case. The proof of the next lemma is quite special, having no immediate generalization to $\mathbb{R}^N$.

Lemma 14.6.
$$\mathrm{Med}_k[x+sy^2](0) = \frac{s}{3} + O(|s|^3).$$

Proof. Represent Med_k by $\mathrm{Med}_k u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda u\}$. Then

$$\mathrm{Med}_k[x+sy^2](0) = \sup\{\lambda \mid 0 \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda [x+sy^2]\,\}.$$

By definition, $0 \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda[x+sy^2]$ if and only if $|\mathcal{X}_\lambda[x+sy^2]|_k \geq 1/2$. This implies that $\mathrm{Med}_k[x+sy^2](0) = m(s)$, where $|\mathcal{X}_{m(s)}[x+sy^2]|_k = 1/2$, and this is true if and only if the graph of $x+sy^2 = m(s)$ divides $D(0,1)$ into two sets that have equal area. Of course, we are only considering small s, say $|s| \leq 1/2$. The geometry of this situation is illustrated in Figure 14.3. The signed area between the y-axis and the parabola $P(s)$ for $|y| \leq 1$ is

$$\int_{-1}^{1} (m(s) - sy^2)\,\mathrm{d}y = 2m(s) - \frac{2s}{3}.$$

Thus, $m(s)$ is the proper value if and only if

$$m(s) - \frac{s}{3} = \ \mathrm{Area}(ABE), \tag{14.21}$$

where ABE denotes the curved triangle bounded by the parabola, the circle, and the line $y=-1$. This area could be computed, but it is sufficient to bound it by $\mathrm{Area}(ABCD)$. The length of the base AB is $|m(s)-s|$, and an easy computation shows that the length of the height BC is less than $(m(s)-s)^2$. This and (14.21) imply that

$$\left| m(s) - \frac{s}{3} \right| \leq |m(s)-s|^3.$$

From this we conclude that $m(s) = s/3 + O(|s|^3)$, which proves the lemma. $\square$

Theorem 14.7. *If $u : \mathbb{R}^2 \to \mathbb{R}$ is C^2, then we have the following expansions:*

(*i*) *On every compact set $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$,*

$$\mathrm{Med}_{k_h} u(\mathbf{x}) = u(\mathbf{x}) + \frac{1}{6}|Du(\mathbf{x})|\mathrm{curv}(u)(\mathbf{x})h^2 + O(\mathbf{x}, h^3),$$

where $|O(\mathbf{x},h^3)| \leq C_K h^3$ for some constant C_K that depends only on u and K.

(*ii*) *On every compact set K in $\mathbb{R}^2$,*

$$|\mathrm{Med}_{k_h} u(\mathbf{x}) - u(\mathbf{x})| \leq C_K h^2$$

where the constant C_K depends only on u and K.

Proof. We have shown (or it is immediate) that the operator $T_h = \mathrm{Med}_{k_h}$ satisfies all of the hypotheses of Theorem 14.5. In particular, $H(0) = \mathrm{Med}_k[x+$

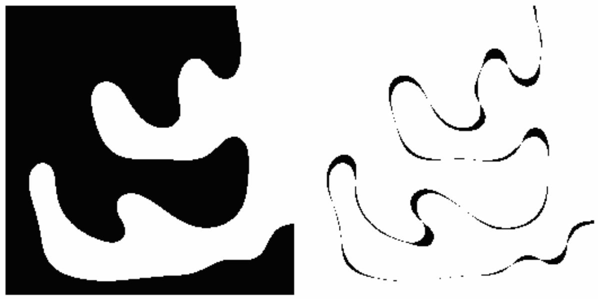

Figure 14.2: Median filter and the curvature of level lines. Smoothing with a median filter is related to the curvature of the level lines. Left: image of a simple shape. Right: difference of this image with itself after it has been smoothed by one iteration of the median filter. We see, in black, the points which have changed. The width of the difference is proportional to the curvature, as indicated by Theorem 14.7.

$sy^2](0) = 0$ by Lemma 14.6. Also by Lemma 14.6, $H(s) = s/3 + O(|s|^3)$. This means that we have

$$H\Big(\frac{1}{2}h\,\mathrm{curv}(u)(\mathbf{x})\Big) = \frac{1}{6}h\,\mathrm{curv}(u)(\mathbf{x}) + O(h^3\ |\mathrm{curv}(u)(\mathbf{x})|^3).$$

The first result is now read directly from Theorem 14.5(i). Relation (ii) follows immediately from Theorem 14.5(ii). □

Our second example is called the Catté–Dibos–Koepfler scheme. It involves another application of Theorem 14.5.

Theorem 14.8. *Let $\mathcal{B}$ be the set of all line segments of length 2 centered at the origin of $\mathbb{R}^2$. Define the operators SI_h and IS_h by*

$$SI_h u(\mathbf{x}) = \sup_{B\in h\mathcal{B}} \inf_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}) \qquad \text{and} \qquad IS_h u(\mathbf{x}) = \inf_{B\in h\mathcal{B}} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}).$$

If $u : \mathbb{R}^2 \to \mathbb{R}$ is C^2 and $|Du(\mathbf{x})| \neq 0$, then

$$\frac{1}{2}(IS_h + SI_h)u(\mathbf{x}) = u(\mathbf{x}) + h^2\frac{1}{4}\mathrm{curv}(u)(\mathbf{x})|Du(\mathbf{x})| + O(h^3).$$

Proof. The first step is to compute the action of the operators on $u(x, y) = x + sy^2$. Define $H(s) = IS[x + sy^2](0)$ and write $(x, y) = (r\cos\theta, r\sin\theta)$. Then

$$H(s) = \inf_{-\frac{\pi}{2}\le\theta\le\frac{\pi}{2}} \sup_{-1\le r\le 1} (r\cos\theta + sr^2\sin^2\theta).$$

For $s \ge 0$ and $r \ge 0$, the function $r \mapsto r\cos\theta + sr^2\sin^2\theta$ is increasing. Hence,

$$H(s) = \inf_{-\frac{\pi}{2}\le\theta\le\frac{\pi}{2}} (\cos\theta + s\sin^2\theta) = s$$

for sufficiently small s, say, $s < 1/2$. If $s \le 0$, then $H(0) = 0$, since

$$0 \le \sup_{-1\le r\le 1} (r\cos\theta + sr^2\sin^2\theta) \le \cos\theta.$$

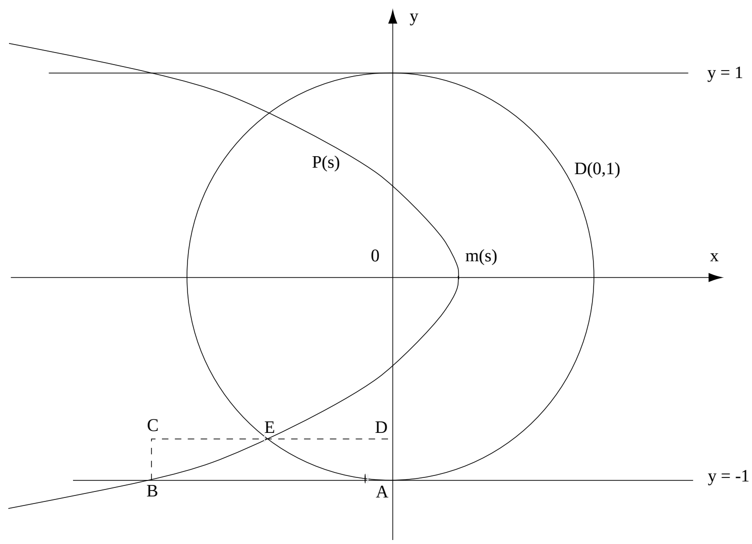

Figure 14.3: When s is small, the parabola $P(s)$ with equation $x + sy^2 = m$ divides $D(0,1)$ into two components. The median value $m(s)$ of $x + sy^2$ on $D(0,1)$ simply is the value m for which these two components have equal area.

If $H^-(s) = SI[x + sy^2](0)$, then it is an easy check that $H^-(s) = -H(-s)$. Thus we have

$$H(s) = \begin{cases} s, & \text{if } s \geq 0; \\ 0, & \text{if } s < 0; \end{cases} \qquad \text{and} \qquad H^-(s) = \begin{cases} 0, & \text{if } s \geq 0; \\ s, & \text{if } s < 0. \end{cases}$$

Thus, $H(s) + H^-(s) = s$ for all small s. Since $H(0) = H^-(0) = 0$, the conclusions of Theorem 14.5 apply. By applying Theorem 14.5(i) to IS_h and SI_h and adding, we have

$$\begin{aligned}(IS_h + SI_h)u(\mathbf{x}) &= 2u(\mathbf{x}) + h(H + H^-)\Big(\frac{h}{2}\mathrm{curv}(u)(\mathbf{x})\Big) + O(h^3) \\ &= 2u(\mathbf{x}) + \frac{h^2}{2}\mathrm{curv}(u)(\mathbf{x}) + O(h^3).\end{aligned}$$

Dividing both sides by two gives the result. $\square$

Exercise 14.3. Prove the relation $H^-(s) = -H(-s)$ used in the above proof. ■

14.3 Exercises

Exercise 14.4. Assume that T is a local translation and contrast invariant operator, but not necessarily isotropic. Show that

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + hT[Du(\mathbf{x}) \cdot \mathbf{x}](0) + O(h^2). \ \blacksquare$$

Exercise 14.5. Let $\mathcal{B}$ be the set of all rectangles in the plane with length two, width $\delta < 1$, and centered at the origin. Define the operators IS_h and SI_h by

$$IS_h u(\mathbf{x}) = \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}) \qquad \text{and} \qquad SI_h u(\mathbf{x}) = \sup_{B \in h\mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}).$$

(i) Let $u : \mathbb{R}^2 \to \mathbb{R}$ be C^2. Compute the expansions of $IS_h u(\mathbf{x})$, $SI_h u(\mathbf{x})$, and $(1/2)(IS_h + SI_h)u(\mathbf{x})$ in terms of small $h > 0$.

(ii) Take $\delta = h$ and compute the same expansions.

(iii) Take $\delta = h^\alpha$ and interpret the expansions for $\alpha > 0$ and for $\alpha < 0$. ■

14.4 Comments and references

Merriman, Bence, and Osher [203] discovered, and gave some heuristic arguments to prove, that a convolution of a shape with a Gaussian followed by a threshold at 1/2 simulated the mean-curvature motion given by $\partial u/\partial t = |Du|\mathrm{curv}(u)$. The consistency of their arguments was checked by Mascarenhas [200]. Barles and Georgelin [36] and Evans [97] also gave consistency proofs; in addition, they showed that iterated weighted Gaussian median filtering converges to the mean curvature motion. An extension of this result to any iterated weighted median filter was given by Ishii in [142]. An interesting attempt to generalize this result to vector median filters was made Caselles, Sapiro, and Chung in [65]. Catté, Dibos, and Koepfler [67] related mean curvature motion to the classic morphological filters whose structuring elements are one-dimensional sets oriented in all directions (see [213] and [261] regarding these filters.)

The importance of the function H in the main expansion theorem raises the following question: Given an increasing continuous function H, are there structuring elements $\mathcal{B}$ such that $H(s) = \inf_{B\in\mathcal{B}} \sup_{(x,y)\in B}(x + sy^2)$? As we have seen is this chapter, the function $H(s) = s$ is attained by a median filter. Pasquignon [227] has studied this question extensively and shown that all of the functions of the form $H(s) = s^\alpha$ are possible using sets of simple structuring elements.

The presentation of the main results of this chapter is mainly original and was announced in the tutorials [133] and [134]. An early version of this work appeared in [130].

Figure 14.4: Fixed point property of the discrete median filter, showing its grid-dependence. Left: original image. 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 this median filter. This is not in agreement with Theorem 14.7, which shows that median filters move images by their curvature : The image on the right clearly has nonzero curvatures! Yet, the discrete median filter that we have applied here operating on a discrete image is grid-dependent and blind to small curvatures.

Figure 14.5: Comparing an iterated median filter and a median filter. Top-left: original image. 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 each image are the level-lines for grey levels equal to multiples of 16. This shows that iterating a small size median filter provides more accuracy and less shape mixing than applying a large size median filter. Compare this with the Koenderink–Van Doorn shape smoothing and the Merriman–Bence–Osher iterated filter in Chapter 4, in particular Figures 4.2, 4.1, and 4.4.

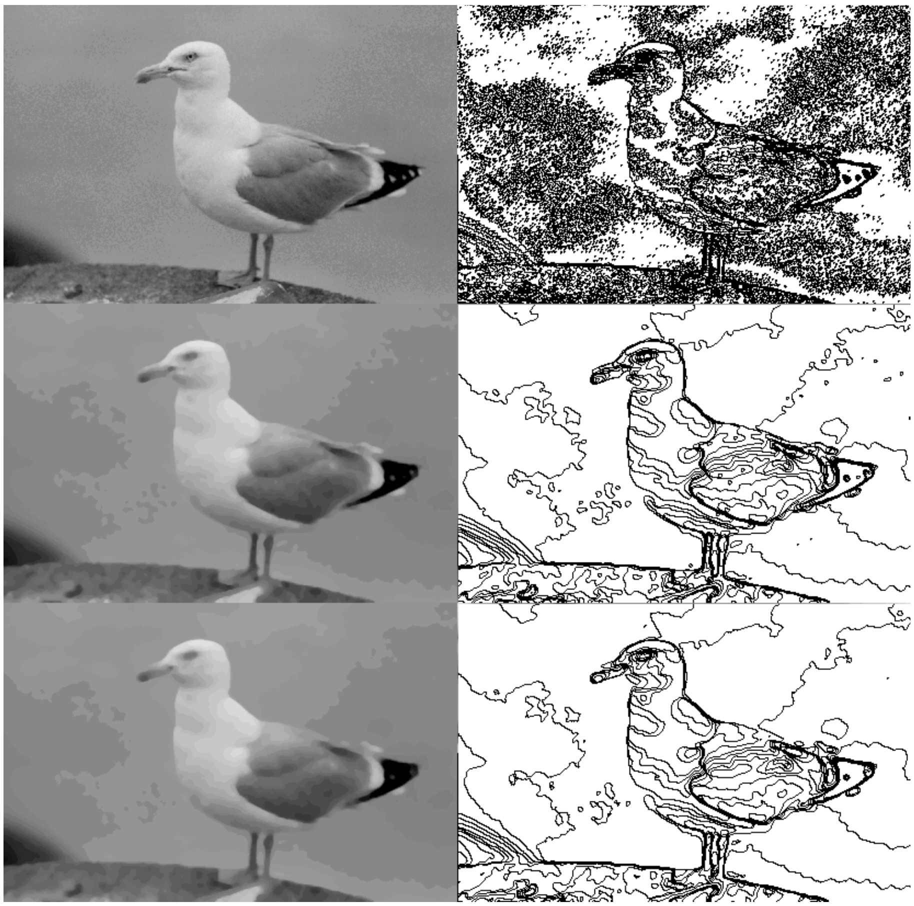

Figure 14.6: Consistency of the median filter and of the Catté–Dibos–Koepfler numerical scheme. Top row: the sea bird image and its level lines for all levels equal to multiples of 12. Second row: a median filter on a disk with radius 2 has been iterated twice. Third row: an inf-sup and then a sup-inf filter based on segments have been applied. On the right: the corresponding level lines of the results, which, according to the theoretical results (Theorems 14.7 and 14.8), must have moved at a speed proportional to their curvature. The results are very close. This yields a cross validation of two very different numerical schemes that implement curvature-motion based smoothing.

Chapter 15

Asymptotic Behavior in Dimension N

We are going to generalize to N dimensions the asymptotic results of Chapter 14. Our aim is to show that the action of any local SMTCII operator, when properly scaled, is a motion of the N-dimensional image that is controlled by its principal curvatures. In particular, we will relate the median filter to the mean curvature of the level surface.

15.1 Asymptotic behavior theorem in $\mathbb{R}^N$

Let $u : \mathbb{R}^N \to \mathbb{R}$ be C^3 and assume that $Du(\mathbf{x}) \neq 0$. Then we denote the vector whose terms are the $N-1$ principal curvatures of the level surface $\{\mathbf{y} \mid u(\mathbf{y}) = u(\mathbf{x})\}$ that passes through $\mathbf{x}$ by $\boldsymbol{\kappa}(u)(\mathbf{x}) = \boldsymbol{\kappa}(u) = (\kappa_2, \dots, \kappa_N)$. The terms $\kappa_i(u)(\mathbf{x})|Du(\mathbf{x})|$ are then the eigenvalues of the restriction of $D^2u(\mathbf{x})$ to $Du(\mathbf{x})^\perp$. (See Definition 11.19.) For $\mathbf{x} \in \mathbb{R}^N$, we write $\mathbf{x} = (x, y_2, \dots, y_N) = (x, \mathbf{y})$, $\mathbf{y} \in \mathbb{R}^{N-1}$ and in the same way $\boldsymbol{s} = (s_2, \dots, s_N)$.

Theorem 15.1. *Let T be a local SMTCII operator. Define*

$$H(\boldsymbol{s}) = T[x + s_2 y_2^2 + \dots + s_N y_N^2](0). \tag{15.1}$$

Then for every C^3 function $u : \mathbb{R}^N \to \mathbb{R}$,

(*i*) $T_h u(\mathbf{x}) = u(\mathbf{x}) + hH(0)|Du(\mathbf{x})| + O(\mathbf{x}, h^2)$;

(*ii*) *If $H(0) = 0$, then on every compact set K contained in $\{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$*

$$T_h u(\mathbf{x}) = u(\mathbf{x}) + hH\left(h\frac{1}{2}\boldsymbol{\kappa}(u)(\mathbf{x})\right)|Du(\mathbf{x})| + O(\mathbf{x}, h^3)$$

where $|O(\mathbf{x}, h^3)| \leq C_K h^3$;

(*iii*) *If $H(0) = 0$, then on every compact set $K \subset \mathbb{R}^N$,*

$$|T_h u(\mathbf{x}) - u(\mathbf{x})| \leq C_K h^2,$$

where C_K denotes some constant that depends only on u and K.

Proof. The proof is the same as the proof of Theorems 14.3 and 14.5. We simply have to relate the notation used for the N-dimensional case to that used in the two-dimensional case. We begin by assuming that $Du(\mathbf{x}) \neq 0$. We then establish the local coordinate system at $\mathbf{x}$ defined by $\boldsymbol{i}_1 = Du(\mathbf{x})/|Du(\mathbf{x})$ and $\boldsymbol{i}_2, \ldots, \boldsymbol{i}_N$, where $\boldsymbol{i}_2, \ldots, \boldsymbol{i}_N$ are the eigenvectors of the restriction of $D^2u(\mathbf{x})$ to the hyperplane $Du(\mathbf{x})^\perp$. Then in a neighborhood of $\mathbf{x}$ we can expand u as follows:

$$u(\mathbf{x}+\mathbf{y}) = u(\mathbf{x}) + px + ax^2 + b_2y_2^2 + \cdots + b_Ny_N^2 + (\mathbf{c}\cdot\mathbf{y})x + R(\mathbf{x},\mathbf{y}), \quad (15.2)$$

where $\mathbf{y} = x\boldsymbol{i}_1 + y_2\boldsymbol{i}_2 + \cdots + y_N\boldsymbol{i}_N$, $p = |Du(\mathbf{x})| > 0$, and for $j = 2, \ldots, N$,

$$\begin{aligned} a &= \frac{1}{2}\frac{\partial^2 u}{\partial x^2}(\mathbf{x}) = \frac{1}{2}D^2u(\mathbf{x})(\boldsymbol{i}_1, \boldsymbol{i}_1), \\ b_j &= \frac{1}{2}\frac{\partial^2 u}{\partial y_j^2}(\mathbf{x}) = \frac{1}{2}D^2u(\mathbf{x})(\boldsymbol{i}_j, \boldsymbol{i}_j), \\ c_j &= \frac{\partial^2 u}{\partial x \partial y_j}(\mathbf{x}) = D^2u(\mathbf{x})(\boldsymbol{i}_1, \boldsymbol{i}_j). \end{aligned} \quad (15.3)$$

We can also write b_j as

$$b_j = \frac{1}{2}|Du(\mathbf{x})|\kappa_j(u)(\mathbf{x}). \quad (15.4)$$

For the proof of (i), we write $u(\mathbf{x}+\mathbf{y}) = u(\mathbf{x}) + px + O(\mathbf{x}, |\mathbf{y}|^2)$ and just follow the steps of the proof of Theorem 15.1. The proof of (ii) and (iii) follows, step by step, the proof of Theorem 14.5. We need only make the following identifications: $cxy \leftrightarrow (\mathbf{c}\cdot\mathbf{y})x$, $by^2 \leftrightarrow b_2y_2^2 + \cdots + b_Ny_N^2$, and $\mathrm{curv}(u) \leftrightarrow \kappa(u)$. □

15.2 Asymptotic behavior of median filters in $\mathbb{R}^N$

The action of median filtering in three dimensions is illustrated in Figures 15.1 and 15.2. The median filters we consider will be defined in terms of a continuous weight function $k : \mathbb{R}^N \to [0, +\infty)$ that is radial, $k(\mathbf{x}) = k(|\mathbf{x}|)$, and that is normalized, $\int_{\mathbb{R}^N} k(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. Recall that, by definition,

$$|B|_k = \int_B k(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

We also assume that k is nonseparable, which is the case if $\{\mathbf{x} \mid k(\mathbf{x}) > 0\}$ is connected. Then by Proposition 10.8, $\mathrm{Med}_k u = \mathrm{Med}_k^- u$ and the median operator can defined by

$$\mathrm{Med}_k u(\mathbf{x}) = \inf_{|B|_k = 1/2} \sup_{\mathbf{y}\in\mathbf{x}+B} u(\mathbf{y}). \quad (15.5)$$

Define the scaled weight function k_h, $0 < h \le 1$, by $k_h(\mathbf{x}) = h^{-N}k(\mathbf{x}/h)$. Then a change of variable shows that $|B|_k = 1/2$ if and only if $|hB|_{k_h} = 1/2$, and this implies that $(\mathrm{Med}_k)_h = \mathrm{Med}_{k_h}$ (see Exercise 14.2). Since we consider

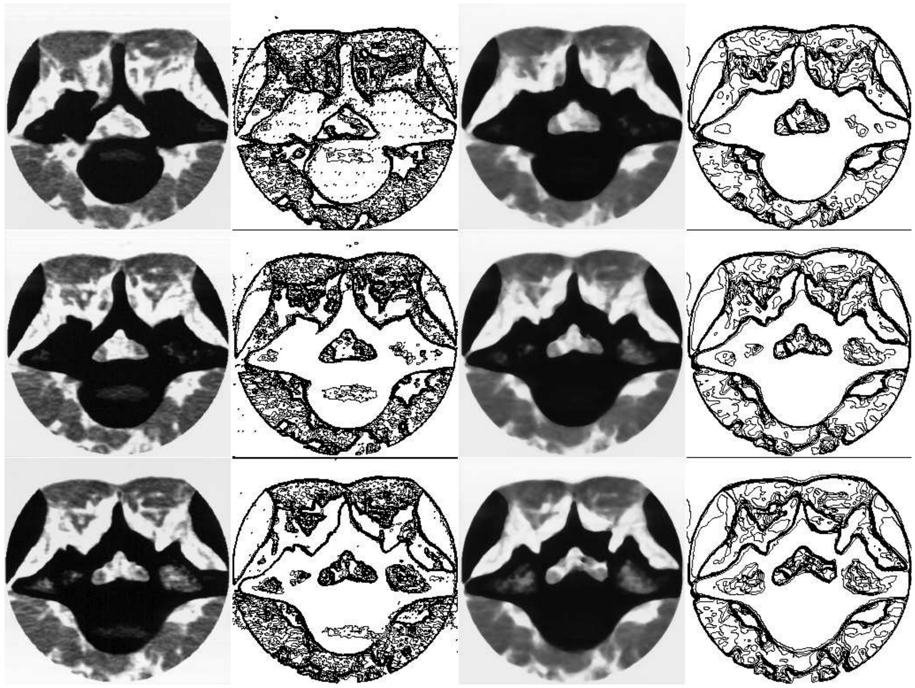

Figure 15.1: Three-dimensional median filter. The original three-dimensional image (not shown) is of 20 slices of a vertebra. Three successive slices are displayed in the left column. The next column shows their level lines (multiples of 20). The third column shows these three slices after one iteration of the median filter based on the three-dimensional ball of radius two. The resulting level lines are shown in the last column.

only one weight function at a time, there should be no confusion if we write Med_h for the scaled operator.

We analyzed the asymptotic behavior of a median filter in $\mathbb{R}^2$ whose weight function was the characteristic function of the unit disk in Chapter 14. This proof can be generalized to $\mathbb{R}^N$ by taking k to be the normalized characteristic function of the unit ball. We will go in a different direction by taking smooth weight functions. Our analysis will not be as general as possible because this would be needlessly complicated. The k we consider will be smooth (C^∞) and have compact support. This means that the considered median filters are local. Thus, the results of Theorem 15.1 apply, provided we get an estimate near 0 of the structure function H of the median filter.

Lemma 15.2. *Let k be a nonnegative radial function belonging to the Schwartz class $\mathcal{S}$. Assume that $\int_{\mathbb{R}^N} k(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$ and that the support of k is connected in $\mathbb{R}^N$. Then the structure function of* Med_k $H(h\mathbf{b}) = \mathrm{Med}_k[x + h(b_2y_2^2 + \cdots + b_Ny_N^2)](0)$ *can be expressed as*

$$H(h\mathbf{b}) = hc_k\Big(\sum_{j=2}^{N} b_j\Big) + O(h^2),$$

where

$$c_k = \frac{\int_{\mathbb{R}^{N-1}} y_2^2 k(\mathbf{y})\,\mathrm{d}\mathbf{y}}{\int_{\mathbb{R}^{N-1}} k(\mathbf{y})\,\mathrm{d}\mathbf{y}},$$

$\mathbf{y} = (y_2, \dots, y_N)$, *and* $\mathbf{b} = (b_2, \dots, b_N)$.

Proof. Before beginning the proof, note that we have not assumed that k has compact support, so the result applies to the Gaussian, for example.

We will use the abbreviation $\mathbf{b}(\mathbf{y},\mathbf{y}) = b_2y_2^2 + \cdots + b_Ny_N^2$, since $\mathbf{b}$ is, in fact, a diagonal matrix. Our proof is based on an analysis of the function $f(\lambda, h) = |\mathcal{X}_\lambda(x + h\mathbf{b}(\mathbf{y},\mathbf{y}))|_k$. Since $\mathcal{X}_\lambda(x + h\mathbf{b}(\mathbf{y},\mathbf{y})) = \{(x,\mathbf{y}) \mid x + h\mathbf{b}(\mathbf{y},\mathbf{y}) \geq \lambda\}$, we can express f as an integral,

$$f(\lambda, h) = \int_{\mathbb{R}^{N-1}} \int_{\lambda - h\mathbf{b}(\mathbf{y},\mathbf{y})}^{\infty} k(x,\mathbf{y})\,\mathrm{d}x\,\mathrm{d}\mathbf{y}.$$

It follows from the assumption that k is in the Schwartz class that $f : \mathbb{R}^2 \to \mathbb{R}$ is bounded and C^∞. Also, for every $h \in \mathbb{R}$, $\lim_{\lambda\to-\infty} f(\lambda, h) = 1$ and $\lim_{\lambda\to+\infty} f(\lambda, h) = 0$. Thus, for every $h \in \mathbb{R}$, there is at least one λ such that $f(\lambda, h) = 1/2$. In fact, there is only one such λ; this is a consequence of the assumption that the k is continuous and that its support is connected, which implies that it is nonseparable (see Exercise 10.5). To see that λ is unique, assume that there are $\lambda < \lambda'$ such that $f(\lambda, h) = 1/2$ and $f(\lambda', h) = 1/2$. Then the two sets $\{(x,\mathbf{y}) \mid x + h\mathbf{b}(\mathbf{y},\mathbf{y}) \geq \lambda'\}$ and $\{(x,\mathbf{y}) \mid x + h\mathbf{b}(\mathbf{y},\mathbf{y}) \leq \lambda\}$ both have k-measure $1/2$, but their intersection is empty. This contradicts the fact that k is nonseparable. This means that the relation $f(\lambda, h) = 1/2$ defines implicitly a well-defined function $h \mapsto \lambda(h)$.

Recall that Med_k was originally defined in terms of the superposition formula

$$\mathrm{Med}_k u(\mathbf{x}) = \sup\{\lambda \mid \mathbf{x} \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda u\}.$$

This translates for our case into the relation

$$\mathrm{Med}_k[x + h\mathbf{b}(\mathbf{y},\mathbf{y})](0) = \sup\{\lambda \mid 0 \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda[x + h\mathbf{b}(\mathbf{y},\mathbf{y})]\} = \lambda(h)$$

because $0 \in \mathcal{M}\mathrm{ed}_k \mathcal{X}_\lambda[x + h\mathbf{b}(\mathbf{y},\mathbf{y})]$ if and only if $|\mathcal{X}_\lambda[x + h\mathbf{b}(\mathbf{y},\mathbf{y})]|_k \geq 1/2$.

We are interested in the behavior of $h \mapsto \lambda(h)$ near the origin. The first thing to note is that $\lambda(0) = 0$. To see this, write

$$f(\lambda(0), 0) = \int_{\mathbb{R}^{N-1}} \int_{\lambda(0)}^{\infty} k(x,\mathbf{y})\,\mathrm{d}x\,\mathrm{d}\mathbf{y} = \frac{1}{2}.$$

Since k is radial, the value $\lambda = 0$ solves the equation $\int_{\mathbb{R}^{N-1}} \int_{\lambda}^{\infty} k(x,\mathbf{y})\,\mathrm{d}x\,\mathrm{d}\mathbf{y} = 1/2$. We have just shown that this equation has a unique solution, so $\lambda(0) = 0$.

Now consider the first partial derivatives of f:

$$\frac{\partial f}{\partial \lambda}(\lambda, h) = -\int_{\mathbb{R}^{N-1}} k\Big(\big((\lambda - h\mathbf{b}(\mathbf{y},\mathbf{y}))^2 + \mathbf{y}\cdot\mathbf{y}\big)^{1/2}\Big)\,\mathrm{d}\mathbf{y}. \tag{15.6}$$

$$\frac{\partial f}{\partial h}(\lambda, h) = \int_{\mathbb{R}^{N-1}} \mathbf{b}(\mathbf{y},\mathbf{y})k\Big(\big((\lambda - h\mathbf{b}(\mathbf{y},\mathbf{y}))^2 + \mathbf{y}\cdot\mathbf{y}\big)^{1/2}\Big)\,\mathrm{d}\mathbf{y}. \tag{15.7}$$

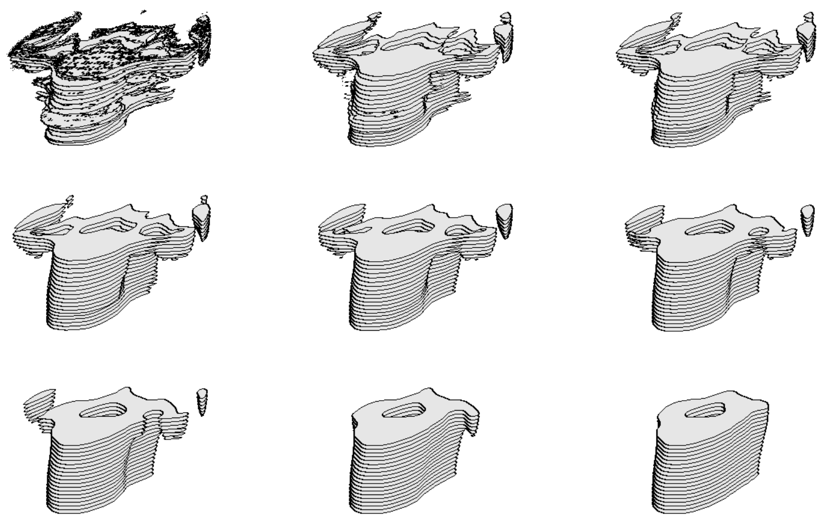

Figure 15.2: Median filtering of a three-dimensional image. The first image is a representation of the horizontal slices of a three-dimensional level surface of the three-dimensional image of a vertebra. Right to left, top to bottom: 1, 2, 5, 10, 20, 30, 60, 100 iterations of a three-dimensional median filter based on a ball with radius three. This scheme is a possible implementation of the mean curvature motion, originally proposed as such by Merriman, Bence and Osher.

These functions are C^∞ because k is in the Schwartz class; also, $(\partial f/\partial\lambda)(0,0) \neq 0$. Then by the implicit function theorem, we know that the function $h \mapsto \lambda(h)$ that satisfies $f(\lambda(h), h) = 1/2$ is also C^∞ and that

$$\lambda'(h)\frac{\partial f}{\partial \lambda}(\lambda(h), h) + \frac{\partial f}{\partial h}(\lambda(h), h) = 0.$$

Thus, for small h,

$$\lambda'(h) = -\frac{\frac{\partial f}{\partial h}(\lambda(h), h)}{\frac{\partial f}{\partial \lambda}(\lambda(h), h)},$$

and, using equations (15.6) and(15.7), we see that

$$\lambda'(0) = \frac{\int_{\mathbb{R}^{N-1}} \mathbf{b}(\mathbf{y},\mathbf{y}) k\big((\mathbf{y}\cdot\mathbf{y})^{1/2}\big)\,\mathrm{d}\mathbf{y}}{\int_{\mathbb{R}^{N-1}} k\big((\mathbf{y}\cdot\mathbf{y})^{1/2}\big)\,\mathrm{d}\mathbf{y}}.$$

Now expand λ for small h:

$$\lambda(h) = \lambda(0) + \lambda'(0)h + O(h^2).$$

Since $\int_{\mathbb{R}^{N-1}} \mathbf{b}(\mathbf{y},\mathbf{y}) k\big((\mathbf{y}\cdot\mathbf{y})^{1/2}\big)\,\mathrm{d}\mathbf{y} = \big(\sum_{j=2}^{N-1} b_j\big) \int_{\mathbb{R}^{N-1}} y_2^2 k\big((\mathbf{y}\cdot\mathbf{y})^{1/2}\big)\,\mathrm{d}\mathbf{y}$, $H(h\mathbf{b}) = \lambda(h)$, and $\lambda(0) = 0$, this proves the lemma. □

Theorem 15.3. *Let k be a nonnegative radial function belonging to the Schwartz class $\mathcal{S}$. Assume that $\int_{\mathbb{R}^N} k(\mathbf{y})\,\mathrm{d}\mathbf{y} = 1$ and that the support of k is compact and connected. Then for every C^3 function $u : \mathbb{R}^N \to \mathbb{R}$:*

(*i*) *On every compact set* $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$,

$$\mathrm{Med}_h u(\mathbf{x}) = u(\mathbf{x}) + h^2 \frac{1}{2} c_k \Big(\sum_{i=2}^{N} \kappa_i(u)(\mathbf{x}) \Big) |Du(\mathbf{x})| + O(\mathbf{x}, h^3),$$

where $|O(\mathbf{x}, h^3)| \leq C_K h^3$ *for some constant that depends only on* u *and* K.

(*ii*) *On every compact set* $K \subset \mathbb{R}^N$, $|\mathrm{Med}_h u(\mathbf{x}) - u(\mathbf{x})| \leq C_K h^2$ *for some constant* C_K *that depends only on* u *and* K.

Proof. Theorem 15.1 is directly applicable. We know from Lemma 15.2 that $H(0) = 0$, so we can read (*ii*) directly from Theorem 15.1(*iii*). By Lemma 15.2,

$$H(h\boldsymbol{\kappa}(u)) = hc_k \Big(\sum_{i=2}^{N} \kappa_i(u) |Du| \Big) + O(h^2).$$

From this and Theorem 15.1(*ii*), we get

$$\mathrm{Med}_h u(\mathbf{x}) = u(\mathbf{x}) + h^2 \frac{1}{2} c_k \Big(\sum_{i=2}^{N} \kappa_i(u)(\mathbf{x}) \Big) |Du(\mathbf{x})| + O(\mathbf{x}, h^3),$$

and we know that the estimate is uniform on any compact set $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$. □

15.3 Exercises : other motions by the principal curvatures

This section contains several applications of Theorem 15.1 in three dimensions. A level surface of a C^3 function in three dimensions has two principal curvatures, and this provides an extra degree of freedom for constructing contrast-invariant operators based on curvature motion. We develop the applications in three exercises. For each case, we will assume that the principal curvatures κ_1 and κ_2 are ordered so that $\kappa_1 \leq \kappa_2$. In each example, the set of structuring elements $\mathcal{B}$ is constructed from a single set B in $\mathbb{R}^2$ by rotating B in all possible ways, that is, $\mathcal{B} = \{RB \mid B \in \mathbb{R}^2, R \in SO(3)\}$. For each example we write

$$SI_h u(\mathbf{x}) = \sup_{B \in \mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x} + hB} u(\mathbf{y})$$

and

$$IS_h u(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x} + hB} u(\mathbf{y}),$$

where $0 < h \leq 1$.

Exercise 15.1. Let B be a segment of length 2 centered at the origin. Our aim is to show that

$$IS_h u = u + h^2 \frac{1}{2} \kappa_1^+(u)|Du| + O(h^3),$$
$$SI_h u = u + h^2 \frac{1}{2} \kappa_2^-(u)|Du| + O(h^3).$$

This implies

$$IS_h u + SI_h u = u + h^2\frac{1}{2}(\operatorname{sign}(\kappa_1(u)) + \operatorname{sign}(\kappa_2(u)))\min(|\kappa_1(u)|, |\kappa_2(u)|) + O(h^3).$$

(i) The first step is to compute $H(h\mathbf{b})$. One way to do this is to write $x = r\sin\phi$, $y_2 = r\cos\phi\cos\theta$, $y_3 = r\cos\phi\cos\theta$, and use an argument similar to that given in the proof of Theorem 14.8 to show that, for a fixed θ and small h, the "inf-sup" of

$$r\sin\phi + hb_2 r^2\cos^2\phi\cos^2\theta + hb_3 r^2\cos^2\phi\cos^2\theta$$

always occurs at $\phi = 0$. Then $H(h\mathbf{b}) = hH(\mathbf{b})$ and

$$H(\mathbf{b}) = \inf_{B\in\mathcal{B}}\sup_{\mathbf{y}\in B}(b_2 y_2^2 + b_3 y_3^2) = \inf_{\theta}\sup_{0\le r\le 1} r^2(b_2\cos^2\theta + b_3\sin^2\theta).$$

Deduce that $b_2 < 0$ or $b_3 < 0$ implies $H(\mathbf{b}) = 0$ and that $0 \le b_2 \le b_3$ implies $H(\mathbf{b}) = b_2$.

(ii) Since $H(0) = 0$, deduce from Theorem 15.1 that

$$IS_h u(\mathbf{x}) = u(\mathbf{x}) + h^2\frac{1}{2}\kappa_1^+(u)(\mathbf{x})|Du(\mathbf{x})| + O(h^3). \tag{15.8}$$

Exercise 15.2. Let B be the union of two symmetric points $(1, 0, 0)$ and $(-1, 0, 0)$. Use the techniques of Exercise 11.2 to show that

$$IS_h u = u + h^2\frac{1}{2}\min\{\kappa_1(u), \kappa_2(u)\}|Du| + O(h^3);$$
$$SI_h u = u + h^2\frac{1}{2}\max\{\kappa_1(u), \kappa_2(u)\}|Du| + O(h^3);$$
$$IS_h + SI_h u = u + h^2\frac{1}{2}(\kappa_1(u) + \kappa_2(u))|Du| + O(h^3).$$

The last formula shows that the operator $IS_h + SI_h$ involves the mean curvature of u at $\mathbf{x}$. ■

Exercise 15.3. Let B consist of two orthogonal segments of length two centered at the origin.

(i) Show that

$$IS_h u = u + h^2\frac{1}{2}\Big(\frac{\kappa_1(u) + \kappa_2(u)}{2}\Big)^+|Du| + O(h^3);$$
$$SI_h u = u + h^2\frac{1}{2}\Big(\frac{\kappa_1(u) + \kappa_2(u)}{2}\Big)^-|Du| + O(h^3).$$

(ii) Show that you can get the mean curvature by simply taking B to be the four end-points of the orthogonal segments. Check that another possibility for obtaining the mean curvature is to alternate these operators or to add them. ■

15.4 Comments and references

The references for this chapter are essentially the same as those for Chapter 14. The main theorem on the asymptotic behavior of morphological filters was first stated and proved in [133] and [134]. The examples developed in Exercises 15.1, 15.2, and 15.3 have not been published elsewhere. The consistency of Gaussian smoothing followed by thresholding and mean-curvature motion was proved in increasing mathematical sophistication and generality by Merriman, Bence, and Osher [203], Mascarenhas [200], Barles and Georgelin [36], and Evans [97]. Our presentation is slightly more general than the ones cited because we allow any nonnegative weight function in the Schwartz class. The most general result was given by Ishii in [142].

Chapter 16

Affine-Invariant Mathematical Morphology

In Chapter 9, we introduced a class of simple set and function operators called erosions and dilations. These operators were defined by a single structuring set. They are contrast invariant and translation invariant, but they are not affine invariant. In this chapter, we introduce set operators, also called erosions and dilations, that are affine invariant.

Our interest in affine-invariant smoothing, like our interest in contrast- and translation-invariant smoothing, is based on practical considerations. When we take a photograph of a plane image, say, a painting, the image is projected onto the focal plane of the camera. If the camera is an ideal pin-hole device, then this is a projective transformation where the center of projection is the pin hole. In any case, it approximates a projective transformation. If we are far removed from the plane of the painting, then the focal plane of the camera approximates the plane at infinity, and the transformation looks like an affine transformation. For a more common example, we note that most digital cameras, copy machines, fax machines, and scanners introduce a slight affine distortion. Thus, we would like the smoothing to be affine invariant so it is "blind" to any deformations introduced by these processes. It would be nice to have a smoothing that is invariant under the full projective group, but we will see later (Chapter 22) that this is not possible.

16.1 Affine invariance

Isometries, by definition, preserve the distance between points, and hence, preserve the angle between vectors. In a finite dimensional space $\mathbb{R}^N$, any isometry can be represented by $\mathbf{x} \mapsto A\mathbf{x} + \boldsymbol{a}$, where A is an orthogonal matrix and $\boldsymbol{a}$ is a fixed vector. These transformations include all of the rigid motions of $\mathbb{R}^N$ plus reflections. Classical Euclidean geometry in $\mathbb{R}^2$ is concerned with the objects that are invariant under these transformations. If we loosen the requirement that the matrix A be orthogonal and assume only that it is nonsingular, then we have generalized Euclidean motions to affine motions, and the distance between points is no longer an invariant. However, there are affine invariants, and the most important from our point of view is that parallel lines are mapped

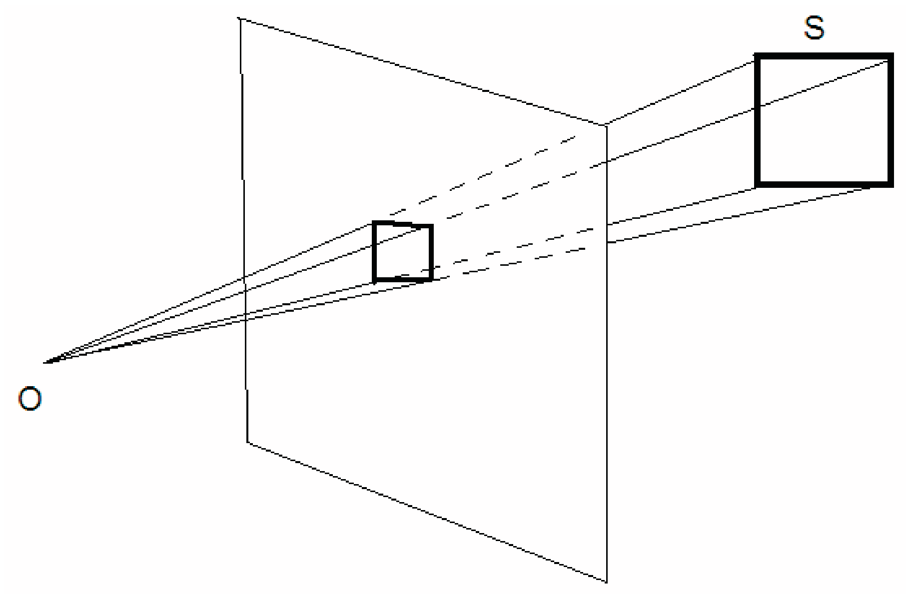

Figure 16.1: A rectangle seen from far enough has its sides roughly parallel and looks like a parallelogram. Thus affine invariance is a particular instance of projective invariance.

into parallel lines and finite points are mapped into finite points. Furthermore, if the determinant of A is one, $|A| = 1$, then the transformation preserves area: Unit squares are mapped into parallelograms whose area is one. Note, however, that the parallelograms can be arbitrarily long.

If we jump to projective geometry and the projective plane, then parallel lines are not necessarily preserved. Thus, affine transformations are a special class of projective transformations. This means that affine geometry can be considered a generalization of Euclidian geometry or a specialization of projective geometry. Incidentally, this is the view taken in classical Chinese drawing, which tends always to display scenes as seen from a distance and to maintain parallelism.

16.2 Affine-invariant erosions and dilations

Everything in this chapter will take place in $\mathbb{R}^2$ so the following definition is given for $\mathbb{R}^2$. The group of all linear affine transformations $A : \mathbb{R}^2 \to \mathbb{R}^2$ with determinant one, $|A| = 1$, is called the special linear group; it is often denoted by $SL(\mathbb{R}^2)$.

Our goal is to define erosions and dilations that are invariant under $SL(\mathbb{R}^2)$. Any attempt to do so using Euclidean distance is doomed to failure, since the distances between points are not affine invariant. It is thus necessary to base the definition of affine-invariant erosions and dilations on some affine invariant, and the most obvious one to use is area. This leads to the notion of the *affine-invariant distance* between a point and a set. We begin with the definition of a *chord-arc set*.

Definition 16.1. *Let X be a subset of $\mathbb{R}^2$ and let Δ be a straight line in $\mathbb{R}^2$. Any connected component of $X \setminus \Delta = X \cap \Delta^c$ is called a chord-arc set defined by X and Δ.*

Exercise 16.1. Chord-arc sets take their name from the case where X is a disk, in which case a chord-arc set is called a segment. What are the chord-arc sets if (i) X is an open disk and (ii) X is a closed disk? What is the situation when X is a closed arc of a circle or a closed segment of a disk? ■

Exercise 16.2. Suppose that $\mathbf{x} \in (\overline{X})^c \cap \Delta$. Show that there are two, and only two, chord-arc sets defined by $(\overline{X})^c = (X^c)^\circ$ and Δ that contain $\mathbf{x}$ in their boundary. (Note that we are assuming that $(\overline{X})^c \neq \emptyset$.) ■

We are interested first in some special chord-arc sets that will be used to define the affine distance from a point $\mathbf{x}$ to a set X. From Exercise 16.2, we know that there are only two chord-arc sets defined by $(\overline{X})^c$ and Δ that contain $\mathbf{x}$ in their boundary, if $\overline{X} \neq \mathbb{R}^2$. In this case, we call these two sets $CA_1(\mathbf{x}, \Delta, X)$ and $CA_2(\mathbf{x}, \Delta, X)$, and we order them so that

$$\text{area}(CA_1(\mathbf{x}, \Delta, X)) \leq \text{area}(CA_2(\mathbf{x}, \Delta, X)).$$

Definition 16.2. *Let X be a subset of $\mathbb{R}^2$ and let $\mathbf{x}$ be an arbitrary point in $\mathbb{R}^2$. We define the affine distance from $\mathbf{x}$ to X to be*

$$\delta(\mathbf{x}, X) = \inf_{\Delta} [\text{area}(CA_1(\mathbf{x}, \Delta, X))]^{1/2} \textit{ if } \mathbf{x} \in (\overline{X})^c \textit{ and } \delta(\mathbf{x}, X) = 0 \textit{ otherwise.}$$

(See Figure 16.2.)

The power $1/2$ is taken so that δ has the "dimension" of a distance. Notice that $\delta(\mathbf{x}, X)$ can be infinite: Take X convex and compact and $\mathbf{x} \notin X$. Then all chord-arc sets defined by a straight line Δ through $\mathbf{x}$ have infinite area. Notice also that $\delta(\mathbf{x}, \emptyset) = +\infty$ and $\delta(\mathbf{x}, \mathbb{R}^2) = 0$.

Definition 16.3. *The affine a-dilation $\tilde{\mathcal{D}}_a$ and the affine a-erosion $\tilde{\mathcal{E}}_a$ are set operators defined for $X \subset \mathbb{R}^2$ by*

$$\tilde{\mathcal{D}}_a X = \{\mathbf{x} \mid \delta(\mathbf{x}, X) \leq a^{1/2}\} \textit{ and } \tilde{\mathcal{E}}_a X = \{\mathbf{x} \mid \delta(\mathbf{x}, X^c) > a^{1/2}\}.$$

They are extended to $\mathcal{M}(S_2)$ by the standard extension (Definition 7.1.)

Exercise 16.3. Check that $\tilde{\mathcal{D}}_a \mathbb{R}^2 = \tilde{\mathcal{E}}_a \mathbb{R}^2 = \mathbb{R}^2$. Show that $\tilde{\mathcal{E}}_a X = (\tilde{\mathcal{D}}_a X^c)^c$. (Recall Exercise 8.1(ii).) This relation shows that eroding a set and dilating its complement yield complementary sets. This is a useful symmetry, since the same shape can appear as an upper level set or as the complement of an upper level set, depending on whether it is darker or lighter than the background. ■

The names we have used for the operators $\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ are not standard nomenclature in mathematical morphology. Indeed, in mathematical morphology, a dilation must commute with set union and an erosion is expected to commute with set intersection. It is easy to check that $\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ do not satisfy these properties.[1] Nevertheless, we will use these names. There should be no confusion, and these operators are natural generalizations of the corresponding Euclidean erosions and dilations discussed in Chapter 9.

[1]We thank Michel Schmitt for pointing this out to us.

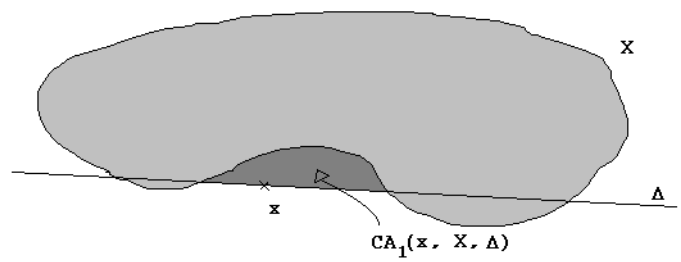

Figure 16.2: Affine distance to a set.

Exercise 16.4. Prove the above statements, that a standard erosion commutes with set intersection, a standard dilation commutes with set union. Give examples showing that this commutation is no more true with affine dilation or erosion. ■

Proposition 16.4. *$\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ are monotone and affine-invariant. More precisely, for every linear map A such that $|A| > 0$, $\tilde{\mathcal{E}}_a A = A\tilde{\mathcal{E}}_{|A|^{-1}a}$ and $\tilde{\mathcal{D}}_a A = A\tilde{\mathcal{D}}_{|A|^{-1}a}$. In particular, $\tilde{\mathcal{E}}_a A = A\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a A = A\tilde{\mathcal{D}}_a$ if $|A| = 1$.*

Proof. If $X \subset Y \subset \mathbb{R}^2$, then it is easily seen from the definitions that $\delta(\mathbf{x}, Y) \leq \delta(\mathbf{x}, X)$ for every $\mathbf{x} \in \mathbb{R}^2$. It then follows from the definition of $\tilde{\mathcal{D}}_a$ that $\tilde{\mathcal{D}}_a X \subset \tilde{\mathcal{D}}_a Y$. Hence, $\tilde{\mathcal{D}}_a$ is monotone. The monotonicity of $\tilde{\mathcal{E}}_a$ follows directly from the relation $\tilde{\mathcal{E}}_a X = (\tilde{\mathcal{D}}_a X^c)^c$.

The transformation A preserves the topological properties of the configuration determined by $\mathbf{x}$, X, and Δ and multiplies all areas by $|A|$. This implies that $\delta(A\mathbf{x}, AX) = |A|^{1/2}\delta(\mathbf{x}, X)$. It follows from the definition of $\tilde{\mathcal{D}}_a$ that $\tilde{\mathcal{D}}_a AX = A\tilde{\mathcal{D}}_{|A|^{-1}a} X$. Thus $\tilde{\mathcal{D}}_a A = A\tilde{\mathcal{D}}_{|A|^{-1}a}$. The result for $\tilde{\mathcal{E}}_a$ follows from the result for $\tilde{\mathcal{D}}_a$ and the relations $AX^c = (AX)^c$ and $\tilde{\mathcal{E}}_a X = (\tilde{\mathcal{D}}_a X^c)^c$. The extension of this relation to subsets of $\mathbb{R}^2 \cup \{\infty\}$ is straightforward, since we have set $A\infty = \infty$. □

The next proposition shows that it is equivalent to erode a set or its interior, and to dilate a set or its closure.

Proposition 16.5. *For any $X \subset \mathbb{R}^2$, $\tilde{\mathcal{E}}_a X = \tilde{\mathcal{E}}_a X^\circ$ and $\tilde{\mathcal{D}}_a X = \tilde{\mathcal{D}}_a \overline{X}$, where X° denotes the interior of X and $\overline{X}$ denotes the closure of X.*

Proof. We will first prove the result about $\tilde{\mathcal{E}}_a$. Since $\tilde{\mathcal{E}}_a$ is monotone, $\tilde{\mathcal{E}}_a X^\circ \subset \tilde{\mathcal{E}}_a X$. Now, $\mathbf{x} \in \tilde{\mathcal{E}}_a X$ if and only if $\delta(\mathbf{x}, X^c) > a^{1/2}$. By the definition, $\delta(\mathbf{x}, Y) = \delta(\mathbf{x}, \overline{Y})$, so $\mathbf{x} \in \tilde{\mathcal{E}}_a X$ if and only if $\delta(\mathbf{x}, \overline{X^c}) > a^{1/2}$. Since $\overline{X^c} = (X^\circ)^c$, this means that $\delta(\mathbf{x}, (X^\circ)^c) > a^{1/2}$, which proves that $\mathbf{x} \in \tilde{\mathcal{E}}_a X^\circ$. That $\tilde{\mathcal{D}}_a X = \tilde{\mathcal{D}}_a \overline{X}$ follows from the two identities $\tilde{\mathcal{E}}_a X = (\tilde{\mathcal{D}}_a X^c)^c$ and $((X^c)^\circ)^c = \overline{X}$.

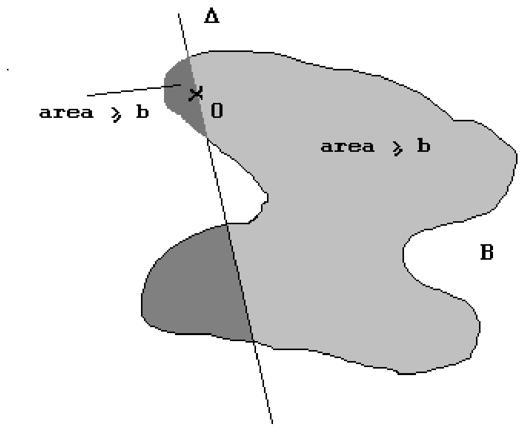

Figure 16.3: An affine structuring element: All lines through 0 divide B into several connected components. The two of them which contain 0 in their boundary have area larger or equal to b.

□

This result means in particular that affine erosion erases all boundary points. It is easy to check that the affine erosion processes independently connected components of a set. This is the object of the next exercise.

Exercise 16.5. Let X_i°, $i \in I$, be the connected components of X°. Then the $\tilde{\mathcal{E}}_a X_i^\circ$ are disjoint and $\tilde{\mathcal{E}}_a X^\circ = \bigcup_{i \in I} \tilde{\mathcal{E}}_a X_i^\circ$.

Lemma 16.6. *$\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ are standard monotone.*

Proof. We have to check five properties, but the three first items of Definition 7.3 are obvious. Let $X \subset \mathbb{R}^2$ be bounded. Then $\tilde{\mathcal{E}}_a X \subset X$ also is bounded. Assume now that X^c is bounded. Then $X \supset \overline{B(0,R)^c}$ for some $R \geq 0$. It is easily checked that $\tilde{\mathcal{E}}_a(\overline{B(0,R)^c}) = \overline{B(0,R)^c}$. Thus $\tilde{\mathcal{E}}_a(X) \supset \overline{B(0,R)^c}$. We conclude that $\tilde{\mathcal{E}}_a$ is standard monotone. By Proposition 7.5 its dual operator $\tilde{\mathcal{D}}_a$ also is standard monotone. □

Definition 16.7. *A set B is called an affine structuring element if B is open and connected and if $\delta(0, B^c) > 1$. We denote the set of all affine structuring elements by $\mathcal{B}_{\text{aff}}$.*

Exercise 16.6. If A is a linear transformation and B an affine structuring element, check that $\delta(0,(AB)^c) = \delta(0, AB^c) = |A|^{1/2}\delta(0,B^c)$. Deduce that $A\mathcal{B}_{\text{aff}} = \mathcal{B}_{\text{aff}}$ for all $A \in SL(\mathbb{R}^2)$. ∎

$$\mathcal{T}X = \bigcup_{B \in \mathcal{B}} \bigcap_{\mathbf{y} \in B} (X - \mathbf{y}) = \{\mathbf{x} \mid \mathbf{x} + B \subset X \text{ for some } B \in \mathcal{B}\},$$

where $\mathcal{B} = \{X \mid 0 \in \mathcal{T}X\}$. The task reduces to characterizing the set of structuring elements $\mathcal{B}$. We know that $0 \in \tilde{\mathcal{E}}_a X$ if and only if $\tilde{\mathcal{E}}_a X \neq \emptyset$ and $\delta(0, X^c) > a^{1/2}$, and this, and the results stated above, lead to the following definition.

Proposition 16.8. *$\mathcal{B}_1 = a^{\frac{1}{2}}\mathcal{B}_{aff}$ is a standard set of structuring elements for $\tilde{\mathcal{E}}_a$. Thus for every set $X \subset \mathbb{R}^2$,*

$$\tilde{\mathcal{E}}_a X = \bigcup_{B \in \mathcal{B}_{\text{aff}}} \bigcap_{\mathbf{y} \in a^{1/2}B} (X - \mathbf{y}) = \{\mathbf{x} \mid \mathbf{x} + a^{1/2}B \subset X \textit{ for some } B \in \mathcal{B}_{\text{aff}}\}.$$

Proof. $\tilde{\mathcal{E}}_a$ is translation invariant and standard monotone. Thus we can apply Matheron Theorem 8.2. The canonical set of structuring elements of $\tilde{\mathcal{E}}_a$ is

$$\mathcal{B}_0 = \{B \mid \tilde{\mathcal{E}}_a \ni 0\} = \{B \mid \delta(0, B^c) > a^{\frac{1}{2}}.\}.$$

Then $\mathcal{B}_1 = a^{\frac{1}{2}}\mathcal{B}_{aff}$ is the subset of elements in $\mathcal{B}_0$ which are open and connected. By Proposition 8.4 we only need to show that every element B_0 in $\mathcal{B}_0$ contains some element B_1 of $\mathcal{B}_1$. Let us choose for B_1 the open connected component of 0 in B_0. For every line Δ passing by 0, $CA_1(0, \Delta, B_0^c)$ is a connected open set contained in $B_0 \setminus \Delta$. Thus it is also contained in $B_1 \setminus \Delta$. This implies that $CA_1(0, \Delta, B_1^c) \supseteq CA_1(0, \Delta, B_0^c)$ and, by taking the infimum of the areas of these sets, that $\delta(0, B_1^c) \geq \delta(0, B_0^c) > a^{\frac{1}{2}}$.

□

Remark 16.9. *An alternative way to prove the above proposition is the following. By Proposition 16.5 and the result of Exercise 16.5 we know that*

$$\tilde{\mathcal{E}}_a B_0 = \tilde{\mathcal{E}}_a B_0^\circ = \cup_{i \in I} \tilde{\mathcal{E}}_a \left[(B_0^\circ)_i\right],$$

where $(B_0^\circ)_i$ are the open connected components of B_0°. Now one of them $(B_0)_i = B_1$ is the open connected component containing 0. Thus $\tilde{\mathcal{E}}_a B_0 \ni 0 \Leftrightarrow \tilde{\mathcal{E}}_a B_1 \ni 0$.

Let us give a more practical characterization for the affine structuring elements, which follows immediately from Definitions 16.2 and 16.7.

Proposition 16.10. *A set B is an affine structuring element if it is open, connected, and contains the origin, and if for some $b > 1$ and for every straight line Δ through the origin, the two connected components of $B \setminus \Delta$ that contain the origin in their boundary each have area greater than some number $b > 1$. (See Figure 16.3.)*

It is easy to see that $\tilde{\mathcal{E}}_a$ is not upper semicontinuous on $\mathcal{L}$. $\tilde{\mathcal{E}}_a$ sends a closed disk on an open disk and therefore doesn't map $\mathcal{L}$ into itself. As for $\tilde{\mathcal{D}}_a$, it is not known whether it is upper semicontinuous or not and this is a good question! All the same we can define a stack filter for $\tilde{\mathcal{D}}_a$ or $\tilde{\mathcal{E}}_a$ by the superposition principle. Thus we set for $u \in \mathcal{F}$,

$$\tilde{D}_a u(\mathbf{x}) = \sup\{\lambda \in \mathbb{R} \mid \mathbf{x} \in \tilde{\mathcal{D}}_a(\mathcal{X}_\lambda u)\};$$

$$\tilde{E}_a u(\mathbf{x}) = \sup\{\lambda \in \mathbb{R} \mid \mathbf{x} \in \tilde{\mathcal{E}}_a(\mathcal{X}_\lambda u)\}$$

and we call them respectively *affine function dilation* and *affine function erosion.*

Proposition 16.11. *The affine function dilation and erosion are translation invariant, contrast invariant and standard monotone from $\mathcal{F}$ to $\mathcal{F}$. In addition they are affine invariant. Finally the commutation almost everywhere holds:*

$$\mathcal{X}_\lambda(\tilde{D}_a u) = \tilde{\mathcal{D}}_a(\mathcal{X}_\lambda u) \text{ and } \mathcal{X}_\lambda(\tilde{E}_a u) = \tilde{\mathcal{E}}_a(\mathcal{X}_\lambda u), \tag{16.1}$$

almost everywhere for almost every $\lambda \in \mathbb{R}$.

Proof. Since $\tilde{\mathcal{D}}_a$ and $\tilde{\mathcal{E}}_a$ are standard monotone and translation invariant, this is a direct application of Theorem 7.16. Since the set operators are not upper semicontinuous, there is no chance to get a full commutation with thresholds. However, the commutation with thresholds almost everywhere holds by Proposition 8.18. □

Let us finally point out that the affine function erosion and dilation are dual of each other.

Proposition 16.12. *For $u \in \mathcal{F}$, $\tilde{E}_a u = -\tilde{D}_a(-u)$.*

Proof. We wish to use the duality relations between $\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ and the superposition principle. This leads us to deal with upper level sets of $-u$ which are lower level sets of u. Thus we will need the following relations,

$$\mathcal{X}_\lambda(-u) \subset (\mathcal{X}_{-\lambda+\varepsilon} u)^c \text{ and} \tag{16.2}$$

$$(\mathcal{X}_{-\lambda+\varepsilon}(-u))^c \subset \mathcal{X}_{\lambda-\varepsilon} u, \tag{16.3}$$

for $\varepsilon > 0$. We then have

$$\begin{aligned}\mathcal{X}_\lambda(-\tilde{D}_a(-u)) &\overset{(16.2)}{\subset} \left[\mathcal{X}_{-\lambda+\varepsilon}(\tilde{D}_a(-u))\right]^c \overset{\text{a.e.}}{=} \left[\tilde{\mathcal{D}}_a \mathcal{X}_{-\lambda+\varepsilon}(-u)\right]^c \\ &\overset{\text{def.}\tilde{\mathcal{E}}_a}{=} \tilde{\mathcal{E}}_a\left(\mathcal{X}_{-\lambda+\varepsilon}(-u)\right)^c \overset{(16.3)}{\subset} \tilde{\mathcal{E}}_a(\mathcal{X}_{\lambda-\varepsilon} u) \overset{\text{a.e.}}{=} \mathcal{X}_{\lambda-\varepsilon}(\tilde{E}_a u),\end{aligned}$$

where the a.e. relations are true for every λ and almost every $\varepsilon > 0$ by the commutation with thresholds almost everywhere (16.1). By using the relation $\mathcal{X}_\lambda v = \cap_{\varepsilon>0} \mathcal{X}_{\lambda-\varepsilon} v$ with $v = \tilde{E}_a u$, we obtain $\mathcal{X}_\lambda(-\tilde{D}_a(-u)) \subset \mathcal{X}_\lambda(\tilde{E}_a u)$ almost everywhere for almost every λ. By taking $\varepsilon < 0$ it is easily checked that all inclusions in the above argument reverse. Thus almost all level sets of $\tilde{E}_a u$ and $-\tilde{D}_a(-u)$ are equal almost everywhere. By Corollary 8.17 and its consequence in Exercise 8.7 this implies that $\tilde{E}_a u$ and $-\tilde{D}_a(-u)$ coincide almost everywhere. Since in addition these functions belong to $\mathcal{F}$ and are therefore continuous, they coincide everywhere.

□

Exercise 16.7. Prove the relations (16.2) and (16.3) used in the proof of Proposition 16.12. ■

16.3 Principles for an algorithm

This section won't give an explicit algorithm for performing affine erosions or dilations, but rather a general principle from which algorithms can be derived.

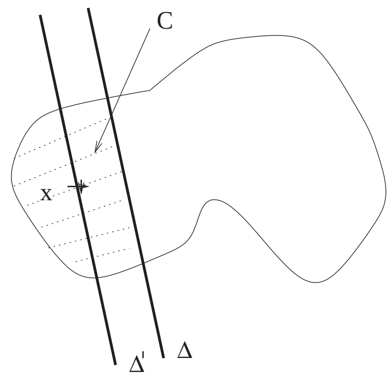

Figure 16.4: Illustration for Proposition 16.13.

Since $\tilde{\mathcal{D}}_a$ is obtained by duality from $\tilde{\mathcal{E}}_a$ we'll focus on the implementation of affine erosion. Since as we shall see a curve will be split into convex parts to apply the erosion, we can restrict ourselves to the case where X is convex.

The intuitive idea is that $\tilde{\mathcal{E}}_a X$ could be obtained from X by removing all chord-arc sets C defined by X and Δ that have area less than or equal to a, but this is not quite true. A simple example is given by taking for X the open disk $D(0,1)$. Removing all chord-arc sets with area less than or equal to $a < \frac{\pi}{2}$ will leave a closed disk $D(0,r)$, whereas $\tilde{\mathcal{E}}_a D(0,1)$ is the open disk $D(0,r)$. We need only make a small modification. If C is a chord-arc set defined by X and Δ, we define C^* by $C^* = C \cup (\overline{C} \cap \Delta)$.

Proposition 16.13. *Assume that X is an open, convex and bounded subset of $\mathbb{R}^2$. Then $\tilde{\mathcal{E}}_a X$ can be obtained from X by removing all of the modified chord arc sets C^* with area less than or equal to a.*

Proof. Let C be any chord-arc set defined by X and a line Δ such that $\text{area}(C) \le a$. Then we claim that $\mathbf{x} \in C^*$ implies that $\delta(\mathbf{x}, X^c) \le a^{1/2}$. To see this, let Δ' be the line parallel to Δ that contains $\mathbf{x}$. The lines Δ and Δ' each define two open half-planes, which we denote by R and L and R' and L'. Without loss of generality, we may assume that $\Delta' \subset R$, or equivalently, that $\Delta \subset L'$. (See Figure 16.4.)

Consider the sets $CA_i(\mathbf{x}, \Delta', X^c)$, $i = 1, 2$. These are the connected components of $\mathbb{R}^2 \setminus (\overline{X^c} \cup \Delta') = (\overline{X^c})^c \cap (\Delta')^c = X^\circ \setminus \Delta'$ that contain $\mathbf{x}$ in their boundaries. One of these sets, say, $CA_1(\mathbf{x}, \Delta', X^c)$ is in $R \cap R'$. Now $\overline{CA_1(\mathbf{x}, \Delta', X^c)} \cap C \neq \emptyset$, since both sets contain $\mathbf{x}$. This means that $CA_1(\mathbf{x}, \Delta', X^c) \cup C$ is connected, and since C was a maximal connected set in $X \setminus \Delta$ we conclude that $CA_1(\mathbf{x}, \Delta', X^c) \subset C$. Thus the area of $CA_1(\mathbf{x}, \Delta', X^c)$ is less than or equal to a, so $\delta(\mathbf{x}, X^c) \le a^{1/2}$, and $\mathbf{x} \notin \tilde{\mathcal{E}}_a X$.

For the converse, we must show that if a point was eroded, then it had to have been in a set C^*, where C is a chord-arc set with area less than or equal to a. Our proof uses a result that we prove in the next exercise 16.10, namely, that if $\delta(x, X^c) \le a^{1/2}$, then there is a line Δ that contains $\mathbf{x}$ such that $\delta(x, X^c) = [\text{area}(C)]^{1/2}$, where C is a chord-arc set defined by X and Δ. Then for this C, $\mathbf{x} \in C^*$ and $\text{area}(C) \leq a$. □

16.4 Affine Plane Curve Evolution Scheme.

Curve evolution applied to image level lines also yields an image evolution. Osher and Sethian proposed in order to simulate the evolution of a surface by the curvature motion by evolving its distance function by a finite difference scheme of the curvature motion. This strategy is quite well justified in dimension 3 but less in dimension 2 where level lines are simple Jordan curves. One can instead extract all level lines of the image and compute their evolution by the affine shortening. Each curve will be numerically represented as a polygon. The affine shortening is numerically defined as an *alternate filter*, which alternates affine erosion and affine dilation with a small parameter a.

16.4.1 A fast algorithm

The affine erosion of a set X is not simple to compute, because it is a strongly non local process. However, if X is convex, it has been shown in [205] 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 alternate affine erosion and dilation of X by computing the affine erosion of each *convex or concave 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 or concave parts.** 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 numerical issue is to take into account the finite precision of the computer in order to avoid spurious (small and almost straight) convex components.

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 an affine 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

$$d_i = (x_i - x_{i-1})(y_{i+1} - y_i) - (y_i - y_{i-1})(x_{i+1} - x_i) \tag{16.4}$$

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

$$\varepsilon_0 = \max\{x > 0,\ (1.0 \oplus x) \ominus 1.0 = 0.0\}. \tag{16.5}$$

In this definition, $\oplus$ (resp. $\ominus$) represent the computer addition (resp. substraction), which is not associative. When computing d_i using (16.4), the computer gives a result $\tilde{d}_i$ such that $|d_i - \tilde{d}_i| \leq e_i$, with

$$\begin{aligned} e_i = \varepsilon_0 \Big(& \quad |x_i - x_{i-1}|(|y_{i+1}| + |y_i|) + (|x_i| + |x_{i-1}|)|y_{i+1} - y_i| \\ & + \quad |y_i - y_{i-1}|(|x_{i+1}| + |x_i|) + (|y_i| + |y_{i-1}|)|x_{i+1} - x_i| \quad \Big). \end{aligned}$$

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 (16.5)) 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 $P_0P_1...P_{i-1}P_{i+1}...P_{n-1}$)
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 is the computation 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 sequence of middle points of all segments $[AB]$ 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.

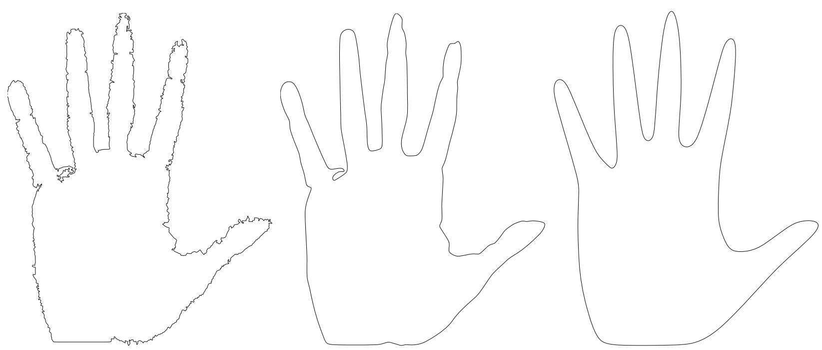

Figure 16.5: Affine scale space of a “hand” curve, performed with the alternate affine erosion-dilation scheme. (scales 1, 20, 400). Experiment : Lionel Moisan.

• 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,

$$E_{(h_1-h_0)^{3/2}} \circ E_{(h_2-h_1)^{3/2}} \circ \ldots \circ E_{(h_n-h_{n-1})^{3/2}} \ \Big(c_0\Big) \ \longrightarrow \ c_T$$

as $|h| = \max_i h_{i+1} - h_i \to 0$, where c_T is the affine shortening of c_0 up to scale T, described by the evolution equation (12.11). We refer to Chapters 18 and 20 for a proof of the equivalence between this affine invariant curve evolution and the above iterated alternate affine erosion-dilation scheme.

• **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.

16.5 Exercises

Exercise 16.8. The aim of this exercise is to prove that a one-to-one mapping $\tilde{A} : \mathbb{R}^2 \to \mathbb{R}^2$ that preserves parallelism must be of the form $\tilde{A}(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, where A is a linear mapping and $\mathbf{b}$ is a fixed vector. The preservation of parallelism is defined as follows: 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 $\lambda \in \mathbb{R}$, then there exists a $\mu \in \mathbb{R}$ such that $\tilde{A}\mathbf{x}_1 - \tilde{A}\mathbf{x}_2 = \mu(\tilde{A}\mathbf{x}_3 - \tilde{A}\mathbf{x}_4)$.

(i) Let $\boldsymbol{i}$ and $\boldsymbol{j}$ be the usual orthonormal basis for $\mathbb{R}^2$ and write $\mathbf{x} = x\boldsymbol{i} + y\boldsymbol{j}$. Define A by $A\mathbf{x} = \tilde{A}\mathbf{x} - \tilde{A}0$. Show that there are two real function $\mu_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2$, such that $A(x\boldsymbol{i}) = \mu_1(x)A\boldsymbol{i}$ and $A(y\boldsymbol{j}) = \mu_2(y)A\boldsymbol{j}$.

(ii) Notice that A preserves parallelism and that $A0 = 0$.

(iii) Show that $A\mathbf{x} = \mu_1(x)A\boldsymbol{i} + \mu_2(y)A\boldsymbol{j}$.

(iv) Show that $\mu_1(\lambda) = \mu_2(\lambda)$.

(v) We wish to show that $\mu_1(x) = x$. One way to do this is to prove that $\mu_1 : \mathbb{R} \to \mathbb{R}$ is an isomorphism. This can be done using the fact that $\mathbf{x}_1 - \mathbf{x}_2 = \lambda(\mathbf{x}_3 - \mathbf{x}_4)$ implies that $A\mathbf{x}_1 - A\mathbf{x}_2 = \mu(A\mathbf{x}_3 - A\mathbf{x}_4)$. Once you have shown that $\mu_1 : \mathbb{R} \to \mathbb{R}$ is an isomorphism, you can quote the result that says any isomorphism of $\mathbb{R}$ onto itself must be $x \mapsto x$, or you can prove this result. ■

Exercise 16.9. A justification of the main step of Moisan's algorithm. We refer to the "discrete affine erosion" step of the Moisan algorithm described in Section 16.4.1. It is said that a polygon approximating a the affine erosion of a convex set can be obtained by finding many chord-arc sets and taking the middle points of their chords as vertices of the eroded polygon. This follows from the fact that the chords of two nearby chord-arc sets with same area tend to meet at their common middle point.

1. Let $abcd$ be a quadrilateral with diagonals ac and bd. Let i be the crossing point of ac and bd. Assume that the areas of the triangles abi and icd are equal and that a and c are fixed points while b and d move in such a way that $d \to c$ and $b \to a$. Prove that the lengths $|ia|, |ib|, |ic|, |id|$ all tend to $\frac{|ac|}{2}$. Hint: to do so, prove that the area of the triangle idc is equivalent to $\frac{\theta.|id|^2}{2}$, where θ is the angle of id with ic.

2. Let C be a convex Jordan curve surrounding a convex set X and let Δ be a straight line meeting C at a and c. Call CA one of the two chord-arc sets defined by Δ and C. Let b be a point close to a on C and d a point close to c chosen in such a way that the chord-arc CA' defined by the line $\Delta' = bd$ and C has the same area as CA. Apply the result of the first question with $b \to a$.

3. Deduce from this and Proposition 16.13 that the Moisan algorithm computes an approximation to an affine erosion of a polygon.

■

Exercise 16.10. Assume that X is convex, open and bounded. We refer to Figure 16.6 below for the definitions of the various objects. Thus, $\Delta(0)$ is an arbitrary line that contains $\mathbf{x}$ and $C(0)$ is the connected component of $X \setminus \Delta(0)$ on the arrow-side of $\Delta(0)$ whose boundary contains $\mathbf{x}$. $C(\varphi)$ is the connected component of $X \setminus \Delta(\varphi)$ on the arrow-side of $\Delta(\varphi)$ whose boundary contains $\mathbf{x}$. Since X is open, there is an $r > 0$ such that the disk $D(x, r)$ is contained in X; since X is assumed to be bounded, there is an $R > 0$ such that $X \subset D(x, R)$.

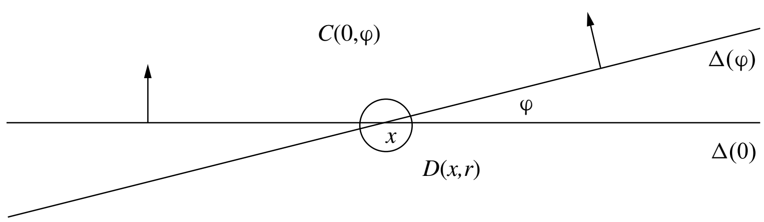

Figure 16.6: Definition of $C(0, \varphi)$. The set $C(0, \varphi)$ is the connected component of $X \setminus (\Delta(0) \cup \Delta(\varphi))$ that lies in the direction of the arrows.

(i) Show that $C(0, \varphi) \subset C(0)$ and $C(0, \varphi) \subset C(\varphi)$.

(ii) Show that $\text{area}(C(0) \cap C(0, \varphi)^c) \to 0$ as $\varphi \to 0$ and similarly that $\text{area}(C(\varphi) \cap C(0, \varphi)^c) \to 0$ as $\varphi \to 0$.

(iii) Deduce that $\text{area}(C(\varphi)) \to \text{area}(C(0))$ as $\varphi \to 0$.

This shows that $\text{area}(C(\varphi))$ is a continuous function of φ. Thus, the $\inf_{\Delta(\varphi)} \text{area}(C(\varphi))$ is attained, which means that for every $\mathbf{x} \in X$ there is some φ such that $\delta(\mathbf{x}, X^c) = [\text{area}(C(\varphi))]^{1/2}$.

Is the above result still true if X is not convex? ■

16.6 Comments and references

Shape-recognition algorithms in the plane are clearly more robust if they are affine invariant, if only because most optical devices that copy plane images (photocopiers) or that convert plane images to digital information (scanners, faxes) create a slight affine distortion. Also, all diffeomorphisms are locally affine. Affine-invariant techniques for matching shapes are described in [154]; discussions of the role of affine and projective invariance for object recognition can be found in [43], [285], and [164]. Corners and T-junctions can appear in images with arbitrary angles, and the detection of angles between straight lines should be affine invariant. Algorithms for affine-detection of angles are proposed in [41], [13], [275], and [90]. See Merriman, Bence, and Osher [204] for a very original numerical view for filtering multiple junctions. Because of the relevance to computer vision, there has been considerable research devoted to looking for affine-invariant definitions of classical concepts in geometric measure theory and integral geometry. An interesting attempt to define an "affine-invariant length" and an "affine-invariant dimension" analogous to Hausdorff lengths and dimensions is given in [91]. The diameters of the sets of a Hausdorff covering are simply replaced by their areas. Several attempts to define affine-invariant analyses of discrete sets of points are described in [121] and [247]. An affine-invariant symmetry set (skeleton) for shapes is defined in [123]; the 1/3 power law of planar motion perception and generation is related to affine invariance in [233]. Some of the techniques on affine erosions and dilations presented in this chapter were announced in [186]. We have made liberal use of the Matheron formalism for monotone set operators [202].

The fully invariant affine curve evolution geometric algorithm which we presented was found by Moisan [205]. Its implementation for *all* level lines of an image was realized in Koepfler [**?**]. 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. [186] and later Cao, Gousseau, Sur and Musé [**?**] have used the affine curve evolution scheme for shape recognition and image comparison algorithms.

Chapter 17

Localizable Structuring Elements and the Local Maximum Principle

Given a set of structuring elements $\mathcal{B}$, the scaled operators IS_h defined by

$$IS_h u(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y})$$

are immediately translation invariant and contrast invariant. Furthermore, if the elements of $\mathcal{B}$ are uniformly bounded, then the operators satisfy an important local property that we have not yet emphasized: If two functions u and v are such that $u(\mathbf{y}) \le v(\mathbf{y})$ for all $\mathbf{y}$ in some disk $D(\mathbf{x}, r)$, then for sufficiently small h, $hB \in D(0, r)$, and $IS_h u(\mathbf{x}) \le IS_h v(\mathbf{x})$. This is a special case of the *local maximum principle,* which for bounded structuring elements goes almost un-noticed. It might seem at first glance that we would not have a local maximum principle if the structuring elements were not bounded. It turns out, however, that for a large class of unbounded structuring elements, the operators IS_h behave as if they were local operators—in the sense that they satisfy a local maximum principle. For example, if the operators IS_h are affine invariant, then the affine-invariant structuring elements $\mathcal{B}$ cannot be bounded. Indeed, affine invariance allows an element $B \in \mathcal{B}$ to be stretched arbitrarily far in any direction: The matrix $A = \begin{pmatrix} \varepsilon & 0 \\ 0 & 1/\varepsilon \end{pmatrix}$, where ε is small, followed by a rotation, does the job. Nevertheless, there are affine-invariant operators that satisfy a local maximum principle. In particular, we will show that this is the case for $\mathcal{B} = \mathcal{B}_{\text{aff}}$. In general, the application of operators that satisfy this property involves an error term:

$$IS_h u(\mathbf{x}) \le IS_h v(\mathbf{x}) + o(h^\beta).$$

We are going to define a property of structuring elements $\mathcal{B}$ called *localizability*, and even though $\mathcal{B}$ may contain arbitrarily large elements, or even unbounded elements, if it is localizable, then the inf-sup operators defined by $\mathcal{B}$ will satisfy a local maximum principle. The importance of the local maximum principle for our program will become clear in the chapter on viscosity solutions. Since our focus is on affine-invariant operators, we will apply the concept of localizability only to families of affine-invariant structuring elements, but the reader should keep in mind that concept is applicable to other situations. For example, the structuring elements associated with a median filter defined by a k that does not have compact support, are unbounded, but they may be localizable. A case in point is the Gaussian: It does not have compact support, but it can be shown that the structuring elements are localizable.

17.1 Localizable sets of structuring elements

Recall that the Euclidean distance d between a point $\mathbf{x}$ and a set Y is defined by

$$d(\mathbf{x}, Y) = \inf_{\mathbf{y} \in Y} |\mathbf{x} - \mathbf{y}|.$$

It will be convenient to use the following notation: $D(0,\rho)$ will denote the open disk (or ball) $\{\mathbf{x} \mid |\mathbf{x}| < \rho\}$, and D_a will denote the dilation operator defined by

$$D_a(X) = \{\mathbf{x} \mid d(\mathbf{x}, X) < a\}.$$

In the notation of Chapter 9, this means that $D_a = \mathcal{D}_a$, where the structuring element for $\mathcal{D}_a$ is $D(0,1)$. Note that if X is open and connected, then $D_a(X)$ is open and connected. We will write ∂X to denote the boundary of X. These definitions and notation are used to define the concept of a set of structuring elements being localizable.

For convenience, we introduce two set operators:

- $\mathcal{C}_0[X] =$ the connected component of X that contains the origin.
- $\mathcal{C}_0^{\partial}[X] =$ the connected component of X that contains the origin in its boundary.

The set X will always be open and either contain the origin or contain the origin in its boundary. Note that these operators commute with scaling: For example, $\mathcal{C}_0[hX] = h\mathcal{C}_0[X]$. Note also that these operators are monotone.

Definition 17.1. *Let $\alpha > 0$ be a positive constant. Assume that $\mathcal{B}$ is a set of structuring elements whose members are open and contain the origin. $\mathcal{B}$ is said to be α-localizable if there are two constants $c > 0$ and $R > 0$, where c and R depend on $\mathcal{B}$, such that for every $\rho > R$ the following conditions holds: For each $B \in \mathcal{B}$, there is a $B' \in \mathcal{B}$ such that*

(i) $B' \subset D(0,\rho)$;

(ii) $B' \subset D_{c/\rho^\alpha}(\mathcal{C}_0[B_\rho])$, *where* $B_\rho = B \cap D(0,\rho)$.

The constant α is called the exponent of localizability.

We wish to emphasize that our definition of localizable includes the assumption that the elements of $\mathcal{B}$ are open and contain the origin. Note also that a localizable set $\mathcal{B}$ contains bounded members. Also, it may happen that $\mathcal{C}_0[D_{c/\rho^\alpha}(\mathcal{C}_0[B_\rho]) \cap D(0,\rho)]$ is itself a member of $\mathcal{B}$ for all $B \in \mathcal{B}$; indeed, the proof of Proposition 13.4 shows that this is the case if $\mathcal{B} = \mathcal{B}_{\text{aff}}$. The next result shows how the concept of localizability scales.

Proposition 17.2. *Let h, $0 < h \leq 1$, be a scaling factor and assume the notation of Definition 13.1. A set of structuring elements $\mathcal{B}$ is α-localizable if and only if there are constants $c > 0$ and $R > 0$ such that for all $r > 0$ and all $h < r/R$ the following conditions holds: For each $B \in h\mathcal{B}$, there is a $B' \in h\mathcal{B}$ such that*

(*i'*) $B' \subset D(0, r)$;

(*ii'*) $B' \subset D_{ch^{\alpha+1}/r^\alpha}(\mathcal{C}_0[B_r])$, *where* $B_r = B \cap D(0, r)$.

Proof. Assume the conditions of the proposition. Then $r/R > 1$ for all $r > R$. Since h can be any number in the range $0 < h < r/R$, the conditions of the proposition are true for $h = 1$. By letting $\rho = r$, we have the conditions of Definition 13.1.

To prove that the conditions of Definition 13.1 imply the conditions of the proposition, let $\rho = r/h$. Then the statement "all $\rho > R$" is equivalent to the statement "all $r > 0$ and all $h < r/R$." Next, we need to see how $D_{c/\rho^\alpha}(X)$ scales:

$$\begin{aligned} hD_{c/\rho^\alpha}(X) &= \{h\mathbf{x} \mid d(\mathbf{x}, X) < c/\rho^\alpha\} \\ &= \{h\mathbf{x} \mid d(h\mathbf{x}, hX) < ch/\rho^\alpha\} \\ &= \{\mathbf{y} \mid d(\mathbf{y}, hX) < ch/\rho^\alpha\} \\ &= D_{ch/\rho^\alpha}(hX). \end{aligned}$$

If $X = \mathcal{C}_0[B_\rho]$, then $hD_{c/\rho^\alpha}(\mathcal{C}_0[B_\rho]) = D_{ch^{\alpha+1}/r^\alpha}(\mathcal{C}_0[hB_\rho])$, where $hB_\rho = B_r$. In other words, $B' \in D_{c/\rho^\alpha}(\mathcal{C}_0[B_\rho])$ implies that $hB' \in D_{ch^{\alpha+1}/r^\alpha}(\mathcal{C}_0[B_r])$, which shows that (ii) implies (ii'). □

We will use Definition 13.1 and its scaled version, Proposition 13.1, to prove two results: The first is that if $\mathcal{B}$ is localizable, then the IS_h satisfy the local maximum principle; the second is that $\mathcal{B}_{\text{aff}}$ is 1-localizable.

17.2 The local maximum principle

While the notion of α-localizability has an important role in mathematical morphology, we are concerned in this book only with the 1-localizable families of structuring elements. Thus, from this point, we assume that $\alpha = 1$ and leave the general cases as exercises.

Lemma 17.3 (local maximum principle). *Let $\mathcal{B}$ be a 1-localizable set of structuring elements with the associated constants $c > 0$ and $R > 0$. Assume that the functions u and v satisfy a Lipschitz condition on a disk $D(\mathbf{x}, r)$ with Lipschitz constant L. If $u(\mathbf{y}) \leq v(\mathbf{y})$ for $\mathbf{y} \in D(\mathbf{x}, r)$ and if $h < r/R$, then*

$$IS_h u(\mathbf{x}) \leq IS_h v(\mathbf{x}) + Lc\frac{h^2}{r} \quad \text{and} \quad SI_h u(\mathbf{x}) \leq SI_h v(\mathbf{x}) + Lc\frac{h^2}{r}.$$

Proof. For notational convenience, we take $\mathbf{x} = 0$. Then

$$IS_h v(0) = \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B} v(\mathbf{y}) \geq \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B \cap D(0,r)} v(\mathbf{y}) \geq \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B \cap D(0,r)} u(\mathbf{y}). \tag{17.1}$$

By Proposition 13.2, for all $B \in h\mathcal{B}$, there exists $B' \in h\mathcal{B}$ such that $B' \subset D(0,r)$ and $B' \subset D_{ch^2/r}(\mathcal{C}_0[B \cap D(0,r)])$. Thus, since u is Lipschitz with constant L and since for every point $b' \in B'$ there is a point $b \in B \cap D(0,r)$ such that $|b' - b| < ch^2/r$, we have

$$\sup_{\mathbf{y} \in B \cap D(0,r)} u(\mathbf{y}) \geq \sup_{\mathbf{y} \in B'} u(\mathbf{y}) - Lc\frac{h^2}{r}. \tag{17.2}$$

These last two inequalities imply that

$$IS_h v(0) + Lc\frac{h^2}{r} \geq \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B'} u(\mathbf{y}),$$

where B' is any set such that $B' \in h\mathcal{B}$, $B' \subset D(0,r)$, and $B' \subset D_{ch^2/r}(\mathcal{C}_0[B \cap D(0,r)])$. If we denote the family of all such sets B' associated with B by $\mathcal{B}' = \mathcal{B}'(B)$, then a more precise statement is that

$$IS_h v(0) + Lc\frac{h^2}{r} \geq \inf_{B \in h\mathcal{B}} \Big(\sup_{B' \in \mathcal{B}'} \sup_{\mathbf{y} \in B'} u(\mathbf{y}) \Big).$$

Let $\inf_{B \in h\mathcal{B}} \Big(\sup_{B' \in \mathcal{B}'} \sup_{\mathbf{y} \in B'} u(\mathbf{y}) \Big) = \lambda$. Then we claim that $\inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B} u(\mathbf{y}) \leq \lambda$. To see this, let $\varepsilon > 0$ be arbitrary. By the definition of λ, there is some set $B \in h\mathcal{B}$ such that $\sup_{\mathbf{y} \in B} u(\mathbf{y}) \leq \lambda + \varepsilon$. Thus, $\inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B} u(\mathbf{y}) \leq \lambda + \varepsilon$, which, since ε is arbitrary, implies that $\inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in B} u(\mathbf{y}) \leq \lambda$. This yields the result:

$$IS_h u(\mathbf{x}) = \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}) \leq IS_h v(\mathbf{x}) + Lc\frac{h^2}{r}.$$

The result for SI_h follows from the relation $-IS_h(-u) = SI_h(u)$. □

Note that the proof for IS_h does not use the fact that v is locally Lipschitz. In fact, the proof works for any v. The problem with this is that we would not have the result for SI_h if we did not assume that v is locally Lipschitz. Also, taking $\mathbf{x} = 0$ in the proof is indeed only a notational convenience; knowing that the assumptions hold at $\mathbf{x}$ does not imply they hold elsewhere. Thus, as stated, the result of Lemma 13.3 is strictly local.

Lemma 13.3 and the next lemma provide the links between the localizability of structuring elements and the local properties of the associated operators. As such, they lie near the heart of our program. Their use is the key to demonstrating the asymptotic behavior of inf-sup operators defined by 1-localizable families of affine-invariant structuring elements. The local maximum principle is also used in the proof of the Barles–Souganidis theorem, Proposition 15.13, which is essential for relating the inf-sup operators to their associated PDEs via viscosity solutions of the PDEs.

Exercise 17.1. Prove the general form of Lemma 13.3: Replace the hypothesis that $\mathcal{B}$ is 1-localizable with the hypothesis that it is α-localizable and conclude that

$$IS_h u(\mathbf{x}) \leq IS_h v(\mathbf{x}) + Lc\frac{h^{\alpha+1}}{r^\alpha}. \ \blacksquare$$

Lemma 13.3 compares the action of IS_h on two functions u and v. In the next lemma, we consider an operator IS_h^r that approximates IS_h and examine its action on a single function u. The approximate operator IS_h^r is defined by truncating the structuring elements $\mathcal{B}$: We replace the family $\mathcal{B}$ with the family $\mathcal{B}_r = \{B_r \mid B_r = B \cap D(0,r), B \in \mathcal{B}\}$. Thus,

$$IS_h^r u(\mathbf{x}) = \inf_{B_r \in h\mathcal{B}_r} \sup_{\mathbf{y} \in \mathbf{x}+B_r} u(\mathbf{y}) = \inf_{B \in h\mathcal{B}} \sup_{\mathbf{y} \in (\mathbf{x}+B) \cap D(\mathbf{x},r)} u(\mathbf{y}).$$

The local properties of IS_h^r have been imposed by definition. Later, when we apply this result, we will take $r = h^{1/2}$, so the error term will be $Lch^{3/2}$.

Lemma 17.4 (localization lemma). *Let $\mathcal{B}$ be a 1-localizable set of structuring elements with constants $c > 0$ and $R > 0$ and assume that $h < r/R$. If u satisfies a Lipschitz condition in $D(\mathbf{x}, r)$ with constant L, then*

(*i*) $IS_h^r u(\mathbf{x}) \le IS_h u(\mathbf{x}) \le IS_h^r u(\mathbf{x}) + Lch^2/r$;

(*ii*) $|IS_h^r u(\mathbf{x}) - IS_h u(\mathbf{x})| \le Lch^2/r$;

(*iii*) $|SI_h^r u(\mathbf{x}) - SI_h u(\mathbf{x})| \le Lch^2/r$.

(*iv*) $|SI_h^r IS_h^r u(\mathbf{x}) - SI_h IS_h u(\mathbf{x})| \le 2Lch^2/r$, *if u is L-Lipschitz on $\mathbb{R}^N$*

Proof. By taking $u = v$ in inequality (17.1), we see that $IS_h u(\mathbf{x}) \ge IS_h^r u(\mathbf{x})$. This half of (*i*) does not depend on u being Lipschitz on $D(\mathbf{x}, r)$, but the other half of (*i*) does depend on u being Lipschitz on $D(\mathbf{x}, r)$. To prove the other half of (*i*), we are going to follow the proof of Lemma 13.3, including the notational convenience that $\mathbf{x} = 0$. In particular, we use the 1-localizability of $\mathcal{B}$ to establish the inequality

$$\sup_{\mathbf{y} \in B \cap D(0,r)} u(\mathbf{y}) \ge \sup_{\mathbf{y} \in B'} u(\mathbf{y}) + Lc\frac{h^2}{r},$$

which is (17.2). The remainder of the proof shows that

$$IS_h^r u(\mathbf{x}) + Lc\frac{h^2}{r} \ge IS_h u(\mathbf{x}),$$

and this proves the other half of (*i*).

Inequality (*ii*) is just a restatement of (*i*). Inequality (*iii*) is deduced from (*ii*) by using the relation $IS_h(-u) = -SI_h u$. To prove (*iv*), first recall from Lemma 6.5 that if u is Lipschitz with constant L, then $IS_h u$ and $IS_h^r u$ are Lipschitz with constants no greater than L. By (*ii*) and (*iii*) we have

$$IS_h^r u(\mathbf{x}) \le IS_h u(\mathbf{x}) \le IS_h^r u(\mathbf{x}) + Lch^2/r; \tag{17.3}$$

$$SI_h^r u(\mathbf{x}) \le SI_h u(\mathbf{x}) \le SI_h^r u(\mathbf{x}) + Lch^2/r. \tag{17.4}$$

Replacing u with $IS_h u$ in (17.4) and applying SI_h^r to (17.3) shows that

$$SI_h^r IS_h^r u(\mathbf{x}) \le SI_h IS_h u(\mathbf{x}) \le SI_h^r IS_h^r u(\mathbf{x}) + 2Lch^2/r,$$

which proves (*iv*). □

The statements and proofs of Lemmas 13.3 and 13.4 are strictly local. There are, however, immediate global generalizations, and since these more general results are important for later applications, we give them a precise statement for future reference.

Lemma 17.5. *Let K be an arbitrary set and assume that the function u and v in Lemmas 13.3, 13.4(i), 13.4(ii), and 13.4(iii) are L-Lipschitz in $D(\mathbf{x}, r)$ for every $\mathbf{x} \in K$. Then the results of these lemmas are true uniformly for $\mathbf{x} \in K$.*

These uniform results need no special proofs. One merely rereads the proofs of the local lemmas and notes that, if the same hypotheses hold at each point $\mathbf{x} \in K$, then the results are true for each $\mathbf{x}$ with exactly the same error term. When we apply Lemma 13.5, K will be compact, but clearly this is not a necessary condition for the lemma.

The next result is a direct consequence of Lemma 13.3. It allows us to fix an optimal relation between the localization scale r and the operator scale h. Again, it is a local result that can easily be made uniform.

Lemma 17.6. *Let $\mathcal{B}$ be a 1-localizable set of structuring elements with constants $c > 0$ and $R > 0$. Let u and v be two continuous functions that satisfy Lipschitz conditions with the same constant L on a disk $D(0, r)$. If*

$$|u(\mathbf{x}) - v(\mathbf{x})| \leq C|\mathbf{x}|^3$$

for $\mathbf{x} \in D(0, r)$, and if $h \leq r^2$ and $h < 1/R^2$, then

$$|IS_h u(0) - IS_h v(0)| \leq (C + Lc)h^{3/2}.$$

Proof. The relation $v(\mathbf{x}) - Cr^3 \leq u(\mathbf{x}) \leq v(\mathbf{x}) + Cr^3$ is true for all $\mathbf{x} \in D(0, r)$, so we can apply Lemma 13.3 and conclude that

$$IS_h v(0) - Cr^3 - Lc\frac{h^2}{r} \leq IS_h u(0) \leq IS_h v(0) + Cr^3 + Lc\frac{h^2}{r}$$

for $h < r/R$. This argument is also true for $0 < s \leq r$, if we have $h < s/R$. So, in particular, if we take $s = h^{1/2} \leq r$ and $h < s/R$, that is, $h < 1/R^2$, we have

$$IS_h v(0) - Ch^{3/2} - Lch^{3/2} \leq IS_h u(0) \leq IS_h v(0) + Ch^{3/2} + Lch^{3/2},$$

which proves the result. □

Here, we have taken the point of view that r is given, and we ask that $h = r^2$. In other situation, we may take the opposite view and ask that r be determined by h. This is the case, for example, if we are able to choose the size of r for the localized operator IS_h^r.

The main application of Lemma 13.4 is to reduce the asymptotic analysis of the operator IS_h as $h \to 0$ to the case where it is applied to quadratic polynomials. (We have seen in Chapters 10 and 11 how this kind of analysis works in the case of structuring elements that are bounded and isotropic.)

17.3 $\mathcal{B}_{\text{aff}}$ is 1-localizable

We are going to prove that $\mathcal{B}_{\text{aff}}$ is 1-localizable, but to make things as transparent as possible, we first do some geometry. Thus, consider Figure 17.1.

We are interested in the area of $D_{c/\rho}([0, z]) \cap D(0, \rho) \cap \Delta^c$. This is the area of the figure $ABCD$. In what follows, $c > 0$ is a constant. For the figure to

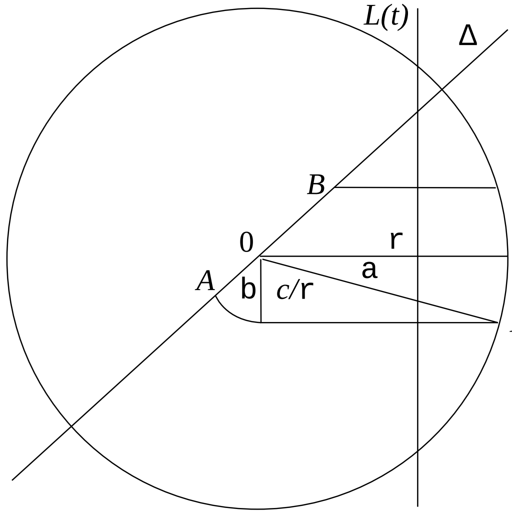

Figure 17.1: Dilation of $[0, z]$.

make sense, we must assume that $\rho^2 \geq c$. We want a lower bound for the area of the set $ABCD$ that will hold for any $\beta \in [0, \pi/2]$. (In the limit case $\beta = \pi/2$, and $z \in \Delta$.) We also wish to compare this area with that of another set, and the way we do this limits us to using the set $A0zD$. (The reason for this will become clear.) Denote the area of $A0zD$ by $\mathcal{A} = \mathcal{A}(\alpha, \beta)$. Thus, assuming $\rho^2 \geq c$, we have

$$\mathcal{A}(\alpha, \beta) = \frac{c}{2}\Big(\beta \sin\alpha + \frac{\alpha}{\sin\alpha} + \cos\alpha\Big).$$

This is written in terms of α and β because it is easy to see what happens as α ranges from zero to $\pi/2$. This is equivalent to ρ going from $+\infty$ to $\sqrt{c}$, which is just the range of interest. The smallest value of the function

$$f(\alpha, \beta) = \beta \sin\alpha + \frac{\alpha}{\sin\alpha} + \cos\alpha$$

for $0 \leq \alpha \leq \pi/2$, $0 \leq \beta \leq \pi/2$ occurs at $\alpha = \pi/2$, $\beta = 0$, and $\mathcal{A}(\pi/2, 0) = (\pi/4)c$ for these values.

To avoid repeating it, we assume that the set $D(0, \rho) \cap \Delta^c$ always denotes the same open half-disk, and all of the sets we consider are understood to lie on the same side of Δ as $D(0, \rho) \cap \Delta^c$.

Parameterize the segment $[0, z]$ so the points are represented by tz, $t \in [0, 1]$. Let $L(t)$ denote the line orthogonal to $[0, z]$ at tz. Let y_t be any point on $L(t)$ such that $y_t \in D(0, \rho) \cap \Delta^c$, and consider the set $D(y_t, c/\rho) \cap D(0, \rho) \cap \Delta^c$. Then the open segment $L(t) \cap D(y_t, c/\rho) \cap D(0, \rho) \cap \Delta^c$ is always at least as long as the segment on the same line defined by $D(tz, c/\rho) \cap L(t) \cap (A0zD)$. Note that this is true for all $t \in [0, 1]$, even for those t close to one.

Lemma 17.7. *Assume the geometry and notation of Figure 17.1. Let Γ be any Jordan arc connecting the origin and z and lying completely in the set $D(0, \rho) \cap \Delta^c$, except for the end points. Then*

$$\operatorname{area}\big(D_{c/\rho}(\Gamma) \cap D(0, \rho) \cap \Delta^c\big) \geq \operatorname{area}(A0zD) \geq \frac{\pi}{4}c$$

whenever $\rho^2 \geq c$.

Proof. Let f denote the characteristic function of $D_{c/\rho}(\Gamma) \cap D(0,\rho) \cap \Delta^c = G$. Then

$$\operatorname{area}(G) = \int_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

By Fubini's theorem,

$$\operatorname{area}(G) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x,$$

and

$$\operatorname{area}(G) \geq \frac{c}{2}\beta \sin\alpha + \int_0^1 \left(\int_{\mathbb{R}} f(tz,y)\,\mathrm{d}y \right) \mathrm{d}t.$$

For fixed t,

$$\int_{\mathbb{R}} f(tz,y)\,\mathrm{d}y = \int_{l(t)} f(tz,y)\,\mathrm{d}y,$$

where $l(t) = G \cap L(t)$. We know that $l(t)$ contains a line segment at least as long as $(A0zD) \cap L(t)$, so

$$\operatorname{area}(G) \geq \frac{c}{2}\beta \sin\alpha + \operatorname{area}(0zDE) = \operatorname{area}(A0zD).$$

Note that in adding the term $(c/2)\beta \sin\alpha$ to the area of G, we use the fact that the origin is a point of Γ. □

Ostensibly, this lemma has little to do with $\mathcal{B}_{\mathrm{aff}}$. The lemma only compares the area of $D_{c/\rho}(\Gamma) \cap D(0,\rho) \cap \Delta^c$, where Γ is a Jordan arc that connects the origin to $z \in \partial D(0,\rho)$, with the area of $D_{c/\rho}([0,z]) \cap D(0,\rho) \cap \Delta^c$. This is a purely geometric result, however, the application to $\mathcal{B}_{\mathrm{aff}}$ is direct.

Before stating and proving the theorem, we note that all of the connected sets involved in the proof are open and thus arcwise connected. As usual, Δ denotes a straight line through the origin, and Δ^c always denotes the same open half-plane. If A is an open set that contains the origin, then the set $A \cap \Delta^c$ will contain a half-neighborhood $D(0,\varepsilon) \cap \Delta^c$ for some $\varepsilon > 0$.

Proposition 17.8. $\mathcal{B}_{\mathrm{aff}}$ *is* 1*-localizable.*

Proof. We must exhibit a $c >$ and an $R > 0$ such that the conditions of Definition 13.1 hold. Taking a clue from Lemma 13.6, we wish to have $c > 4/\pi$, so we take $c = 2$ and $R = \sqrt{2}$. These are not the "best" constants; we only claim that they work.

Let B be any element of $\mathcal{B}_{\mathrm{aff}}$. Then B is open and connected, B contains the origin, and $\delta(0,B^c) > 1$, or equivalently, given any Δ through the origin, the two connected components of $B \cap \Delta^c$ that contain the origin in their boundaries always have areas greater than or equal to $b = \delta(0,B^c)$.

Three open connected sets enter the proof:

$$B' = \mathcal{C}_0[D_{c/\rho}(\mathcal{C}_0[B \cap D(0,\rho)]) \cap D(0,\rho)].$$

$$C' = \mathcal{C}_0^{\partial}[B' \cap \Delta^c].$$

$$C = \mathcal{C}_0^{\partial}[B \cap \Delta^c].$$

By our convention regarding the use of Δ^c, the sets C and C' lie on the same side of Δ.

The plan is to show that B' is in $\mathcal{B}_{\text{aff}}$. Since B' is open and connected, it remains to exhibit a $c' > 1$ that does not depend on Δ such that $\text{area}(C') \geq c' > 1$.

There are two cases: $C \subset D(0,\rho)$ and $C \not\subset D(0,\rho)$. If $C \subset D(0,\rho)$, then, by the definition of $\mathcal{B}_{\text{aff}}$, $\text{area}(C) \geq b$. (This is the only place where the definition of $\mathcal{B}_{\text{aff}}$ is used.) In this case, $C \subset C'$, and $\text{area}(C') \geq b > 1$. Thus, $c' = b$ works for this case, and we are done. (The proof that $C \subset C'$ is left as an exercise.)

If $C \not\subset D(0,\rho)$, then there is a point $\boldsymbol{a} \in C$ such that $|\boldsymbol{a}| > \rho$. There is a Jordan arc $\gamma \in C$ that connects the origin to $\boldsymbol{a}$. (In fact, we may assume that this Jordan arc is piecewise linear.) Let $t : [0,1] \to \gamma$ be a parameterization such that $\gamma(0) = 0$ and $\gamma(1) = \boldsymbol{a}$. Then there is a smallest $t = t_0$ such that $\gamma(t_0) \in \partial D(0,\rho)$. Call this point z and let Γ denote the part of γ defined by $0 \leq t \leq t_0$. The arc Γ lies in $C \cap D(0,\rho) \cap \Delta^c$, and in particular, $\Gamma \subset B \cap D(0,\rho) \cap \Delta^c$. It follows that $D_{c/\rho}(\Gamma) \cap D(0,\rho) \cap \Delta^c \subset C'$.

By Lemma 13.6, $\text{area}(D_{c/\rho}(\Gamma) \cap D(0,\rho) \cap \Delta^c) \geq (\pi/4)c$, so $\text{area}(C') > \pi/2$, by the definition of c. Thus, by taking $c' = \min\{\pi/2, b\}$, it is always true that $\text{area}(C') \geq c' > 1$, and c' does not depend on Δ. It follows from Definition 13.1 that $\mathcal{B}_{\text{aff}}$ is 1-localizable. □

Exercise 17.2. We skipped over two points in the proof that the reader should check. The first was when we stated that $C \subset C'$ (the case $C \subset D(0,\rho)$), and the second was when we claimed that $D_{c/\rho}(\Gamma) \cap D(0,\rho) \cap \Delta^c \subset C'$ (the case $C \not\subset D(0,\rho)$). (Hint: All of the sets in sight are open and connected, so they are arcwise connected.) ■

The next two exercises show that there are other affine-invariant families of structuring elements that are 1-localizable.

Exercise 17.3. Let $\mathcal{B}$ be an affine-invariant family of open convex sets, each of which contains the origin and has area less than one. The goal is to show that $\mathcal{B}$ is 1-localizable. Here is one way to do this. Suppose $\rho > R$, where $R > 0$ is a constant to be determined, and that B is an element of $\mathcal{B}$. If $B \subset D(0,\rho)$, take $B' = B$, and we are done. If not, let $\mathbf{x}$ be a vector such that $|\mathbf{x}| = \xi = \sup_{\mathbf{y} \in \overline{B}} |\mathbf{y}|$. Establish the coordinate system based on $\boldsymbol{i} = \mathbf{x}/|\mathbf{x}|$ and $\boldsymbol{j} = \boldsymbol{i}^{\perp}$. Consider the affine transformation defined by $A = \begin{pmatrix} \rho/\xi & 0 \\ 0 & \xi/\rho \end{pmatrix}$, and show that if ρ is greater than some constant, then $AB \subset D(\rho, 0)$. Let $B' = AB$, which belongs to $\mathcal{B}$ by assumption. (Hint: Let η be the longest perpendicular distance from the x-axis to B^c. Relate the product $\eta\xi$ to the area of B and determine a value for R that works.) Having found an R such that $B' \subset D(0,\rho)$ for $\rho > R$, look for a $c > 0$ such that $B' \subset D_{c/\rho}(B)$. ■

Exercise 17.4. Let B be a bounded, open, and connected set that contains the origin. Define the affine-invariant family $\mathcal{B}$ by $\mathcal{B} = \{AB \mid A \in SL(\mathbb{R}^2)\}$. Use the methods of Exercise 3.2 to show that $\mathcal{B}$ is 1-localizable. ■

17.4 Comments and references

We were vague about the domains of the various functions that appear in section 13.2. In fact, the results are true in $\mathbb{R}^N$ even though we spoke of "disks" rather than "balls." The results in section 13.3 are, however, strictly limited to $\mathbb{R}^2$.

The mathematical techniques for localizing a set of structuring elements developed in this chapter were first explained in [134], [133], and [130]. The version presented here is much simpler. The doctoral dissertations by Frédéric Cao and Denis Pasquignon contain related techniques [52, 53, 227].

Chapter 18

Asymptotic Behavior of Affine-Invariant Filters

We are going to analyze the asymptotic behavior of affine-invariant operators in much the same way we analyzed contrast-invariant isometric operators in Chapters 10 and 11. The analysis in this chapter will be in $\mathbb{R}^2$. Recall that when we say an operator T is affine invariant, we mean that T commutes with all elements of the special linear group $SL(\mathbb{R}^2)$. Thus, for $A \in SL(\mathbb{R}^2)$, we have $ATu = TAu$ for all functions u in the domain of T, where Au is defined by $Au(\mathbf{x}) = u(A\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$. At this point, there are two possible scenarios: Assume we are given an affine-invariant operator T that is also contrast and translation invariant and then use Theorem 7.3 to conclude that T can be represented as

$$Tu(\mathbf{x}) = \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}), \tag{18.1}$$

where the set of structuring elements $\mathcal{B}$ may be taken to be $\{X \mid 0 \in \mathcal{T}X\}$, $\mathcal{T}X = \mathcal{X}_1 T\mathbf{1}_X$, and where (14.1) holds almost everywhere for u in the domain of T. The other approach, which is the one we take, is to assume the set of structuring elements $\mathcal{B}$ is given and to define T by (14.1). This places the focus on $\mathcal{B}$. With this approach, we know immediately that T is contrast and translation invariant and that it is defined on all $u : \mathbb{R}^2 \to \mathbb{R}$. We are, however, left with the task of proving that T is affine invariant if and only if $\mathcal{B}$ is affine invariant. (This is the content of Exercise 14.1.). Again, it is understood that, in our context, "affine invariant" always means "invariant with respect to $SL(\mathbb{R}^2)$."

We assume that $\mathcal{B}$ is a set of affine-invariant structuring elements, and we define for every $u : \mathbb{R}^2 \to \mathbb{R}$,

$$\begin{aligned} SI_h u(\mathbf{x}) &= \sup_{B \in \mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y}); \\ IS_h u(\mathbf{x}) &= \inf_{B \in \mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+hB} u(\mathbf{y}). \end{aligned} \tag{18.2}$$

$SI_h u$ is considered to be an *affine erosion* of u, and $IS_h u$ is considered to be an *affine dilation* of u. (Note that this nomenclature is consistent with the definitions of $\tilde{\mathcal{E}}_a$ and $\tilde{\mathcal{D}}_a$ in Chapter 12.) Since we have the relation $SI_h u = -IS_h(-u)$, it suffices to study just one of these operators, and we choose to investigate IS_h. Our main concern is the behavior of $IS_h u(\mathbf{x})$ as $h \to 0$ for $u \in C^3(\mathbb{R}^2)$. We will prove that, if $\mathcal{B}$ is affine invariant and 1-localizable, then

$$\lim_{h\to 0} \frac{IS_h u(\mathbf{x}) - u(\mathbf{x})}{h^{4/3}} = c_{\mathcal{B}} |Du(\mathbf{x})| \Big(\frac{1}{2}\mathrm{curv}(u)(\mathbf{x})^+\Big)^{1/3},$$

where $c_{\mathcal{B}}$ is a suitable constant. (As before, $r^{1/3}$ means $(r/|r|)|r|^{1/3}$.)

Exercise 18.1. Show that T is affine invariant if and only if $\mathcal{B}$ is affine invariant. ■

18.1 The analysis of IS_h

The analysis of affine-invariant operators IS_h will follow the general plan outlined in section 10.1.1 and exemplified by Theorem 10.2, with the important difference that the structuring elements $\mathcal{B}$ are not bounded. They are, however, isometric, since the group of isometries is a subgroup of the $SL(\mathbb{R}^2)$. This means that given a C^3 function u, we can expand it in the form

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

where $\mathbf{y} = (x, y)$ and the linear term is px. (We use the notation and conventions of section 4.5.) If we assume that this expansion holds for $\mathbf{y} \in D(0, r)$, then the analysis of the error term R given in the proof of Theorem 10.2 implies that

$$|u(\mathbf{x}+\mathbf{y}) - u(\mathbf{x}) - (px + ax^2 + by^2 + cxy)| \le \sup_{|\mathbf{y}|\le r} \|D^3(\mathbf{x}+\mathbf{y})\| |\mathbf{y}|^3.$$

Define v for $\mathbf{y} \in D(0,r)$ by $v(\mathbf{y}) = u(\mathbf{x}) + px + ax^2 + by^2 + cxy$. If $\mathcal{B}$ is 1-localizable, and if $h \le r^2$ and $h < 1/R^2$, then we know from Lemma 13.6 that

$$|IS_h u(\mathbf{x}) - IS_h v(\mathbf{x})| \le (C + Kc)h^{3/2}.$$

(Here and elsewhere we use the fact that, if u is locally C^3, then it is locally Lipschitz. Also, refer to sections 13.1 and 13.2 for the meaning of the constants.) This implies that the analysis of IS_h can be reduced to analyzing the action of IS_h on polynomials of degree two. This analysis will be done in Theorem 14.4, but before we get there, we need to consider the action of IS_h on two specific polynomials. Because the cases $a = c = 0$ and $b = 1$ or $b = -1$ play key roles, we introduce special notation:

$$c_{\mathcal{B}}^+ = \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B} (x + y^2) \quad \text{and} \quad c_{\mathcal{B}}^- = \inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B} (x - y^2),$$

where "$IS_h(px+ax^2+by^2+cxy)$" always means "$IS_h(px+ax^2+by^2+cxy)(0)$." Since our main results use these constants, it is worth examining some examples.

Lemma 18.1. *(i) Let $\mathcal{B}$ be an affine-invariant family of open convex sets that have area one and that are symmetric with respect to the origin. Then $c_{\mathcal{B}}^+ > 0$ and $c_{\mathcal{B}}^- = 0$. (ii) If $\mathcal{B} = \mathcal{B}_{\mathrm{aff}}$, then $c_{\mathcal{B}}^+ > 0$ and $c_{\mathcal{B}}^- = 0$. (iii) If a set of structuring elements $\mathcal{B}$ is affine invariant and contains one bounded element that is open and contains the origin, then $c_{\mathcal{B}}^- = 0$. In particular, this is the case if $\mathcal{B}$ is affine invariant and 1-localizable.*

Proof. We prove (i) first. Let B be an element of $\mathcal{B}$, and let $l(\alpha)$ denote the line segment defined in polar coordinates by (α, ρ), $0 \leq \rho < +\infty$. Both B and $l(\alpha)$ are convex, so their intersection $B \cap l(\alpha)$ is convex. Since B is open, this set has the form $[0, d(\alpha))$, where $d(\alpha)$ is the distance from the origin to the boundary of B in the direction α. The function $\alpha \mapsto d(\alpha)$ is continuous (Exercise 14.2). Since B has area one, this function can be a constant only if $d(\alpha) = 1/\sqrt{\pi}$. In all other cases, $d(\alpha)$ takes values greater than $1/\sqrt{\pi}$ and less than $1/\sqrt{\pi}$, and since it is continuous, it must take the value $1/\sqrt{\pi}$. In fact, since B is symmetric, d must assume the value $1/\sqrt{\pi}$ four times. In any case, there is a point $(x, y) \in \partial B$ such that $x > 0$ and $x^2 + y^2 = 1/\pi$.

Now consider the disk $D = D(0, 1/\sqrt{\pi})$. We have just seen that there is a point $(x, y) \in \partial B \cap \partial D$ such that $x > 0$. If $(x, y) \in D$ and $x > 0$, then $x > x^2$, and we have the following inequalities:

$$\sup_{(\mathbf{x},\mathbf{y})\in B} (x + y^2) \geq \sup_{(\mathbf{x},\mathbf{y})\in B\cap D, x>0} (x + y^2) \geq \sup_{(\mathbf{x},\mathbf{y})\in B\cap D, x>0} (x^2 + y^2) = \frac{1}{\pi}.$$

The right-hand term does not depend on B, so we have $c^+_{\mathcal{B}} \geq 1/\pi$. The value $1/\pi$ is not significant for our purposes; we just wish to show that $c^+_{\mathcal{B}} > 0$.

To prove that $c^-_{\mathcal{B}} = 0$, first note that since B is open and contains the origin, there are points $(x, y) \in B$ with $x > 0$ and $y = 0$. Thus, $c^-_{\mathcal{B}} \geq 0$. Fix $B \in \mathcal{B}$ and consider the sets obtained by "squeezing" B onto the line $x = 0$:

$$B_\varepsilon = \{(x', y') \mid x' = \varepsilon x, y' = y/\varepsilon, (x, y) \in B\}.$$

Then B_ε is an affine transform of B, so $B_\varepsilon \in \mathcal{B}$. Therefore,

$$c^-_{\mathcal{B}} \leq \sup_{(x,y)\in B_\varepsilon} (x - y^2) \leq \sup_{(x,y)\in B_\varepsilon} (x) \leq C\varepsilon.$$

Thus, $c^-_{\mathcal{B}} = 0$.

We turn now to the proof of (ii). Assume that B is in $\mathcal{B}_{\text{aff}}$. Then by definition, B is open, connected, and contains the origin, and the connected components of $B \cap \Delta^c$ that contain the origin in their boundaries have areas greater than or equal to $b = \delta(0, B^c) > 1$. If we let Δ be the y-axis and H be the open half-plane defined by $x > 0$, then the definition implies that $\text{area}(B \cap H) > 1$. This implies that

$$\Big(\sup_{(x,y)\in B\cap H} (x)\Big)\Big(\sup_{(x,y)\in B\cap H} (|y|)\Big) \geq \frac{1}{2}. \tag{18.3}$$

We wish to find a lower bound for $\sup_{(x,y)\in B}(x + y^2)$, so we may assume that B is bounded, and define $\mu = \sup_{(x,y)\in B\cap H}(x)$ and $\nu = \sup_{(x,y)\in B\cap H}(y^2)$. Then

$$\sup_{(x,y)\in B} (x + y^2) \geq \sup_{(x,y)\in B\cap H} (x + y^2) \geq \inf\{\mu, \nu\}.$$

Thus, $c^+_{\mathcal{B}_{\text{aff}}} \geq \inf\{\mu, \nu\}$. Using the constraint (18.3), we conclude that

$$c^+_{\mathcal{B}_{\text{aff}}} \geq \inf_{\mu\geq 0} \left\{\mu, \frac{1}{4\mu^2}\right\} = 2^{-2/3} > 0.$$

Finally, we must show that $c^-_{\mathcal{B}_{\text{aff}}} = 0$. If $B \in \mathcal{B}_{\text{aff}}$ and if B is bounded, then, as in the proof of (i),

$$c^-_{\mathcal{B}_{\text{aff}}} \le \sup_{(x,y)\in B_\varepsilon} (x - y^2) \le \sup_{(x,y)\in B_\varepsilon} (x) \le C\varepsilon,$$

and $c^-_{\mathcal{B}_{\text{aff}}} = 0$. The proof of (iii) is exactly the same as showing that $c^-_{\mathcal{B}_{\text{aff}}} = 0$. □

Exercise 18.2. The purpose of this exercise is to show that the function $\alpha \mapsto d(\alpha)$ in Lemma 14.1 is continuous. (Hint: The assumption that d is not continuous leads to a contradiction of the fact that the line segment $l(\alpha)$ intersects ∂B in one and only one point.) ■

Exercise 18.3. This exercise is to show that it is possible to have $c^-_{\mathcal{B}} < 0$ for a simple set of structuring elements. Let $\mathcal{B} = \{AC \mid A \in SL(\mathbb{R}^2)\}$, where C is a square with one side missing defined as follows: $C = \{(x,y) \mid x = -2, -2 \le y \le 2; y = +2, -2 \le x \le 2; y = -2, -2 \le x \le 2\}$. Show that $c^+_{\mathcal{B}} > 0$ and $c^-_{\mathcal{B}} < 0$. ■

As one can imagine, the polynomial $px + by^2$, with $p > 0$, is particularly important in the affine-invariant theory. Fortunately, an invariance argument allows us to compute explicitly the action of IS_h on $px + by^2$.

Lemma 18.2. *Let $\mathcal{B}$ be an affine-invariant set of structuring elements and assume that at least one $B \in \mathcal{B}$ is bounded. Let IS_h be the associate inf-sup operator and assume that $p > 0$. Then*

$$IS_h(px + by^2) = c^+_{\mathcal{B}} \Big(\frac{b}{p}\Big)^{1/3} p h^{4/3} \quad \text{if} \quad b > 0;$$
$$IS_h(px + by^2) = c^-_{\mathcal{B}} \Big(\frac{-b}{p}\Big)^{1/3} p h^{4/3} \quad \text{if} \quad b \le 0.$$

Proof. The existence of a bounded structuring element ensures that $\sup_{(x,y)\in B}(px + by^2)$ is not always infinite. If $b \ne 0$, then

$$B \in \mathcal{B} \Longleftrightarrow h \begin{pmatrix} h^{1/3}|b|^{1/3} & 0 \\ 0 & h^{-1/3}|b|^{-1/3} \end{pmatrix} B \in h\mathcal{B}.$$

Thus,

$$\begin{aligned}\inf_{B\in h\mathcal{B}} \sup_{(x,y)\in B} (x + by^2) &= \inf_{B\in\mathcal{B}} \sup_{(x,y)\in B} (|b|^{1/3} h^{4/3} x + b(|b|^{-2/3} h^{4/3} y^2)) \\ &= |b|^{1/3} h^{4/3} \inf_{B\in\mathcal{B}} \sup_{(x,y)\in B} (x + (b/|b|) y^2).\end{aligned}$$

Then we have

$$IS_h(x + by^2) = \begin{cases} c^+_{\mathcal{B}}\, b^{1/3} h^{4/3} & \text{if} \quad b > 0; \\ c^-_{\mathcal{B}} (-b)^{1/3} h^{4/3} & \text{if} \quad b < 0. \end{cases}$$

Since $p > 0$, $IS_h(px + by^2) = pIS_h(x + (b/p)y^2)$, and we deduce that

$$IS_h(px + by^2) = \begin{cases} c^+_{\mathcal{B}} (b/p)^{1/3} p h^{4/3} & \text{if} \quad b > 0; \\ c^-_{\mathcal{B}} (-b/p)^{1/3} p h^{4/3} & \text{if} \quad b < 0. \end{cases}$$

Finally, we must deal with the case $b = 0$. Let B be a bounded element of $\mathcal{B}$ and assume it is contained in the square $[-R, R] \times [-R, R]$. Then the element $h\begin{pmatrix} h\varepsilon & 0 \\ 0 & h^{-1}\varepsilon^{-1} \end{pmatrix}B$ belongs to $h\mathcal{B}$ and is contained in the rectangle $\mathcal{R} = [-R\varepsilon h^2, R\varepsilon h^2] \times [-R/\varepsilon, R/\varepsilon]$. Hence,

$$0 \le IS_h(px) \le \sup_{(x,y)\in\mathcal{R}} (px) \le pRh^2\varepsilon.$$

Since we can take $\varepsilon > 0$ arbitrarily small, $IS_h(px) = 0$. $\square$

When we studied the asymptotic behavior of an operator T applied to a smooth function u in Chapters 10 and 11, we usually assumed that $Du(\mathbf{x}) \neq 0$; $Du(\mathbf{x}) = 0$ was a special case. This is also true for affine-invariant inf-sup operators, and the next lemma deals with this case.

Lemma 18.3. *Let $\mathcal{B}$ be an affine-invariant set of structuring elements, one of which is bounded and all of which contain the origin, and let K be a compact subset of $\mathbb{R}^2$. Then for $\mathbf{x} \in K$ the following inequality holds for every C^3 function u:*

$$|IS_h u(\mathbf{x}) - u(\mathbf{x})| \le C(\|D^2u(\mathbf{x})\| + |Du(\mathbf{x})|)h^{4/3} + C_K h^2,$$

where $0 < h \le 1$, $C > 0$ is a constant that depends only on $\mathcal{B}$, and the constant C_K depends only on $\mathcal{B}$, u, and K. If $Du(\mathbf{x}) = 0$, then

$$|IS_h u(\mathbf{x}) - u(\mathbf{x})| \le C'\|D^2u(\mathbf{x})\|h^2 + C'_K h^3,$$

where C' depends only on $\mathcal{B}$ and C'_K depends only on $\mathcal{B}$, u, and K.

Proof. We use the notation of sections 4.5 and 10.1. Let B be an arbitrary element of $\mathcal{B}$. Since B contains the origin, $\sup_{\mathbf{y}\in\mathbf{x}+hB} u(\mathbf{y}) \ge u(\mathbf{x})$, which implies that $IS_h u(\mathbf{x}) \ge u(\mathbf{x})$.

Now expand u in the familiar local coordinate system in a neighborhood of $\mathbf{x}$:

$$u(\mathbf{x} + h\mathbf{y}) = u(\mathbf{x}) + phx + ah^2x^2 + bh^2y^2 + ch^2xy + R(\mathbf{x}, h\mathbf{y}),$$

where $\mathbf{y} = (x, y)$. Then for any $B \in \mathcal{B}$,

$$\sup_{\mathbf{y}\in B} u(\mathbf{x} + h\mathbf{y}) \le u(\mathbf{x}) + h \sup_{\mathbf{y}\in B} (px + ahx^2 + bhy^2 + chxy) + \sup_{\mathbf{y}\in B} |R(\mathbf{x}, h\mathbf{y})|.$$

Now assume that B^* is bounded. Then all of the suprema are finite, and $\sup_{\mathbf{y}\in B^*} u(\mathbf{x} + h\mathbf{y})$ is a finite upper bound for $\inf_{B\in\mathcal{B}} \sup_{\mathbf{y}\in B} u(\mathbf{x} + h\mathbf{y}) = IS_h u(\mathbf{x})$. Thus we have

$$0 \le IS_h u(\mathbf{x}) - u(\mathbf{x}) \le h \sup_{\mathbf{y}\in B^*} (px + ahx^2 + bhy^2 + chxy) + \sup_{\mathbf{y}\in B^*} |R(\mathbf{x}, h\mathbf{y})|, \quad (18.4)$$

which is true for any bounded set $B^* \in \mathcal{B}$. We are now going to use the affine invariance of $\mathcal{B}$ to manipulate B^* and thereby obtain a good estimate for the terms on the right-hand side of (18.4). Since B^* is bounded, it is contained in the square $[-S,S]\times[-S,S]$, where $S=\sup_{\mathbf{y}\in B^*}|\mathbf{y}|$. By the affine invariance of $\mathcal{B}$, the set $B'=\begin{pmatrix} h^{1/3} & 0 \\ 0 & h^{-1/3}\end{pmatrix}B^*$ belongs to $\mathcal{B}$ and is contained in the rectangle

$$\mathcal{R}=[-Sh^{1/3},Sh^{1/3}]\times[-Sh^{-1/3},Sh^{-1/3}].$$

We replace B^* with B' in (18.4) and proceed to estimate the terms on the right-hand side.

$$\begin{aligned}
&\sup_{\mathbf{y}\in B'}(px+ahx^2+bhy^2+chxy)\le\sup_{\mathbf{y}\in\mathcal{R}}(px+ahx^2+bhy^2+chxy)\\
&\le pSh^{1/3}+|a|S^2h^{5/3}+|b|S^2h^{1/3}+|c|S^2h\le pSh^{1/3}+(|a|+|b|+|c|)S^2h^{1/3}\\
&=|Du(\mathbf{x})|Sh^{1/3}+(1/2)\|D^2u(\mathbf{x})\|S^2h^{1/3}\le C(\|D^2u(\mathbf{x})\|+|Du(\mathbf{x})|)h^{1/3},
\end{aligned}$$

where $C=\max\{S,S^2/2\}$. Note that C depends only on $\mathcal{B}$; in particular, it does not depend on $\mathbf{x}$. Note also that this holds for all h, $0<h\le 1$. We now turn to the other term:

$$\begin{aligned}
\sup_{\mathbf{y}\in B'}|R(\mathbf{x},h\mathbf{y})|&\le\sup_{\mathbf{y}\in\mathcal{R}}|R(\mathbf{x},h\mathbf{y})|\le\sup_{\mathbf{y}\in\mathcal{R}}\|D^3u(\mathbf{x}+h\mathbf{y})\|h^3|\mathbf{y}|^3\\
&\le\sup_{\mathbf{y}\in\mathcal{R}}\|D^3u(\mathbf{x}+h\mathbf{y})\|h^3(S^2h^{2/3}+S^2h^{-2/3})^{3/2}\\
&\le\sup_{\mathbf{y}\in\mathcal{R}}\|D^3u(\mathbf{x}+h\mathbf{y})\|2^{3/2}S^3h^2.
\end{aligned}$$

If K is an arbitrary compact set, then $\sup_{\mathbf{x}\in K}\sup_{\mathbf{y}\in\mathcal{R}}\|D^3u(\mathbf{x}+h\mathbf{y})\|2^{3/2}S^3\le C_K$ for some constant that depends only on K, u, and $\mathcal{B}$. This proves that

$$|IS_hu(\mathbf{x})-u(\mathbf{x})|\le C(|Du(\mathbf{x})|+\|D^2u(\mathbf{x})\|)h^{4/3}+C_Kh^2.$$

If $Du(\mathbf{x})=0$, we do the same computation, but we treat both axes the same: $\mathcal{R}$ is replaced with the square $[-Sh,Sh]\times[-Sh,Sh]$, and

$$\sup_{\mathbf{y}\in\mathcal{R}}(ax^2+by^2+cxy)+\sup_{\mathbf{y}\in\mathcal{R}}|R(\mathbf{x},h\mathbf{y})|\le C'\|D^2u(\mathbf{x})\|h^2+C'_Kh^3. \qquad \square$$

$\square$

Since $SI_hu=-IS_h(-u)$, it is clear that the lemma is true for SI_h. Everything is now in place to state and prove the first major result of this chapter.

Theorem 18.4. *Let $\mathcal{B}$ be a 1-localizable affine-invariant family of structuring elements. Assume that $u:\mathbb{R}^2\to\mathbb{R}$ is a Lipschitz function that is C^3 in a neighborhood of $\mathbf{x}$. Then*

$$IS_hu(\mathbf{x})-u(\mathbf{x})=h^{4/3}c_{\mathcal{B}}|Du(\mathbf{x})|\Big(\frac{1}{2}\mathrm{curv}(u)(\mathbf{x})^+\Big)^{1/3}+o(\mathbf{x},h^{4/3}),$$

where $c_{\mathcal{B}}=c_{\mathcal{B}}^+$. If u is C^3 in a neighborhood of a compact set K and $Du(\mathbf{x})\ne 0$ on K, then the result holds for all $\mathbf{x}\in K$ and $o(\mathbf{x},h^{4/3})/h^{4/3}\to 0$ as $h\to 0$ uniformly for $\mathbf{x}\in K$.

has a local maximum at $(t_\varepsilon, \mathbf{x}_\varepsilon)$. This means that Lemma 19.3 applies to ϑ_ε. Since $D\varphi_\varepsilon(t_\varepsilon, \mathbf{y}_\varepsilon) = D\vartheta_\varepsilon(t_\varepsilon, \mathbf{y}_\varepsilon) \neq 0$ and $D^2\varphi(t_\varepsilon, \mathbf{y}_\varepsilon) = D^2\vartheta_\varepsilon(t_\varepsilon, \mathbf{y}_\varepsilon)$, we have

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(t_\varepsilon, \mathbf{y}_\varepsilon) &\leq F(D^2\varphi(t_\varepsilon, \mathbf{y}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{y}_\varepsilon), \mathbf{y}_\varepsilon, t_\varepsilon) \\ &\leq G^+(D^2\varphi(t_\varepsilon, \mathbf{y}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{y}_\varepsilon), \mathbf{y}_\varepsilon, t_\varepsilon),\end{aligned}$$

using (19.5) and (19.2). Again, letting ε tend to zero gives (19.9). □

Exercise 19.3. Prove the inequality (19.11). ■

The next lemma provides another very useful simplification for verifying that a function u is a viscosity solution of (19.4).

Lemma 19.5. *To show that u is a viscosity subsolution (or supersolution), it suffices to use test functions $\varphi \in C^\infty([0,\infty) \times \mathbb{R}^N)$ of the form $\varphi(t, \mathbf{x}) = f(\mathbf{x}) + g(t)$.*

Proof. This is the assumption: If a function of the form $\varphi(t, \mathbf{x}) = f(\mathbf{x}) + g(t)$ is such that $u - \varphi$ attains a local maximum at $(t_0, \mathbf{x}_0)$, then (19.5) and (19.9) follow. From this assumption, we must prove: If φ is any function in $C^\infty([0,\infty) \times \mathbb{R}^N)$ such that $u - \varphi$ attains a local maximum, then (19.5) and (19.9) follow. The technique for doing this is to develop φ as a Taylor series and separate the variables. To keep the notation manageable, we will assume without loss of generality that $(t_0, \mathbf{x}_0) = (0, 0) = 0$. With this assumption, the Taylor expansion of φ is

$$\varphi(t, \mathbf{x}) = a + bt + \langle p, \mathbf{x}\rangle + ct^2 + \langle Q\mathbf{x}, \mathbf{x}\rangle + t\langle q, \mathbf{x}\rangle + o(|\mathbf{x}|^2 + t^2),$$

where $a = \varphi(0)$, $b = \partial\varphi/\partial t(0)$, $c = (1/2)\partial^2\varphi/\partial t^2(0)$, $Q = (1/2)D^2\varphi(0)$, and

$$q = \Big(\frac{\partial^2\varphi}{\partial x_1 \partial t}(0), \dots, \frac{\partial^2\varphi}{\partial x_N \partial t}(0)\Big).$$

For $\varepsilon > 0$, we define

$$f(\mathbf{x}) = a + \langle p, \mathbf{x}\rangle + \langle Q\mathbf{x}, \mathbf{x}\rangle + \varepsilon|\mathbf{x}|^2 + \varepsilon|q||\mathbf{x}|^2$$

and

$$g(t) = bt + \frac{|q|}{\varepsilon}t^2 + \varepsilon t^2 + ct^2.$$

This means that

$$\varphi(t, \mathbf{x}) = f(\mathbf{x}) + g(t) - \Big(\varepsilon|q||\mathbf{x}|^2 + \frac{|q|}{\varepsilon}t^2 - t\langle q, \mathbf{x}\rangle + \varepsilon(|\mathbf{x}|^2 + t^2)\Big) + o(|\mathbf{x}|^2 + t^2).$$

Since, by the Cauchy–Schwartz inequality, $\varepsilon|q||\mathbf{x}|^2 + (|q|/\varepsilon)t^2 - t\langle q, \mathbf{x}\rangle \geq 0$, we have $\varphi(t, \mathbf{x}) \leq f(\mathbf{x}) + g(t)$ for all sufficiently small $(t, \mathbf{x})$. Thus, in some neighborhood of $(0, 0)$, $u(t, \mathbf{x}) - \varphi(t, \mathbf{x}) \geq u(t, \mathbf{x}) - f(\mathbf{x}) - g(t)$ and this inequality is an equality for $(t, \mathbf{x}) = (0, 0)$. The assumption is that $u - \varphi$ has a local maximum at $(0, 0)$. Hence this last inequality implies that $u - f - g$ has a local maximum at $(0, 0)$. Thus, by assumption, (19.5) and (19.9) hold for $f + g$. More precisely, we have the following two cases.

Case (1): $D(f+g)(0) \neq 0$.

From (19.5),

$$\frac{\partial(f+g)}{\partial t}(0) \leq F(D^2(f+g)(0), D(f+g)(0), 0, 0).$$

It is easy to see that

$$D\varphi(0) = D(f+g)(0) \neq 0 \text{ and } (\partial\varphi/\partial t)(0) = (\partial(f+g)/\partial t)(0),$$

and a short computation shows that $D^2(f+g)(0) = D^2\varphi(0) + 2\varepsilon(1+|q|)I)$. Substituting these values in the expression above shows that

$$\frac{\partial\varphi}{\partial t}(0) \leq F(D^2\varphi(0) + 2\varepsilon(1+|q|)I, D\varphi(0), 0, 0).$$

We let $\varepsilon \to 0$ and use the continuity of F to see that (19.5) holds for φ.

Case (2): $D(f+g)(0) = 0$.

In this case, (19.9) is true for $f+g$:

$$\frac{\partial(f+g)}{\partial t}(0) \leq G^+(D^2(f+g)(0), D(f+g)(0), 0, 0).$$

Letting $\varepsilon \to 0$ and using the continuity of G^+ yields (19.9) for φ. □

We are now in position to see how classical and viscosity solutions are related. The next two propositions show that the notion of viscosity solution is indeed a generalization of that of classical solution.

Proposition 19.6. *Let F be an admissible function that is continuous everywhere, and assume $u : [0,\infty) \times \mathbb{R}^N \to \mathbb{R}$ is C^2 with respect to $\mathbf{x}$ and C^1 with respect to t. If u is a classical solution of*

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

at $(t_0, \mathbf{x}_0)$, then u is a viscosity solution at $(t_0, \mathbf{x}_0)$.

Proof. We prove this for the case $Du(t_0, \mathbf{x}_0) \neq 0$. (The other cases follow immediately.) Thus, let $\varphi \in C^\infty([0,\infty) \times \mathbb{R}^N)$ be such that $u - \varphi$ has a local maximum at $(t_0, \mathbf{x}_0)$. This implies that $(\partial u/\partial t, Du)(t_0, \mathbf{x}_0) = (\partial\varphi/\partial t, D\varphi)(t_0, \mathbf{x}_0)$ and that $D^2(u-\varphi)(t_0, \mathbf{x}_0) \leq 0$, so

$$D^2u(t_0, \mathbf{x}_0) \leq D^2\varphi(t_0, \mathbf{x}_0).$$

Hence,

$$\begin{aligned}\frac{\partial\varphi}{\partial t}(t_0, \mathbf{x}_0) = \frac{\partial u}{\partial t}(t_0, \mathbf{x}_0) &= F(D^2u(t_0, \mathbf{x}_0), Du(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0)\\ &\leq F(D^2\varphi(t_0, \mathbf{x}_0), D\varphi(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0).\end{aligned}$$

This proves that u is a viscosity supersolution. A similar argument shows that it is a viscosity subsolution. □

Proposition 19.7. *Assume that F is admissible and continuous everywhere. Let u be a $C^2(\mathbb{R}^+ \times \mathbb{R}^N)$ viscosity solution of $\partial u/\partial t = F(D^2u, Du, \mathbf{x}, t)$. Then u is a classical solution of the same equation.*

Proof. Assume that u is a viscosity solution at the point $(t_0, \mathbf{x}_0)$. We write the second-order Taylor expansion of u in the $N+1$ variables near $(t_0, \mathbf{x}_0)$ as

$$\begin{aligned} u(t, \mathbf{x}) = {} & u(t_0, \mathbf{x}_0) + \langle \tilde{D}u(t_0, \mathbf{x}_0), (t - t_0, \mathbf{x} - \mathbf{x}_0)\rangle \\ & + \langle \tilde{D}^2 u(t_0, \mathbf{x}_0)(t - t_0, \mathbf{x} - \mathbf{x}_0), (t - t_0, \mathbf{x} - \mathbf{x}_0)\rangle + o(|t - t_0|^2 + |\mathbf{x} - \mathbf{x}_0|^2), \end{aligned}$$

where the operators $\tilde{D}$ and $\tilde{D}^2$ involve all $N+1$ variables. For $\varepsilon > 0$, define φ_ε by

$$\begin{aligned} \varphi_\varepsilon(t, \mathbf{x}) = u(t_0, \mathbf{x}_0) & + \langle \tilde{D}u(t_0, \mathbf{x}_0), (t - t_0, \mathbf{x} - \mathbf{x}_0)\rangle \\ & + \langle (\tilde{D}^2 u(t_0, \mathbf{x}_0) + \varepsilon I)(t - t_0, \mathbf{x} - \mathbf{x}_0), (t - t_0, \mathbf{x} - \mathbf{x}_0)\rangle. \end{aligned}$$

Thus,

$$u(t, \mathbf{x}) - \varphi_\varepsilon(t, \mathbf{x}) = -\varepsilon(|t - t_0|^2 + |\mathbf{x} - \mathbf{x}_0|^2) + o(|t - t_0|^2 + |\mathbf{x} - \mathbf{x}_0|^2)$$

and the point $(t_0, \mathbf{x}_0)$ is a local maximum of $u - \varphi_\varepsilon$ for all $\varepsilon > 0$. Similarly, $(t_0, \mathbf{x}_0)$ is a local minimum of $u - \varphi_{-\varepsilon}$. The test functions φ_ε and $\varphi_{-\varepsilon}$ are C^∞, so we can apply Definition 15.2 directly. Thus,

$$\begin{aligned} \frac{\partial u}{\partial t}(t_0, \mathbf{x}_0) = \frac{\partial \varphi_\varepsilon}{\partial t}(t_0, \mathbf{x}_0) & \le F(D^2\varphi_\varepsilon(t_0, \mathbf{x}_0), D\varphi_\varepsilon(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0) \\ & = F((D^2u + \varepsilon I)(t_0, \mathbf{x}_0), Du(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial u}{\partial t}(t_0, \mathbf{x}_0) = \frac{\partial \varphi_{-\varepsilon}}{\partial t}(t_0, \mathbf{x}_0) & \ge F(D^2\varphi_{-\varepsilon}(t_0, \mathbf{x}_0), D\varphi_\varepsilon(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0) \\ & = F((D^2u - \varepsilon I)(t_0, \mathbf{x}_0), Du(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0). \end{aligned}$$

Letting $\varepsilon \to 0$ and using the continuity of F shows that

$$\frac{\partial u}{\partial t}(t_0, \mathbf{x}_0) = F(D^2u(t_0, \mathbf{x}_0), Du(t_0, \mathbf{x}_0), \mathbf{x}_0, t_0). \qquad \square$$

$\square$

Before discussing several examples, we need a useful further restriction on the test functions φ.

Lemma 19.8. *To show that u is a viscosity subsolution (or supersolution), it suffices to use test functions $\varphi \in C^\infty([0, \infty) \times \mathbb{R}^N)$ that satisfy $\mathbf{x} \to \varphi(t, \mathbf{x}) \in \mathcal{F}(\mathbb{R}^N)$ for all $t \ge 0$ and are globally Lipschitz on $\mathbb{R}^N$.*

Proof. In fact the properties we deal with are all local around a point $(t_0, \mathbf{x}_0)$. Thus we can replace φ by another C^∞ function ψ which coincides with φ on a ball $B(0, (t_0, \mathbf{x}_0))$, belongs to $\mathcal{F}$ for all t, and is globally Lipschitz on $\mathbb{R}^N$. $\square$

Exercise 19.4. Give a detailed construction of ψ from φ. ■

19.2 Application to mathematical morphology

In Proposition 9.6, we showed that $u(t,\mathbf{x}) = D_t u_0(\mathbf{x})$ is a solution of $\partial u/\partial t = |Du|$ at each point $(t_0, \mathbf{x}_0)$ where u is C^1. We are now going to prove that u is a viscosity solution of this equation at all points.

Theorem 19.9. *Assume that $u_0 \in \mathcal{F}(\mathbb{R}^N)$. Let $D(0,1)$ be the unit ball in $\mathbb{R}^N$ and let u be defined by*

$$u(t,\mathbf{x}) = D_t u_0(\mathbf{x}) = \sup_{\mathbf{y}\in tD(0,1)} u_0(\mathbf{x}+\mathbf{y}).$$

Then u is a viscosity solution of

$$\frac{\partial u}{\partial t} = |Du|, \quad u(0,\mathbf{x}) = u_0(\mathbf{x}). \tag{19.12}$$

Proof. We will use the fact that the dilation D_t is recursive, that is, $D_{s+t} = D_s D_t$ (see Proposition 9.5). In particular, for $t > 0$, $D_t = D_h D_{t-h}$, so

$$u(t,\mathbf{x}) = \sup_{|\mathbf{y}|<h} u(t-h, \mathbf{x}+\mathbf{y}). \tag{19.13}$$

Let φ be a C^∞ test function and assume that $u - \varphi$ has a local maximum at $(t_0, \mathbf{x}_0)$. To prove that u is a viscosity subsolution of (19.12), we must show that

$$\frac{\partial \varphi}{\partial t}(t_0,\mathbf{x}_0) - |D\varphi(t_0,\mathbf{x}_0)| \le 0.$$

Since $u - \varphi$ has a local maximum at $(t_0, \mathbf{x}_0)$, we have for sufficiently small h and $|\mathbf{y}|$,

$$u(t_0-h,\mathbf{x}_0+\mathbf{y}) - \varphi(t_0-h,\mathbf{x}_0+\mathbf{y}) \le u(t_0,\mathbf{x}_0) - \varphi(t_0,\mathbf{x}_0)$$

It follows that

$$\sup_{|\mathbf{y}|<h} u(t_0-h,\mathbf{x}_0+\mathbf{y}) \le u(t_0,\mathbf{x}_0) - \varphi(t_0,\mathbf{x}_0) + \sup_{|\mathbf{y}|<h} \varphi(t_0-h,\mathbf{x}_0+\mathbf{y}),$$

and using (19.13) shows that

$$u(t_0,\mathbf{x}_0) \le u(t_0,\mathbf{x}_0) - \varphi(t_0,\mathbf{x}_0) + \sup_{|\mathbf{y}|<h} \varphi(t_0-h,\mathbf{x}_0+\mathbf{y}).$$

Thus,

$$\varphi(t_0,\mathbf{x}_0) \le \sup_{|\mathbf{y}|<h} \varphi(t_0-h,\mathbf{x}_0+\mathbf{y}).$$

Subtracting $\varphi(t_0-h,\mathbf{x}_0)$ from both sides yields

$$\varphi(t_0,\mathbf{x}_0) - \varphi(t_0-h,\mathbf{x}_0) \le \sup_{|\mathbf{y}|<h} \big(\varphi(t_0-h,\mathbf{x}_0+\mathbf{y}) - \varphi(t_0-h,\mathbf{x}_0)\big).$$

By writing $\varphi(t_0-h,\mathbf{x}_0+\mathbf{y}) - \varphi(t_0-h,\mathbf{x}_0) = \langle \tilde{D}\varphi(t_0,\mathbf{x}_0), (0,\mathbf{y})\rangle + o(h+|\mathbf{y}|)$, we see that

$$\varphi(t_0,\mathbf{x}_0) - \varphi(t_0-h,\mathbf{x}_0) \le |D\varphi(t_0,\mathbf{x}_0)|h + o(h).$$

Dividing both sides by h and letting h tend to zero leads to

$$\frac{\partial \varphi}{\partial t}(t_0, \mathbf{x}_0) - |D\varphi(t_0, \mathbf{x}_0)| \leq 0.$$

We have proven this under the assumption that $t_0 > 0$ and $h > 0$ is sufficiently small. But by continuity, the last inequality is true for $t_0 = 0$. This proves that u is a subsolution of (19.12); the proof that it is a supersolution is similar. The fact that $u(0, \mathbf{x}) = u_0(\mathbf{x})$ is a direct consequence of the assumption that u_0 is continuous. □

19.3 Approximation theory of viscosity solutions

For simplicity, we consider slightly less general PDEs, namely, those of the form

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) - F(D^2u(t, \mathbf{x}), Du(t, \mathbf{x}), \mathbf{x}) = 0, \tag{19.14}$$

where $(A, p, \mathbf{x}) \mapsto F(A, p, \mathbf{x})$ is admissible, but independent of t. This is the case for the functions listed at the beginning of the chapter. For these equations, it is reasonable to expect that the operator $S_t : u_0 \mapsto u(t, \cdot)$ could be approximated by iterations of an operator T_h, by which we mean that $(T_h)^n \to S_t$ in some sense as $h \to 0$, $n \to \infty$. We have seen this in Theorem 2.3, where it was shown that a large class of iterated linear operators converge asymptotically to the heat equation. We have promised to show that whole classes of nonlinear operators converge asymptotically to other PDEs. We are mainly interested in operators that have been shown to be useful for image analysis and that have been studied in previous chapters. To include all of these operators in the theory, we shall first state three abstract properties which were proven under various forms for scaled morphological operators T_h.

Definition 19.10. *We say that a family of operators T_h, $h > 0$ is* uniformly consistent with Equation (19.14) *if for every C^3, Lipschitz function u we can assert that*

$$\text{if } Du(\mathbf{x}) \neq 0, (T_hu)(\mathbf{x}) - u(\mathbf{x}) = hF(D^2u, Du, \mathbf{x}) + o_{\mathbf{x}}(h), \tag{19.15}$$

where the convergence of $o_{\mathbf{x}}(h)$ is uniform for $\mathbf{x}$ in every compact set contained in the set $\{\mathbf{x}, Du(\mathbf{x}) \neq 0\}$ and

$$\text{if } Du(\mathbf{x}) = 0, |(T_hu)(\mathbf{x}) - u(\mathbf{x})| \leq hG(D^2u, 0, \mathbf{x}) + o_{\mathbf{x}}(h) \tag{19.16}$$

for a continuous functions G, with $G(0, 0, \mathbf{x}) = 0$, and where the convergence of $o_{\mathbf{x}}(h)$ is uniform for $\mathbf{x}$ in every compact set.

Definition 19.11. *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 $u(\mathbf{y}) \geq v(\mathbf{y})$ on a disk $D(\mathbf{x}, r)$,*

$$(T_hu)(\mathbf{x}) \geq (T_hv)(\mathbf{x}) - o(h), \tag{19.17}$$

where the function $o(h)$ only depends upon the Lipschitz constant L and r.

Notice that if T is a local morphological operator like the median on a ball and T_h its rescaled version, then (19.17) is trivially satisfied, with $o(h) = 0$ for h small enough.

Definition 19.12. *Let T_h, $h > 0$, be a family of operators: $\mathcal{F} \to \mathcal{F}$ which is uniformly consistent with Equation (19.14). We call* **approximate solutions** *of (19.14) with initial condition $u_0(\mathbf{x})$ the functions $u_h(t, \mathbf{x})$ defined for every $h > 0$ by*

$$\forall n \in I\!N, \quad u_h(nh, \mathbf{x}) = (T_h^n u_0)(\mathbf{x}).$$

The functions u_h are only defined on $(h\mathbb{N}) \times \mathbb{R}^N$. All the same, we are interested in their limit on $\mathbb{R}^+ \times \mathbb{R}^N = [0, +\infty) \times \mathbb{R}^N$.

Definition 19.13. *We say that approximate solutions u_h converge uniformly on compacts sets to a function u defined on $\mathbb{R}^+ \times \mathbb{R}^N$ if for every compact subset K of $\mathbb{R}^+ \times \mathbb{R}^N$ and every $\varepsilon > 0$, there is h_0 such that $|u(t, \mathbf{x}) - u_h(t, \mathbf{x})| < \varepsilon$ for all $h \le h_0$ and all $(t, \mathbf{x}) \in K \cap (h\mathbb{N}) \times \mathbb{R}^N$.*

Proposition 19.14 (Barles-Souganidis). *Let $(T_h)_{h \ge 0}$ be a family of translation invariant operators uniformly consistent with (19.14), satisfying a uniform comparison principle and commuting with the addition of constants. Let $u_0 \in \mathcal{F}$ be Lipschitz. Assume that a sequence of approximate solutions u_{h_k} converges uniformly on every compact set to a function u, with $h_k \to 0$. Then u is a viscosity solution of (19.14).*

Before starting with the proof, let us state two obvious but useful lemmas.

Lemma 19.15. *Consider u_h converging to u uniformly on compact sets, as in Definition 19.13. Assume that u is continuous on a ball $B_r = B((t, \mathbf{x}), r)$ and that it attains its strict maximum on B_r at $(t, \mathbf{x})$. Then if $(t_h, \mathbf{x}_h)$ is a maximum point of u_h on B_r, one has $(t_h, \mathbf{x}_h) \to (t, \mathbf{x})$.*

Proof. For every $\varepsilon > 0$, there is $\eta > 0$ such that $\sup_{B_r \setminus B_\varepsilon} u < \sup_{B_r} u - \eta$. Take h small enough so that $\sup_{B_r} |u - u_h| < \frac{\eta}{2}$. Then $\sup_{B_r} |u_h| > \sup_{B_r} u - \frac{\eta}{2}$. On the other hand, $\sup_{B_r \setminus B_\varepsilon} |u_h| < \sup_{B_r} u - \eta + \frac{\eta}{2}$, which proves that the maximum of u_h is attained on B_ε only.

□

Lemma 19.16. *Assume that u_0 is L-Lipschitz. Then for every n, $u_h(nh, \mathbf{x})$ is L-Lipschitz in $\mathbf{x}$.*

Proof. This is a straightforward consequence of the definition of u_h and Lemma 7.11. □

Proof of Proposition 19.14. Without risk of ambiguity, we shall write u_h instead of u_{h_k}. Let $B = B((\mathbf{x}, t), r)$ be a closed ball and $\varphi(s, \mathbf{y})$ a C^∞ and Lipschitz function such that $(u - \varphi)(s, \mathbf{y})$ attains its strict maximum on B at $(t, \mathbf{x})$. Without loss of generality, we can assume by Lemma 19.5 that $\varphi(t, \mathbf{y}) = f(\mathbf{y}) + g(t)$. Notice that the functions $\mathbf{x} \to u_n(nh, \mathbf{x})$ are in $\mathcal{F}$ and therefore continuous. The maximum of $u_h - \varphi$ on $B \cap (h\mathbb{N}) \times \mathbb{R}^N$ is attained, because this function is discrete in time and continuous in $\mathbf{x}$. Since $u_h - \varphi \to u - \varphi$ uniformly on $B \cap (h\mathbb{N}) \times \mathbb{R}^N$, we know by Lemma 19.15 that a sequence $(n_h h, \mathbf{x}_h)$ of maxima of $u_h - \varphi$ on B converges to $(t, \mathbf{x})$. By the maximum property of $(n_h h, \mathbf{x}_h)$, we have

$$u_h((n_h - 1)h, \mathbf{y}) - \varphi((n_h - 1)h, \mathbf{y}) \leq u_h(n_h h, \mathbf{x}_h) - \varphi(n_h h, \mathbf{x}_h).$$

for every $\mathbf{y}$ such that $((n_h - 1)h, \mathbf{y}) \in B$ and therefore

$$u_h((n_h - 1)h, \mathbf{y}) \leq u_h(n_h h, \mathbf{x}_h) - \varphi(n_h h, \mathbf{x}_h) + \varphi((n_h - 1)h, \mathbf{y})$$

for h small enough (i.e. k large enough) and every $\mathbf{y} \in B(\mathbf{x}, \frac{r}{2})$. 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 $T_h(u((n_h - 1)h), .)(\mathbf{x}) = u_h(n_h h, \mathbf{x})$, 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 (19.15), since for h small enough $Df(\mathbf{x}_h) \neq 0$,

$$(T_h f)(\mathbf{x}_h) = f(\mathbf{x}_h) + hF(D^2 f(\mathbf{x}_h), Df(\mathbf{x}_h), \mathbf{x}_h) + o_{\mathbf{x}_h}(h).$$

Thus

$$g(n_h h) - g((n_h - 1)h) \leq hF(D^2 f(\mathbf{x}_h), Df(\mathbf{x}_h), \mathbf{x}_h) + o_{\mathbf{x}_h}(h).$$

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 (19.14) and we prove in exactly the same way that it is a supersolution and therefore a viscosity solution. □

At this point it should be clear that the last important step in our program is to show that the approximate solutions converge uniformly on compact sets. It should also be clear that Definitions 15.9 and 15.10 were fashioned to abstract from previous results about inf-sup operators, conditions that are sufficient to prove the Barles–Souganidis theorem. We will see in the next chapter that these conditions are also sufficient to prove that the approximate solutions converge. This then will close the gap and show that the iterated operators converge to viscosity solutions of their associated equations.

19.4 A uniqueness result for viscosity solutions

Proving uniqueness is technically quite difficult, and we are going to fudge by quoting a uniqueness result without proof. Our statement has been simplified to cover only those admissible functions F that are associated with image operators. References where one can find this and more general results are given in the next section.

Theorem 19.17 (Uniqueness). *Assume that $(A,p) \mapsto F(A,p)$ is admissible. Let $u(t,\mathbf{x})$ and $v(t,\mathbf{x})$ be two continuous functions for $(t,\mathbf{x}) \in \mathbb{R}^+ \times S_N$, such that for all $t \in \mathbb{R}^+$, $\mathbf{x} \to u(t,\mathbf{x})$ and $\mathbf{x} \to v(t,\mathbf{x})$ belong to $\mathcal{F}$. If u and v are continuous viscosity solutions of*

$$\frac{\partial \varphi}{\partial t} = F(D^2u, Du) \tag{19.18}$$

then

$$\sup_{t\in R^+,\mathbf{x}\in\mathbb{R}^N} (u(t,\mathbf{x}) - v(t,\mathbf{x})) \le \sup_{\mathbf{x}\in\mathbb{R}^N} (u(0,\mathbf{x}) - v(0,\mathbf{x})). \tag{19.19}$$

As a consequence, if $u(0,\mathbf{x}) = v(0,\mathbf{x})$ for all $\mathbf{x}$, then $u(t,\mathbf{x}) = v(t,\mathbf{x})$ for all $\mathbf{x}$ and t.

19.5 Exercises

Exercise 19.5. The exercise refers to the proof of Lemma 19.4. Consider $u - \varphi$, continuous and having a strict maximum at $(t_0, \mathbf{x}_0)$. Set

$$\psi_\varepsilon(t,\mathbf{x},\mathbf{y}) = u(t,\mathbf{x}) - \varphi(t,\mathbf{y}) - \frac{|\mathbf{x}-\mathbf{y}|^4}{\varepsilon}, \quad \varepsilon > 0.$$

Prove the existence of the points $(t_\varepsilon, \mathbf{x}_\varepsilon, \mathbf{y}_\varepsilon)$ where ψ_ε has local maxima and that tend to $(t_0, \mathbf{x}_0, \mathbf{y}_0)$ as $\varepsilon \to 0$. Hints: Since $u - \varphi$ has a local maximum at $(t_0, \mathbf{x}_0)$, we can choose an r such that $u(t,\mathbf{x}) - \varphi(t,\mathbf{x}) \le u(t_0,\mathbf{x}_0) - \varphi(t_0,\mathbf{x}_0) = \lambda$ for $(t,\mathbf{x}) \in \overline{D}((t_0,\mathbf{x}_0), r) = \overline{D}$. For any $(t,\mathbf{x}), (t,\mathbf{y}) \in \overline{D}$ we have

$$u(t,\mathbf{x}) - \varphi(t,\mathbf{y}) - \frac{|\mathbf{x}-\mathbf{y}|^4}{\varepsilon} \le \sup_{(t,\mathbf{x}),(t,\mathbf{y})\in\overline{D}} \Big(u(t,\mathbf{x}) - \varphi(t,\mathbf{y}) - \frac{|\mathbf{x}-\mathbf{y}|^4}{\varepsilon}\Big).$$

Let $\mathbf{x} = \mathbf{y} = \mathbf{x}_0$ on the left-hand side, so

$$\lambda \leq \sup_{(t,\mathbf{x}),(t,\mathbf{y})\in\overline{D}} \Big(u(t,\mathbf{x}) - \varphi(t,\mathbf{y}) - \frac{|\mathbf{x}-\mathbf{y}|^4}{\varepsilon} \Big).$$

For each $\varepsilon > 0$, there is some point $(t_\varepsilon, \mathbf{x}_\varepsilon, \mathbf{y}_\varepsilon)$ where the supremum is attained. Now argue that these points must tend to $(t_0, \mathbf{x}_0, \mathbf{y}_0)$ as $\varepsilon \to 0$. (This is where the fact that $(t_0, \mathbf{x}_0, \mathbf{y}_0)$ is "strict" is used.) Hence, for small enough ε the points $(t_\varepsilon, \mathbf{x}_\varepsilon, \mathbf{y}_\varepsilon)$ must all be in the interior of $\overline{D}$ and are therefore *local* maximum points. ■

Exercise 19.6. We have seen in Lemma 19.5 that the test functions in Definition 19.2 can be replaced with functions of the form $\varphi(t,\mathbf{x}) = f(\mathbf{x}) + g(t)$. Prove that the requirement can be weakened further by only requiring that f belongs to any class $\mathcal{C}$ of C^2 functions that has the following property: For any $\mathbf{x} \in \mathbb{R}^N$, any $a \in \mathbb{R}$, any $p \in \mathbb{R}^N$, and any symmetric $N \times N$ matrix A, there exists $f \in \mathcal{C}$ such that

$$f(\mathbf{x}-\mathbf{y}) = a + \langle p, \mathbf{x}-\mathbf{y}\rangle + \frac{1}{2}\langle A(\mathbf{x}-\mathbf{y}), (\mathbf{x}-\mathbf{y})\rangle + o(|\mathbf{x}-\mathbf{y}|^2).$$

Hint: use the techniques used to prove Lemma 19.5. ■

19.6 Comments and references

The simple definition of viscosity solution given in this chapter was originally proposed by Michael G. Crandall and Pierre-Louis Lions [**?**] for first-order PDEs associated with control theory. It was then shown to be applicable to second-order equations, in particular the so-called geometric equations like mean curvature motion. The first complete treatise is *User's guide to viscosity solutions of second order partial differential equations* by Crandall, Ishii, and Lions [82]. First-order equations are treated extensively in Barles [33]. An elementary account for first-order equations is given in the textbook by Evans [98]. Crandall's later presentation of the theory for both first- and second-order equations, published in [**?**], is a masterpiece of simplicity and brevity. This book contains a rather complete overview of the techniques, results, and applications, although it does not include applications to image analysis.

The approximation theory for viscosity solutions presented here is based on the seminal paper by Barles and Souganidis [37].

Proving uniqueness for viscosity solutions of second-order parabolic or elliptic equations is the technically difficult part of the theory. The key step that leads to the uniqueness results was made by Robert Jensen in his 1988 paper [160]. See also [187] and [294] for general uniqueness proofs in the parabolic case. Some alternative (or related) existence and uniqueness theories, namely, the nonlinear semigroup theory and De Giorgi's theory of barriers, are discussed in [95] and [39].

19.7 Tables of multiscale differential operators and equations

- All equations have a unique viscosity solution starting from a Lipschitz initial image u_0;
- all iterated radial convolutions converge to the heat equation;
- iterated monotone contrast invariant isotropic filters converge to a curvature motion or an erosion or a dilation;
- iterated contrast invariant affine invariant self-dual filters converge to an affine curvature motion;
- iterated medians converge to a mean curvature or curvature (dim. 2) motion.

operator	$\frac{\partial u}{\partial t} =$	$F(A,p)$
Laplacian	Δu	$trace(A)$
gradient	$\|Du\|$	$\|p\|$
curvature	$\|Du\|curv(u)$	$A(\frac{p^\perp}{\|p\|}, \frac{p^\perp}{\|p\|})$
affine curvature	$\|Du\|curv(u)^{\frac{1}{3}}$	$A(p^\perp, p^\perp)^{\frac{1}{3}}$
snake	$g.\|Du\|curv(u) + (Dg.Du)$	$gA(\frac{p^\perp}{\|p\|}, \frac{p^\perp}{\|p\|}) + Dg.p$
mean curvature	$\|Du\|(\kappa_1(u) + \kappa_2(u))$	$\mathrm{Tr}(A) - A(p/\|p\|, p/\|p\|)$
affine curvature	$\mathrm{sgn}(\kappa_1)t^{1/2}\|Du\|\|G(u)^+\|^{1/4}$	
acceleration	$\|Du\|curv(u)^{\frac{1-q}{3}}(\mathrm{sgn}(curv(u))accel(u)^q)^+$	

operator	*finite difference scheme*	*structuring elements* $\mathcal{B}$
Laplacian	$u_{xx} + u_{yy}$	gaussian convolution
gradient	$(u_x^2 + u_y^2)^{\frac{1}{2}}$	ball
curvature	$\frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{u_x^2+u_y^2}$	median
affine curv.	$(u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2)^{\frac{1}{3}}$	affine inv.
snake	$g.\frac{u_{xx}u_y^2 - 2u_{xy}u_xu_y + u_{yy}u_x^2}{u_x^2+u_y^2} + (g_xu_x + g_yu_y)$	$\mathbf{x}$-dependent median, dilation
mean curv.	$div\left(\frac{Du}{\|Du\|}\right)$	median (BMO)
affine curv.	$\mathrm{sgn}(\kappa_1)t^{1/2}\|Du\|\|G(u)^+\|^{1/4}$	affine inv.
acceleration	$\|Du\|curv(u)^{\frac{1-q}{3}}(\mathrm{sgn}(curv(u))accel(u)^q)^+$	galilean invariant

- $k_i(A,p) = \frac{\mu_i}{|p|}$ where μ_i is the i-th eigenvalue of Q_pAQ_p of A to $p^\perp$, with $Q_p = p \otimes p$. In other terms μ_i is the i-th eigenvalue of the restriction of A to $p^\perp$, the hyperplane orthogonal to p.
- $k_i(u) = k_i(D^2u, Du)$.
- $g(\mathbf{x}) = \frac{1}{1+|Du_0(\mathbf{x})|}$ is small on edges of u_0 and large otherwise;
- affine invariant structuring elements (dim. 2) computed by Moisan scheme;
- curvature motion implemented by BMO (iterated median) or finite difference scheme;
- alternatively mean curvature and mean curvature motion computed by *diffusion (heat equation) on the hyperplane orthogonal to the gradient*;
- $accel(u)$ is a bit long to write but the galilean invariant set of structuring elements leads to an easy inf sup computation; $q \in]0,1[$.

Chapter 20

Curvature Equations and Iterated Contrast-Invariant Operators

In this chapter, we apply the viscosity solution theory to the main curvature equations. A first important consistency result is that the viscosity solutions are invariant under contrast changes (Proposition 20.2): If u is a viscosity solution of a curvature equation, then, for any continuous contrast change g, $g(u)$ is also a viscosity solution of the same equation. Our second main focus is to illustrate the general principle that "iterated contrast invariant filters are asymptotically equivalent to a curvature equation." We shall not prove this principle in whole generality. We shall limit ourselves to two cases which were proven of great interest in image analysis. The first example is the iterated median filter, which will be showed to converge to a curvature equation. The second example is the iteration of alternate affine filters, which converges to the AMSS equation.

20.1 The main curvature equations used for image processing

The curvature equation of most interest to image processing have the general form

$$\frac{\partial u}{\partial t} = |Du|\beta(\mathrm{curv}(u)) \tag{20.1}$$

in two dimensions and the form

$$\frac{\partial u}{\partial t} = |Du|\beta(\kappa_1(u), \kappa_2(u), \dots, \kappa_{N-1}(u)) \tag{20.2}$$

in N dimensions. The real-valued function β is continuous and nondecreasing with respect to each of its variables. The $\kappa_i(u)$ denote the principal curvatures of the level surface of u, defined as the eigenvalues of the restriction of $\frac{D^2u}{|Du|}$ to the hyperplane orthogonal to Du (see Definition 11.19.)

Here are some specific examples. In two dimensions, we have shown that the equations

$$\frac{\partial u}{\partial t} = |Du|\text{curv}(u) \tag{20.3}$$

and

$$\frac{\partial u}{\partial t} = |Du|(\text{curv}(u))^{1/3}, \tag{20.4}$$

as well as variants like

$$\frac{\partial u}{\partial t} = |Du|(\text{curv}(u)^+)^{1/3}, \tag{20.5}$$

are relevant for image processing. In three dimensions, we will be concerned with

$$\frac{\partial u}{\partial t} = |Du|(\kappa_1(u) + \kappa_2(u)). \tag{20.6}$$

This is the classical *mean curvature motion* that is important because it appears as a limit of iterated median filters. Our assumptions also cover variants like

$$\frac{\partial u}{\partial t} = |Du| \min(\kappa_1(u), \kappa_2(u)). \tag{20.7}$$

This filter provides a less destructive smoothing of three-dimensional images than the mean curvature motion. Finally, let us mention affine-invariant curvature motion, which is a particularly important equation in three dimensions:

$$\frac{\partial u}{\partial t} = \text{sgn}(\kappa_1)t^{1/2}|Du||G(u)^+|^{1/4}. \tag{20.8}$$

The admissible functions F for these equations were listed in Section 19.1 and in the synoptic tables of Section 19.7.

20.2 Contrast invariance and viscosity solutions

We are going to show that the concepts of contrast invariance and viscosity solution are compatible. Proposition 20.2 will show that f u is a viscosity solution of (20.1) or (20.2), then for all continuous nondecreasing functions g, $g(u)$ is also a viscosity solution of the same equation.

Lemma 20.1. *Assume that $(A,p) \mapsto F(A,p)$ is an admissible function of the form $F(D^2(u), D(u)) = |D(u)|\beta(\kappa_1(u), \kappa_2(u), \dots, \kappa_{N-1}(u))$ and that $g : \mathbb{R} \to \mathbb{R}$ is C^2 with $g'(s) > 0$ for all $s \in \mathbb{R}$.*

If $Du \neq 0$, then

$$F(D^2(g(u)), D(g(u))) = g'(u)F(D^2u, Du) \tag{20.9}$$

for any C^2 function $u : \mathbb{R}^N \to \mathbb{R}$.

If $D^2u = 0$ and $Du = 0$, then

$$F(D^2(g(u)), D(g(u))) = F(0,0) = 0. \tag{20.10}$$

Proof. If $D(u) \neq 0$, then we know from Proposition 11.16 that $\operatorname{curv}(g(u)) = \operatorname{curv}(u)$ in two dimensions and that $\kappa_i(g(u)) = \kappa_i(u)$ in the N-dimensional case. Thus,

$$\begin{aligned} F(D^2(g(u)), D(g(u))) =& |D(g(u))|\beta(\kappa_1(g(u)), \kappa_2(g(u)), \dots, \kappa_{N-1}(g(u))) \\ =& g'(u)|D(u)|\beta(\kappa_1(u), \kappa_2(u), \dots, \kappa_{N-1}(u)) \\ =& g'(u)F(D^2u, Du), \end{aligned}$$

as announced. In general, $D(g(u)) = g'(u)Du$ and

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

Thus, if $D^2u = 0$ and $Du = 0$, then $D^2(g(u)) = 0$ and $D(g(u)) = 0$, and

$$F(D^2(g(u)), D(g(u))) = F(0, 0) = 0.$$

□

Exercise 20.1. Check the formula $D^2(g(u)) = g'(u)D^2u + g''(u)Du \otimes Du$. ■

The proof of the next result, the main one of this section, is slightly more involved because we drop the assumption that $g'(s) > 0$.

Proposition 20.2. *Assume that u is a viscosity solution of the equation*

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

where F satisfies the conditions of Lemma 20.1. If $g : \mathbb{R} \to \mathbb{R}$ is continuous and nondecreasing, then $g(u)$ is also a viscosity solution of this equation.

Proof. We begin by assuming that g is C^∞ and that $g'(s) > 0$, and we write $f = g^{-1}$ for convenience. Let $(t, \mathbf{x})$ be a strict local maximum of $g(u) - \varphi$. Without loss of generality, we can assume that $g(u(t, \mathbf{x})) - \varphi(t, \mathbf{x}) = 0$: Just replace φ with $\varphi - g(u(t, \mathbf{x}))$. Then $(t, \mathbf{x})$ is also a strict local maximum of $u - f(\varphi)$. To see this, note that

$$g(u(s, \mathbf{y})) - \varphi(s, \mathbf{y}) < g(u(t, \mathbf{x})) - \varphi(t, \mathbf{x}) = 0$$

for $(s, \mathbf{y})$ sufficiently close to (but not equal to) $(t, \mathbf{x})$. Thus,

$$u(s, \mathbf{y}) < f(\varphi(s, \mathbf{y}))$$

and

$$u(s, \mathbf{y}) - f(\varphi(s, \mathbf{y})) < 0 = u(t, \mathbf{x}) - f(\varphi(t, \mathbf{x})),$$

again, for $(s, \mathbf{y})$ sufficiently close to (but not equal to) $(t, \mathbf{x})$.

Since $f(\varphi)$ is C^∞ and u is a viscosity solution, it follows from the definition of viscosity solution that, for $D(f(\varphi)(t, \mathbf{x}) \neq 0$,

$$\frac{\partial(f(\varphi))}{\partial t}(t, \mathbf{x}) \leq F(D^2(f(\varphi))(t, \mathbf{x}), D(f(\varphi))(t, \mathbf{x})).$$

This implies by Lemma 20.1 that

$$f'(\varphi)\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq f'(\varphi)F(D^2(\varphi)(t, \mathbf{x}), D(\varphi)(t, \mathbf{x}));$$

since $f'(s) > 0$,

$$\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq F(D^2(\varphi)(t, \mathbf{x}), D(\varphi)(t, \mathbf{x})).$$

If $D(f(\varphi)) = 0$ and $D^2(f(\varphi)) = 0$, then, by Definition 19.2, $(\partial f(\varphi)/\partial t)(t, \mathbf{x}) \leq 0$, and so $(\partial \varphi/\partial t)(t, \mathbf{x}) \leq 0$. This proves that $g(u)$ is a viscosity subsolution when $g \in C^\infty$ and $g'(s) > 0$.

Now assume that g is simply continuous and nondecreasing. We replace g with g_ε, $\varepsilon > 0$, a C^∞ function such that $g'_\varepsilon(s) \geq \varepsilon$ and $g_\varepsilon \to g$ uniformly on compact subsets of $\mathbb{R}$ as $\varepsilon \to 0$. (See Exercise 20.5.)

We know from Lemma 19.15 that there is a sequence of points $(t_{\varepsilon(k)}, \mathbf{x}_{\varepsilon(k)})$, $k \in \mathbb{N}$, with the following properties: $\varepsilon(k) \to 0$ as $k \to \infty$, $(t_{\varepsilon(k)}, \mathbf{x}_{\varepsilon(k)}) \to (t, \mathbf{x})$ as $k \to \infty$, and $g_{\varepsilon(k)}(u) - \varphi$ has a local maximum at $(t_{\varepsilon(k)}, \mathbf{x}_{\varepsilon(k)})$. (Having fixed this sequence, we will now simplify the notation by writing $\varepsilon(k) = \varepsilon$.)

If $D\varphi(t, \mathbf{x}) \neq 0$, then $D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon) \neq 0$ for all sufficiently small ε, that is, all sufficiently large k. Since we have shown in the first part of the proof that $g_\varepsilon(u)$ is a viscosity solution of

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

it follows from Lemma 19.3 that

$$\frac{\partial \varphi}{\partial t}(t_\varepsilon, \mathbf{x}_\varepsilon) \leq F(D^2\varphi(t_\varepsilon, \mathbf{x}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon))$$

for all sufficiently small ε. Since both sides of this inequality are continuous, we can pass to the limit as $\varepsilon \to 0$ and conclude that

$$\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq F(D^2\varphi(t, \mathbf{x}), D\varphi(t, \mathbf{x})).$$

In case $D^2\varphi(t, \mathbf{x}) = 0$ and $D\varphi(t, \mathbf{x}) = 0$, we call on Lemma 19.4 and write

$$\frac{\partial \varphi}{\partial t}(t_\varepsilon, \mathbf{x}_\varepsilon) \leq G^+(D^2\varphi(t_\varepsilon, \mathbf{x}_\varepsilon), D\varphi(t_\varepsilon, \mathbf{x}_\varepsilon)),$$

where G^+ satisfies the conditions of Definition 19.1. By passing to the limit and using the fact that $G^+(0, 0) = 0$, we see that

$$\frac{\partial \varphi}{\partial t}(t, \mathbf{x}) \leq 0.$$

This proves that $g(u)$ is a viscosity subsolution of $\partial\varphi/\partial t = F(D^2u, Du)$; the same proof adapts to prove that it is a viscosity supersolution. □

20.3 Uniform continuity of approximate solutions

Lemma 20.3. *Consider scaled monotone translation invariant operators T_h defined on the set of Lipschitz functions on $\mathbb{R}^N$. Assume that they commute with the addition of constants and that there exists a continuous real function, $\epsilon(t)$ satisfying $\epsilon(0) = 0$ and such that for $nh \leq t$, $((T_h)^n(L|x|))(0) \leq L\epsilon(t)$ and $((T_h)^n(-L|x|))(0) \geq -L\epsilon(t)$. Then for every L-Lipschitz function u_0, one has $-L\epsilon(t) \leq ((T_h)^n u_0)(\mathbf{x}) - u_0(\mathbf{x}) \leq L\epsilon(t)$.*

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

$$-L|\mathbf{x}| \le u_0(\mathbf{x}) - u_0(0) \le L|\mathbf{x}|$$

Applying $(T_h)^n$, using its monotonicity and its commutation with the addition of constants and taking the value at 0,

$$((T_h)^n(-L\mathbf{x}))(0) \le ((T_h)^n u_0)(0) - u_0(0) \le ((T_h)^n(L\mathbf{x}))(0),$$

that is, by assumption if $nh \le t$,

$$-L\epsilon(t) \le ((T_h)^n u_0)(0) - u_0(0) \le L\epsilon(t).$$

□

Lemma 20.4. *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 Lemma 20.3. Assume in addition that the associated function $\varepsilon(h)$ is concave near 0. Then the approximate solutions $u_h(t,\mathbf{x})$ associated with T_h are uniformly equicontinuous when we restrict t to the set $h\mathbb{N}$. More precisely, for all $n, m \in \mathbb{N}$ and all $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^N$,*

$$|u_h(nh,\mathbf{x}) - u_h(mh,\mathbf{y})| \le L|\mathbf{x} - \mathbf{y}| + \epsilon(|n-m|h). \tag{20.11}$$

We can extend u_h into functions $\tilde{u}_h$ on $\mathbb{R}^+ \times \mathbb{R}^N$ which are uniformly equicontinuous. As a consequence, there are sequences u_{h_n}, with $h_n \to 0$, which converge uniformly on every compact subset of $\mathbb{R}^+ \times \mathbb{R}^N$.

Proof. Since by definition $u_h(nh,\mathbf{x}) = ((T_h)^n u_0)(\mathbf{x})$, the result is a direct consequence of Lemmas 7.11 and 20.3 : By the first mentioned lemma,

$$|u_h(nh,\mathbf{x}) - u_h(nh,\mathbf{y})| \le L|\mathbf{x} - \mathbf{y}|$$

and by the second one applied with $(T_h)^{n-m}$,

$$|u_h(nh,\mathbf{x}) - u_h(mh,\mathbf{x})| = |((T_h)^{n-m} u_h(mh,.))(\mathbf{x}) - u_h(mh,\mathbf{x})| \le \epsilon(|n-m|h).$$

Thus, we obtain (20.11) by remarking that

$$|u_h(nh,\mathbf{x}) - u_h(mh,\mathbf{y})| \le |u_h(nh,\mathbf{x}) - u_h(nh,\mathbf{y})| + |u_h(nh,\mathbf{y}) - u_h(mh,\mathbf{y})|. \tag{20.12}$$

Consider the linear interpolation of u_h,

$$\tilde{u}_h(t,\mathbf{x}) = \frac{t-nh}{h} u_h((n+1)h,\mathbf{x}) + \frac{(n+1)h-t}{h} u_h(nh,\mathbf{x}).$$

Since ϵ is concave, the function $\frac{\epsilon(h)}{h}$ is nonincreasing. It follows that

$$|\tilde{u}_h(t,\mathbf{y}) - \tilde{u}_h(s,\mathbf{y})| \le \frac{\epsilon(h)}{h}|t-s| \le \epsilon(|t-s|) \le \text{ for } |t-s| \le h \tag{20.13}$$

and

$$|\tilde{u}(t,\mathbf{y}) - \tilde{u}(s,\mathbf{y})| \le 3\epsilon(|t-s|) \text{ for } |t-s| \ge h. \tag{20.14}$$

(See Exercise 20.2.) By using again (20.12) we conclude that the family of functions $\tilde{u}_h$ is uniformly equicontinuous on all of $[0,+\infty[\times\mathbb{R}^N$. Notice that $\tilde{u}_h(0,\mathbf{x}) = u_0(\mathbf{x})$ is fixed. Thus, we can apply Ascoli-Arzela Theorem which asserts that under such conditions, the family of functions $\tilde{u}_h(t,\mathbf{x})$ has a subsequence converging uniformly on every compact set of $[0,+\infty]\times\mathbb{R}^N$ towards a uniformly continuous function $u(t,\mathbf{x})$. The same conclusion holds for $u_h(t,\mathbf{x})$. □

Exercise 20.2. Proof of (20.13), (20.14).
a) Assume first that t, s belong to some $[nh,(n+1)h[$ and prove (20.13) in that case.
b) Assume that $|t-s| \leq h$ and $t \leq nh \leq s$. By using $|u(t)-u(s)| \leq |u(t)-u(nh)| + |u(nh)-u(s)|$ prove again (20.13).
c) If $|t-s| > h$ there are m, n such that $(n-1)h \leq t < nh \leq mh \leq s < (m+1)h$. By using again the triangular inequality and the fact that $|(m-n)h| \leq |t-s|$ and $h < |t-s|$, prove (20.14). ■

Exercise 20.3. Consider the assumptions of Lemmas and 20.3 and see whether the results of these lemmas can be extended to the case where u_0 is assumed to be uniformly continuous instead of Lipschitz. More precisely, assume that there exists a continuous increasing function $\varepsilon : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\varepsilon(0) = 0$ and $|u_0(\mathbf{x}) - u_0(\mathbf{y})| \leq \varepsilon(||\mathbf{x}-\mathbf{y}||)$. Hint: use Corollary 7.12.
■

20.4 Convergence of iterated median filters to the mean curvature motion

We shall prove in Theorem 20.6 one of the main practical and theoretical results of this book : the iterated median filters converge to the mean curvature motion equation. The action of iterated median filters and the action of the corresponding PDE are illustrated and compared in Figures 20.1 and 20.2 and show how true this theorem is.

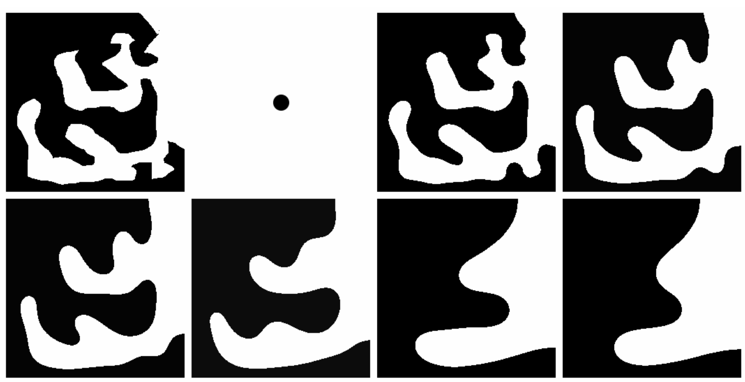

Figure 20.1: Scale-space based on iterations of the median filter. From left to right and top to bottom: original shape, size of the disk used for the median filter, and the results of applying the iterated median filter for an increasing number of iterations.

Figure 20.2: Comparing an iterated median filter with a curvature motion. Numerically, the iterated median filter and the curvature motion must be very close, at least when the curvatures of the level lines are not too small. Indeed, the iterated median filter converges towards the curvature motion. Left: the initial shape of Figure 20.1 has been smoothed by a finite difference scheme of the curvature motion; middle: smoothing with a median filter at the same scale; right: difference between left and middle images. The difference is no greater than the width of one pixel. To have a rigorous comparison, scales have been calibrated by ensuring that for both schemes and all r a circle with radius r vanishes at scale r

Lemma 20.5. *(median filter) Let k be a radial, nonnegative, non separable, compactly supported 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, $T_h u(\mathbf{x}) = \mathrm{Med}_{k_h} u(\mathbf{x})$. Set $v_0(\mathbf{x}) = v_0(|\mathbf{x}|) = v_0(r) = Lr$. Then, if $nh^2 \leq t$,*

$$(T_h^n v)(0) \leq L\sqrt{2t} \text{ and } T_h^n(-v)(0) \geq -L\sqrt{2t}$$

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

$$\mathrm{meas}_{k_h}(\{\mathbf{y}, |\mathbf{y}| \geq \sqrt{r^2+h^2}\} \cap B - \mathbf{x}) \leq \mathrm{meas}_{k_h}(\{\mathbf{y}, |\mathbf{y}| \leq \sqrt{r^2+h^2}\} \cap B - \mathbf{x}). \tag{20.15}$$

As a consequence, v being nondecreasing, we have

$$\mathrm{med}_{k_h} v(\mathbf{x}) \leq v(\sqrt{r^2+h^2}). \tag{20.16}$$

Let us set for brevity $f_h(r) = \sqrt{r^2+h^2}$ and $r_{n+1}(r) = f_h(r_n)$, $r_0 = r$. Then we obviously have from (20.16) and the monotonicity of T_h

$$(T_h^n v)(r) \leq v(r_n(r)). \tag{20.17}$$

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

$$r_n \leq r + \frac{nh^2}{2r}. \tag{20.18}$$

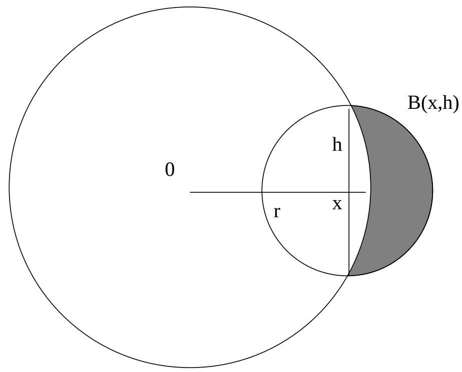

Figure 20.3: Illustrating the inequality (20.15).

Let us assume that $nh^2 \leq t$. Taking into account that v is a nondecreasing function, (20.17-20.18) yield

$$(T_h^n v)(r) \leq v(r + \frac{t}{2r}). \tag{20.19}$$

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 $nh^2 \leq t$

$$(T_h^n v)(0) \leq L\sqrt{2t}.$$

The kernel k_h being non separable, the second announced inequality comes from the self-duality of the median, namely $T_h(-v) = -T_h(v)$. Applying this iteratively we deduce that $(T_h^n(-v))(0) = -T_h^n v(0) \geq -L(\sqrt{2t})$.

□

Exercise 20.4. Fill in the details of the arguments leading to Equations (20.16) and (20.17).

■

Theorem 20.6. Convergence of iterated weighted median filter. *Let k_h be either a C^∞ compactly supported non separable radial function (in any dimension), or the uniform distribution on the unit disk in $\mathbb{R}^2$. and $(T_h u) = \mathrm{Med}_{k_h} u$. Let $u_0 \in \mathcal{F}$. Then the approximate solutions u_h associated with u_0 and Med_{k_h} converge to a viscosity solution u of*

$$\frac{\partial u}{\partial t} = \frac{1}{2} c_k |Du| \mathrm{curv}(u), \tag{20.20}$$

where $c_k = \frac{1}{3}$ if k is the uniform measure on unit disk in $\mathbb{R}^2$ and c_k is the constant specified in Lemma 15.2 otherwise. Incidentally, this proves the existence of a (unique) viscosity solution to the curvature equation.

Proof. We know by Theorems 14.7 and 15.3 that the weighted median is uniformly consistent with (20.20). 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 20.5, so that the assumption of Lemma 20.3 is true. In addition, we know that Med_{k_h} is monotone, satisfies the local comparison principle (19.17), commutes with translations and the addition of constants. Thus, we can apply Lemma 20.3 which asserts that a subsequence of the approximate solutions u_h converges uniformly on compact sets of $\mathbb{R}^+ \times \mathbb{R}^N$ to a function u. In addition, by Proposition 19.14, u is a viscosity solution of (20.20). Since by Theorem 19.17, this solution is unique, we deduce that the whole sequence u_h converges to u. We have thus proved both existence of a viscosity solution for the mean curvature motion and the convergence of the iterated median filters. □

20.5 Convergence of iterated affine-invariant operators to affine-invariant curvature motion

In this section we consider any affine invariant contrast invariant filter associated with an affine invariant, 1-localizable structuring set $\mathcal{B}$. Let

$$IS_h u(\mathbf{x}) = \inf_{B \in h^{\frac{3}{2}}\mathcal{B}} \sup_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}) \quad \text{and} \quad SI_h u(\mathbf{x}) = \sup_{B \in h^{\frac{3}{2}}\mathcal{B}} \inf_{\mathbf{y} \in \mathbf{x}+B} u(\mathbf{y}),$$

and let T_h denote one to the operators IS_h, SI_h, or $SI_h IS_h$. We recall that we have defined u_h, the approximate solutions generated by T_h with an initial function $u_0 \in \mathcal{F}$, by

$$u_h(\mathbf{x}, (n+1)h) = T_h u_h(\mathbf{x}, nh), \quad u_h(\mathbf{x}, 0) = u_0(\mathbf{x}).$$

Theorem 20.7. *Let $\mathcal{B}$ an affine invariant, 1-localizable structuring set such that $c^+_{\mathcal{B}} > 0$ and that every $B \in \mathcal{B}$ contains 0. Then the sequence $\{u_h\}$ converges, when $h \to 0$, uniformly on compact sets of $\mathbb{R}^+ \times \mathbb{R}^2$ to the unique viscosity solution u of*

$$\frac{\partial u}{\partial t} = c_{\mathcal{B}} |Du| g(\mathrm{curv}(u)),$$

where

$$g(r) = \begin{cases} (r^+/2)^{1/3} & \text{if} \quad T_h = IS_h, \\ (r^-/2)^{1/3} & \text{if} \quad T_h = SI_h, \\ (r/2)^{1/3} & \text{if} \quad T_h = SI_h IS_h, \end{cases} \tag{20.21}$$

and $c_{\mathcal{B}} = c^+_{\mathcal{B}}$.

By Barles-Souganidis principle, Theorem 20.7 essentially is a consequence of Lemma **??** and Theorem **??**, which state a consistency result for the schemes SI_h, IS_h, $SI_h IS_h$. In order to achieve the proof of Theorem 20.7, we need to check that the assumptions of Lemma 20.3 are satisfied.

Lemma 20.8. *Consider any radial nondecreasing function $v(\mathbf{x}) = v(|\mathbf{x}|) = v(r) \geq 0$. Then for $nh \leq t$,*

$$0 \leq ((IS_h)^n v)(0) \leq v(at + 2a\sqrt{t}).$$

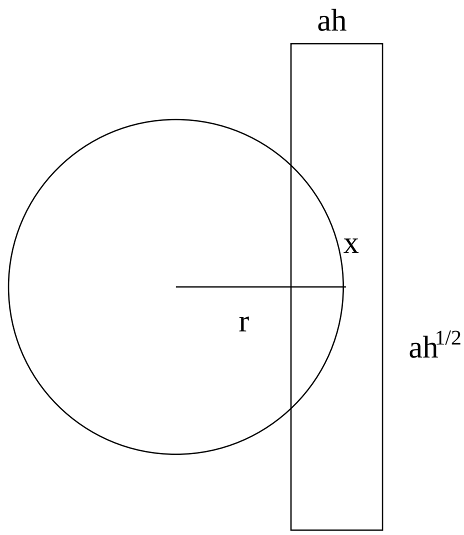

Figure 20.4: Illustration of the proof of the inequality (**??**).

Proof. $(SI_h)^n v$ is easily shown to be radial and nondecreasing, like v. Since $\mathcal{B}$ is localizable, it can be assumed to contain, by Lemma **??**, a square with area a^2. Set $\mathbf{x} = (r, 0)$. Since $\mathcal{B}$ is affine invariant, $\mathcal{B}_h = h^{\frac{3}{2}}\mathcal{B}$ contains a rectangle R_h with sides parallel to the axes, the side parallel to the x-axis having length ah and the other one $a^2 h^{\frac{1}{2}}$. Then

$$IS_h v(\mathbf{x}) \le \sup_{\mathbf{x}+R_h} v(\mathbf{y}).$$

Thus

$$IS_h v(\mathbf{x}) \le v((r + \frac{ah}{2})^2 + a^2 h)^{\frac{1}{2}}) \tag{20.22}$$

We set for conciseness $f_h(r) = ((r + \frac{ah}{2})^2 + a^2 h)^{\frac{1}{2}}$ and $r_{n+1}(r) = f_h(r_n)$, $r_0 = r$. Since $IS_h v$ is a radial nondecreasing function, we can replace v by $IS_h v$ in (20.22). By the monotonicity of IS_h, we obtain

$$(T_h^n v)(r) \le v(r_n(r)) \tag{20.23}$$

In addition, since $(r^2 + \varepsilon)^{\frac{1}{2}} \le r + \frac{1}{2r}\varepsilon$ for all $r, \varepsilon > 0$, we have for $h \le 1$

$$f_h(r) \le (r^2 + ahr + a^2 h + \frac{1}{4}a^2 h^2)^{\frac{1}{2}} \le (r^2 + 2a^2 h + ahr)^{\frac{1}{2}} \le r + \frac{1}{2r}(2a^2 h + ahr),$$

which yields

$$f_h(r) \le r + ah + \frac{a^2 h}{r}. \tag{20.24}$$

Thus $r_{n+1} = f_h(r_n) \le r_n + ah + \frac{a^2 h}{r_n} \le r_n + ah + \frac{a^2 h}{r}$, because r_n is an increasing sequence. Finally, $r_n \le r + n(ah + \frac{a^2 h}{r})$ and, by (20.23),

$$(SI_h^n v)(r) \le v(r + n(ah + \frac{a^2 h}{r})).$$

Let us assume that $nh \leq t$. Then

$$(SI_h^n v)(r) \leq v(r + (a + \frac{a^2}{r})t).$$

Considering that the minimum value of $r \to r + (a + \frac{a^2}{r})t$ is attained at $r = a\sqrt{t}$ and that $r \to (IS_h)^n v(r)$ is nondecreasing, we obtain for $nh \leq t$,

$$0 \leq (IS_h^n v)(0) \leq v(2a\sqrt{t} + at).$$

□

Corollary 20.9. *The operators $T_h = IS_h$, SI_h, $IS_h SI_h$ all satisfy*

$$-v(at + 2a\sqrt{t}) \leq (T_h^n v)(0) \leq v(at + 2a\sqrt{t})$$

for every continuous radial nonincreasing function $v \geq 0$.

Proof. We claim that $SI_h v = v$. By Lemma **??**, for every ε, $\mathcal{B}_h$ contains a rectangle whose side parallel to the x axis has length ε. Thus

$$SI_h v(\mathbf{x}) \geq \inf_{\mathbf{y} \in \mathbf{x} + R_h} v(\mathbf{y}) = v(r - \frac{\varepsilon}{2}) \to v(\mathbf{x}) \text{ as } \varepsilon \to 0.$$

Since every $B \in \mathcal{B}$ contains 0, we also have $SI_h v(\mathbf{x}) \leq v(\mathbf{x})$, which proves the claim.

Since $IS_h v$ is a radial nondecreasing continuous function like v, we also have $SI_h(IS_h v) = IS_h v$ and by iterating and using Lemma 20.8,

$$((SI_h IS_h)^n v)(0) = ((IS_h)^n)v(0) \leq v(at + 2a\sqrt{t}).$$

We also have by the same lemma,

$$((SI_h IS_h)^n(-v))(0) = -((IS_h SI_h)^n v)(0) = -((IS_h)^n v)(0) \geq -v(at + 2a\sqrt{t}.$$

Finally, $(SI_h)^n v = -(IS_h)^n(-v)$, which yields the same inequalities for $(SI_h)^n v(0)$ as for $(IS_h)^n v(0)$. □

Figures 20.5 and 20.6 illustrate numerical results showing that affine-invariant filters really are affine invariant. A finite difference scheme is used to compute the action of the PDE in Figure 20.5. Figure 20.6 illustrates the same invariance using an iterated. inf-sup operator.

of Theorem 20.7. By Lemma **??**, Theorem **??** and Theorem **??** the operators T_h are consistent with their corresponding partial differential equations $\frac{\partial u}{\partial t} = c_{\mathcal{B}}|Du|g(\text{curv}(u))$, and satisfy a uniform local maximum principle. Being contrast invariant, they commute with the addition of constants. Thus, by Proposition 19.14, if a sequence of approximate uniformly continuous solutions u_{h_n} converges uniformly on every compact set to a function u, then u is a viscosity solution of (19.14).
By Lemmas 20.3 and 20.8, the approximate solutions u_h are equicontinuous on every compact set of $\mathbb{R}^+ \times \mathbb{R}^N$ and therefore have subsequences which converge

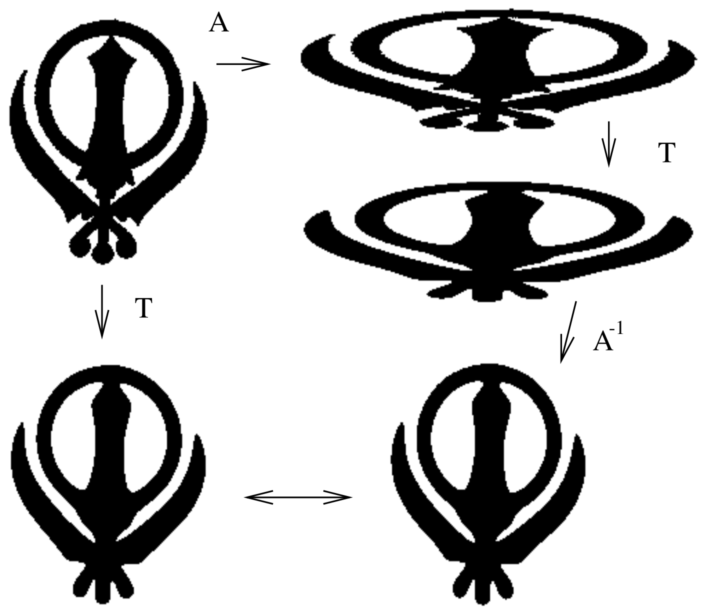

Figure 20.5: Affine invariance of (AMSS). We check the affine invariance of the affine and morphological scale space (AMSS). A simple shape (top-left) is smoothed using a finite differences discretization of (AMSS) followed by thresholding (bottom-left). We apply an affine transform, with determinant equal to 1, on the same shape (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.

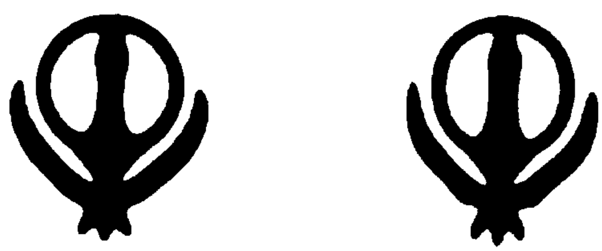

Figure 20.6: Checking the invariance of an affine-invariant inf-sup operator. The images are the final outcomes of the same comparison process shown in Figure 20.5, with T replaced with an affine-invariant inf-sup operator. The structuring set $\mathcal{B}$ is an approximately affine-invariant set of 49 ellipses, all with same area. The inf-sup computation is costly and proves to be less affine invariant than the one obtained by a finite difference scheme. This is due to grid effects.

uniformly to a function u on every compact subset of $\mathbb{R}^+ \times \mathbb{R}^N$. Thus, u is a viscosity solution. In addition, we know that a viscosity solution of (20.21) is unique (Theorem 19.17). Thus the limit u does not depend on the particular considered subsequence and the whole sequence u_h converges to u. So we have proven both the existence of a viscosity solution for the affine invariant equations and the convergence of u_h to this solution. □

20.6 Exercises

Exercise 20.5. Let $g : \mathbb{R} \to \mathbb{R}$ be continuous and nondecreasing and set

$$g_\varepsilon(s) = \int_{\mathbb{R}} \psi_\varepsilon(s-t)(g(t) + \varepsilon t)\,\mathrm{d}t,$$

where the support of ψ_ε is in $[-\varepsilon, \varepsilon]$, ψ_ε is C^∞, $\psi_\varepsilon(s) \geq 0$, and $\int_{\mathbb{R}} \psi_\varepsilon(t)\,\mathrm{d}t = 1$. Show that g_ε is C^∞, that $g'_\varepsilon(s) \geq \varepsilon$, and that $g_\varepsilon \to g$ uniformly on compact subsets of $\mathbb{R}$ as $\varepsilon \to 0$. ■

20.7 Comments and references

The existence and uniqueness theory for the viscosity solutions of mean curvature motion and the relations of these solutions with other kinds of solutions (classical, variational) was developed independently by Evans and Spruck [99, 100, 101, 102] and by Chen, Gigo, and Goto [70, 71]. We do not follow their existence proofs, but rather the elegant numerical approximation schemes invented by Merriman, Bence, and Osher [203] and the subsequent convergence proof to the viscosity solution by Barles and Georgelin [36]. Other proofs of the convergence of the iterated Gaussian median filter toward the mean curvature equation are given in [97] using semigroups and by Ishii [142]. Finally, we note the importance of iterated median filters for denoising applications [23] and [156].

Part V

An Axiomatic Theory of Scale Spaces

Chapter 21

Scale Spaces and Partial Differential Equations

This chapter and the next one are devoted to an axiomatic development of image smoothing. Our approach is based on the notion of a *scale space*:

Definition 21.1. *A* scale space *is a family of image (function) operators* $\{T_t\}$, $t \in \mathbb{R}^+$*, defined on* $\mathcal{F}$*.*

Although this concept is completely abstract, it is clearly based on work presented in previous chapters. One can think of an operator T_t belonging to a scale space as the asymptotic limit of iterated filters, and in fact, the main purpose for developing this abstract theory is to classify and model the possible asymptotic behaviors of iterated filters. The program proceeds as follows: We first introduce several properties that smoothing operators are reasonably expected to have. These properties will be recognized as abstractions of results about iterated filters that we have already encountered; in particular, the function F that has played such an important role in relating iterated operators to differential equations appears in Definition 21.6. At this stage, we will have formally identified scale spaces with the operator mapping u_0 to $u(t)$, where $u(t, \mathbf{x})$ is a solution of a parabolic partial differential equation

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

The next step is to define the now-familiar invariants for scale spaces, one at a time, and then deduce properties that F must have based on the assumed invariants of $\{T_t\}$. This will lead, for example, to a complete characterization of the function F for a linear scale space as the Laplacian.

The chapter contains seven figures that illustrate some of the concepts. Figures 21.4, 21.5, 21.6, and 21.7 are placed at the end of the chapter. Figure 21.4 illustrates numerically that linear smoothing is not contrast invariant. On the other hand, Figure 21.5 shows experimentally that the scale space AMSS is contrast invariant. The significance of contrast invariance for smoothing T-junctions is illustrated in Figure 21.6. Figure 21.7 illustrates one of the most important twentieth century discoveries about human vision and compares it with computer vision.

21.1 Basic assumptions about scale spaces

In our context, the operators T_t are smoothing operators and the functions u are images. Thus, given an image u_0, $T_t u_0 = u(t, \cdot)$ is the image u_0 smoothed at scale t. It is natural to abstract the idea that an image smoothed at scale t can be obtained from the image smoothed at scale s, $s < t$, without having to "go back" to the original image u_0. This concept is illustrated in Figure 21.1 and formulated in the next definition.

Definition 21.2. *A scale space $\{T_t\}$ is said to be* pyramidal *if there is another family of operators $\{T_{t+h,t}\} : \mathcal{F} \to \mathcal{F}$, $h \geq 0$, called transition operators, such that*

$$T_{t+h} = T_{t+h,t} T_t \quad \textit{and} \quad T_0 = I,$$

where I denotes the identity operator.

We will sometimes denote the transition operators by $\{T_{s,t}\}$, $0 \leq t \leq s$. Then $T_s = T_{s,t} T_t$, $h = s - t$, and $T_{t,t} = I$. Most, but not all, results are about pyramidal scale spaces. An important exception is Lemma 21.21, which is a key result in our program.

A strong version of "pyramidal" is the semigroup property. Recall that we have already encountered this idea in Chapter 9 in connection with a dilation or an erosion generated by a convex set.

Definition 21.3. *A scale space $\{T_t\}$ is said to be* recursive *if $T_0 = I$ and*

$$T_s T_t = T_{s+t} \quad \textit{for all} \quad s, t \in \mathbb{R}.$$

Note that if $\{T_t\}$ is recursive, then T_t can be obtained by iterating $T_{t/n}$ n times. Another intuitive concept is that of "locality." The thought that the action of a smoothing operator on a function u at $\mathbf{x}$ would be sensitive to what the function did far from $\mathbf{x}$ just does not make sense. This means that we want the action of the transition operators to depend essentially on the values of $u(\mathbf{y})$ for $\mathbf{y}$ near $\mathbf{x}$. Furthermore, we have had ample opportunity in earlier chapters to see the technical importance of locality. The related property of being monotonic is also intuitively and technically important. We combine these notions is the next definition.

Definition 21.4. *A scale space $\{T_t\}$ satisfies a* local comparison principle *if the following implications are true: For all u and v in the domain of definition, $u(\mathbf{y}) \leq v(\mathbf{y})$ for $\mathbf{y}$ in some neighborhood of $\mathbf{x}$ implies that*

$$T_{t+h,t} u(\mathbf{x}) \leq T_{t+h,t} v(\mathbf{x}) + o(h) \quad \textit{for all sufficiently small } h.$$

If $u(\mathbf{y}) \leq v(\mathbf{y})$ for all $\mathbf{y} \in \mathbb{R}^N$, then

$$T_{t+h,t} u(\mathbf{x}) \leq T_{t+h,t} v(\mathbf{x}) \quad \textit{for all } \mathbf{x} \in \mathbb{R}^N \textit{ and all } h > 0.$$

Our goal is to establish a classification of scale spaces. To do this, we need an assumption stating that a smooth image evolves smoothly with the scale space. From what we have seen in previous chapters, it should not be surprising that it is sufficient to assume this kind of property for quadratic functions. Now, quadratic functions are not allowed to us, as they do not belong to $\mathcal{F}$. Now, there are functions in $\mathcal{F}$ which coincide locally with every quadratic functions and this is enough for our scopes.

Definition 21.5. *We say that u is a "quadratic function around $\mathbf{x}$" if it belongs to $\mathcal{F}$ and if for all $\mathbf{y}$ in some $B(\mathbf{x}, r)$, $r > 0$, one has*

$$u(\mathbf{y}) = \frac{1}{2}\langle A(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \langle p, \mathbf{y} - \mathbf{x}\rangle + c,$$

where $A = D^2u(\mathbf{x})$ is an $N \times N$ matrix, $p = Du(\mathbf{x})$ is a vector in $\mathbb{R}^N$, and $c = u(\mathbf{x})$ is a constant.

From the semigroup point of view, the next assumption implies the existence of an infinitesimal generator for the semigroup T_t.

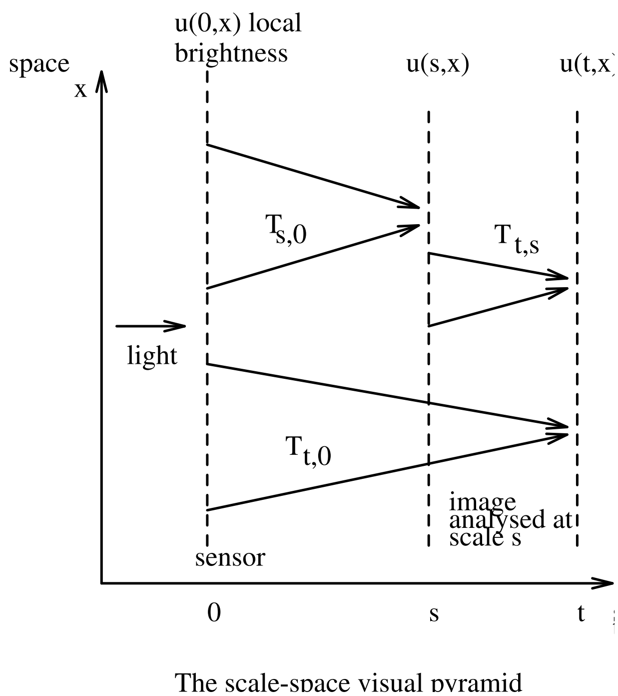

Figure 21.1: The visual pyramid of scale space. Perception is thought of as a flow of images passing through transition operators $T_{t,s}$. These operators receive an image previously analyzed at scale s and deliver an image analyzed at a larger scale t. The scale $t = 0$ corresponds to the original percept. In this simple model, the perception process is irreversible: There is no feedback from coarse scales to fine scales.

Definition 21.6. *A scale space $\{T_t\}$ is said to be* regular *if there exists a function*

$$F : (A, p, \mathbf{x}, c, t) \mapsto F(A, p, \mathbf{x}, c, t)$$

that is continuous with respect to A and such that for every quadratic function u around $\mathbf{x}$,

$$\frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} \to F(D^2u(\mathbf{x}), Du(\mathbf{x}), \mathbf{x}, u(\mathbf{x}), t) \quad \textit{as} \quad h \to 0. \tag{21.1}$$

It is useful to write (21.1) as $T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x}) = hF(A, p, \mathbf{x}, c, t) + o(h)$. Then, replacing t with $t - h$ shows that

$$T_{t,t-h}u(\mathbf{x}) - u(\mathbf{x}) = hF(A, p, \mathbf{x}, c, t - h) + o(h). \tag{21.2}$$

Then, if F is continuous in t,

$$\frac{T_{t,t-h}u(\mathbf{x}) - u(\mathbf{x})}{h} \to F(A, p, \mathbf{x}, c, t) \quad \text{as} \quad h \to 0. \tag{21.3}$$

We encountered the notion of causality in section **??**. The idea was quite simple, if not precise: As scale increases, no new features should be introduced by the smoothing operators. The image at scale $t' > t$ should be simpler than the image at scale t. Since we will constantly be considering scale spaces that are pyramidal and regular, and that satisfy the local comparison principal, it will be convenient to give these scale spaces a name. The causality entails a further property for F:

Definition 21.7. *A scale space $\{T_t\}$ is said to be* causal *if it is pyramidal and regular, and if it satisfies the local comparison principle.*

Lemma 21.8. *If the scale space $\{T_t\}$ is causal, then the function F is nondecreasing with respect to its first argument, that is, if $A \leq B$, where A and B are symmetric matrices, then*

$$F(A, p, c, \mathbf{x}, t) \leq F(B, p, c, \mathbf{x}, t). \tag{21.4}$$

Proof. Let A and B be any $N \times N$ symmetric matrices with $A \leq B$, and let p be any N-dimensional vector. Consider the quadratic functions Q_A and Q_B around $\mathbf{x}$ defined by

$$Q_A(\mathbf{y}) = c + \langle p, \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\langle A(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle;$$
$$Q_B(\mathbf{y}) = c + \langle p, \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\langle B(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle.$$

Then for fixed $\mathbf{x}$ and all $\mathbf{y}$ in a neighborhood of $\mathbf{x}$, $Q_A(\mathbf{y}) \leq Q_B(\mathbf{y})$. Using the local comparison principle, we conclude that $T_{t+h,t}Q_A(\mathbf{x}) \leq T_{t+h,t}Q_B(\mathbf{x})$. Noting that $Q_A(\mathbf{x}) = Q_B(\mathbf{x}) = c$ and using the regularity of $\{T_t\}_{t \in \mathbb{R}^+}$, we see that

$$\lim_{h \to 0} \frac{T_{t+h,t}Q_A(\mathbf{x}) - Q_A(\mathbf{x})}{h} \leq \lim_{h \to 0} \frac{T_{t+h,t}Q_B(\mathbf{x}) - Q_B(\mathbf{x})}{h},$$

which is the inequality $F(A, p, c, \mathbf{x}, t) \leq F(B, p, c, \mathbf{x}, t)$. □

We will see in the next section that the causality assumption implies that the scale space is governed by a PDE.

21.2 Causal scale spaces are governed by PDEs

The next result, Theorem 21.9, should be no surprise. It just says that for causal scale spaces, the regularity condition, which is defined in terms of quadratic forms, transfers directly to functions u that are C^2. This is a fundamental, although easily established, step in our program. Once we have established Theorem 17.8, we are ready to introduce invariants: Postulate that the scale space has certain invariance properties and conclude that F must have certain properties. This will tell us that causal scale spaces with certain invariances will be governed by a general class of PDEs.

Theorem 21.9. *Assume that the scale space $\{T_t\}$ is causal and that u is C^2 at $\mathbf{x}$. Then there exists a function F such that for all $\mathbf{x} \in \mathbb{R}^N$,*

$$\frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} \to F(D^2u(\mathbf{x}), Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}, t) \quad \textit{as} \quad h \to 0. \tag{21.5}$$

Proof. Since we have assumed that u is C^2 at $\mathbf{x}$, we can expand u near $\mathbf{x}$ as

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

For $\varepsilon > 0$, define the quadratic functions Q^+ and Q^- around $\mathbf{x}$ by

$$Q^+(\mathbf{y}) = u(\mathbf{x}) + \langle Du(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\langle D^2u(\mathbf{x})(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\varepsilon\langle \mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle;$$
$$Q^-(\mathbf{y}) = u(\mathbf{x}) + \langle Du(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{1}{2}\langle D^2u(\mathbf{x})(\mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle - \frac{1}{2}\varepsilon\langle \mathbf{y} - \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle.$$

For sufficiently small $|\mathbf{y} - \mathbf{x}|$,

$$Q^-(\mathbf{y}) \le u(\mathbf{y}) \le Q^+(\mathbf{y}).$$

Use the facts that the scale space $\{T_t\}$ is pyramidal (so the transition operators exist) and that it satisfies the local comparison principle to deduce that

$$T_{t+h,t}Q^-(\mathbf{x}) - o(h) \le T_{t+h,t}u(\mathbf{x}) \le T_{t+h,t}Q^+(\mathbf{x}) + o(h).$$

Since $Q^-(\mathbf{x}) = u(\mathbf{x}) = Q^+(\mathbf{x})$, we have

$$T_{t+h,t}Q^-(\mathbf{x}) - Q^-(\mathbf{x}) - o(h) \le T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x}) \le T_{t+h,t}Q^+(\mathbf{x}) - Q^+(\mathbf{x}) + o(h).$$

Now divide by h and let it tend to zero. Since $\{T_t\}$ is regular we have the following limits:

$$\lim_{h\to 0} \frac{T_{t+h,t}Q^-(\mathbf{x}) - Q^-(\mathbf{x})}{h} \le \liminf_{h\to 0} \frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} \le \limsup_{h\to 0} \frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h}$$
$$\le \lim_{h\to 0} \frac{T_{t+h,t}Q^+(\mathbf{x}) - Q^+(\mathbf{x})}{h}.$$

Thus,

$$\begin{aligned} F(D^2u(\mathbf{x}) - \varepsilon I, Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}, t) &\le \liminf_{h\to 0} \frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} \\ &\le \limsup_{h\to 0} \frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} \\ &\le F(D^2u(\mathbf{x}) + \varepsilon I, Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}, t). \end{aligned}$$

Part of the regularity assumption is that F is continuous in its first argument, so letting ε tend to zero shows that

$$\lim_{h\to 0} \frac{T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x})}{h} = F(D^2u(\mathbf{x}), Du(\mathbf{x}), u(\mathbf{x}), \mathbf{x}, t). \qquad \square$$

$\square$

This is about all we can conclude concerning the function F; to deduce more about F, we must assume more about the scale space. Most, but not all, of these assumptions will be that the scale space is invariant under some group of operations. Some of these invariants, like affine invariance, are rather special. The first one we consider is, however, an invariance we naturally expect all smoothing operators to have: Smoothing should not alter constants and smoothing should commute with the addition of constants.

Definition 21.10. *A pyramidal scale space* $\{T_t\}$ *is said to be* invariant under grey level translations (*or* commutes with the addition of constants) *if*

$$T_{t+h,t}[0](\mathbf{x}) = 0 \quad \textit{and} \quad T_{t+h,t}(u+C)(\mathbf{x}) = T_{t+h,t}u(\mathbf{x}) + C \tag{21.6}$$

for all u*, all constants* C*, and all* $\mathbf{x} \in \mathbb{R}^N$*.*

If $T_{t+h,t}$ is a linear filter defined by $T_{t+h,t}u = \varphi * u$, then this axiom is equivalent to the condition $\int \varphi(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$.

Proposition 21.11. *Let* $\{T_t\}$ *be a causal scale space that is invariant under grey level translations. Then its associated function* $F : (A, p, c, \mathbf{x}, t) \mapsto F(A, p, c, \mathbf{x}, t)$ *does not depend on* c*. Furthermore,* $F(0, 0, c, \mathbf{x}, t) = 0$*.*

Proof. Consider a quadratic function around $\mathbf{x}$, $u(\mathbf{y}) = (1/2)\langle A(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \langle p, \mathbf{y}-\mathbf{x}\rangle + c$, and let C be an arbitrary real number. By the regularity assumption

$$\frac{T_{t+h,t}(u+C)(\mathbf{x}) - (u+C)(\mathbf{x})}{h} \to F(A, p, c+C, \mathbf{x}, t) \quad \text{as} \quad h \to 0.$$

Using the grey level translation invariance and regularity again,

$$\frac{T_{t+h,t}(u+C)(\mathbf{x}) - (u+C)(\mathbf{x})}{h} = \frac{T_{t+h,t}u(\mathbf{x}) + C - u(\mathbf{x}) - C}{h} \to F(A, p, c, \mathbf{x}, t)$$

as $h \to 0$. These last two limits imply that $F(A, p, c+C, \mathbf{x}, t) = F(A, p, c, \mathbf{x}, t)$, so F does not depend on c. If $A = 0$ and $p = 0$, then $T_{t+h,t}(u)(\mathbf{x}) - (u)(\mathbf{x}) = c - c = 0$ and $F(0, 0, c, \mathbf{x}, t) = 0$. $\square$

From now on, we assume that the scale space is causal and invariant under grey level translations. We will thus suppress c and write $F(A, p, c, \mathbf{x}, t) = F(A, p, \mathbf{x}, t)$.

21.3 Scale spaces yield viscosity solutions

We are going to prove a result that connects a causal scale space $\{T_t\}$ with a viscosity solution of the PDE associated with $\{T_t\}$. In fact, we will prove that $T_t u_0$ is a viscosity solution of the equation

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = F(D^2u(\mathbf{x}), Du(\mathbf{x}), \mathbf{x}, t), \tag{21.7}$$

where F is the function associated with $\{T_t\}$ by regularity (Definition 21.6).

Theorem 21.12. *Assume that a scale space $\{T_t\}$ is causal and commutes with grey level translations; assume also that F is continuous in t. Then the function u defined by $u(t,\mathbf{x}) = T_t u_0(\mathbf{x})$ is a viscosity solution of* (21.7).

Proof. We will show that u is a viscosity subsolution; the proof that it is also a viscosity supersolution is a similar argument with the inequalities going in the opposite direction.

Assume that φ is C^∞ and that $(t,\mathbf{x}) \in (0,+\infty)\times\mathbb{R}^N$ is a point at which the function $u-\varphi$ has a strict local maximum. The point $(t,\mathbf{x})$ is fixed, so we will denote the variable by $(s,\mathbf{y})$. We need to show that

$$\frac{\partial \varphi}{\partial t}(t,\mathbf{x}) - F(D^2\varphi(t,\mathbf{x}), D\varphi(t,\mathbf{x}), \mathbf{x}, t) \le 0. \tag{21.8}$$

(Note that we would usually consider two cases: (1) $D\varphi(t,\mathbf{x}) \neq 0$; (2) $D\varphi(t,\mathbf{x}) = 0$ and $D^2\varphi(t,\mathbf{x}) = 0$. Since $F(0,0,\mathbf{x},t) = 0$ by Proposition 21.11, it is sufficient to prove (21.8).)

On the basis of Lemma 19.5, we assume that φ is of the form $\varphi(s,\mathbf{y}) = f(\mathbf{y}) + g(s)$. Since the operators commute with the addition of constants, we may also assume that $u(t,\mathbf{x}) = \varphi(t,\mathbf{x}) = f(\mathbf{x}) + g(t)$. Of course, both f and g are C^∞. Thus, for $(s,\mathbf{y})$ in some neighborhood of $(t,\mathbf{x})$, we have $u(s,\mathbf{y}) \le \varphi(s,\mathbf{y}) = f(\mathbf{y}) + g(s)$. In particular, we have

$$u(t-h,\mathbf{y}) \le f(\mathbf{y}) + g(t-h)$$

for all sufficiently small $h > 0$. Since the operators $T_{t,t-h}$ satisfy the local comparison principle and commute with the addition of constants, we have

$$T_{t,t-h}u(t-h,\cdot)(\mathbf{x}) \le T_{t,t-h}f(\mathbf{x}) + g(t-h).$$

By definition of the transition operators, $T_{t,t-h}u(t-h,\cdot)(\mathbf{x}) = u(t,\mathbf{x}) = f(\mathbf{x}) + g(t)$. Thus, we see that

$$g(t) - g(t-h) \le T_{t,t-h}f(\mathbf{x}) - f(\mathbf{x}),$$

which by (21.2) we can write as

$$g(t) - g(t-h) \le hF(D^2f(\mathbf{x}), Df(\mathbf{x}), \mathbf{x}, t-h) + o(h).$$

Divide by h, use the fact the F is continuous in t, let h tend to zero, and conclude that

$$g'(t) \le F(D^2f(\mathbf{x}), Df(\mathbf{x}), \mathbf{x}, t).$$

Since $\partial\varphi/\partial t = g'$, $D^2\varphi = D^2 f$, and $D\varphi = Df$, this proves the result. □

21.4 Scale space invariants and implications for F

We have already seen the implication for F of assuming that a scale space $\{T_t\}$ is invariant under the addition of constants (Proposition 21.11). This section is devoted to continuing this program: Assume an invariant for $\{T_t\}$ and deduce its implication for F.

21.4.1 Translation, rotation, and reflection

These invariants concern the underlying space $\mathbb{R}^N$, and they are easily defined. Translation invariance for operators was defined in Definition 7.10 using the translation operator $\tau_{\mathbf{z}}$: $\tau_{\mathbf{z}} u(\mathbf{x}) = u(\mathbf{x} - \mathbf{z})$ for all $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$.

Definition 21.13. *A pyramidal scale space $\{T_t\}$ is said to be* translation invariant *if*

$$T_{t+h,t}\tau_{\mathbf{z}} = \tau_{\mathbf{z}} T_{t+h,t} \quad \textit{for all } \mathbf{z} \in \mathbb{R}^N,\ t \geq 0, \textit{ and } h \geq 0. \tag{21.9}$$

Proposition 21.14. *Assume that $\{T_t\}$ is a causal, translation invariant scale space. Then its associated function F does not depend on $\mathbf{x}$.*

Proof. Consider two quadratic functions around 0 and $\mathbf{x}$ respectively defined by

$$u(\mathbf{y}) = \frac{1}{2}\langle A\mathbf{y}, \mathbf{y}\rangle + \langle p, \mathbf{y}\rangle + c \quad \text{and} \quad \tau_{\mathbf{x}} u(\mathbf{y}) = \frac{1}{2}\langle A(\mathbf{y}-\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \langle p, \mathbf{y}-\mathbf{x}\rangle + c$$

By the regularity assumption,

$$T_{t+h,t}u(0) - u(0) = hF(A, p, 0, c, t) + o(h); \tag{21.10}$$
$$T_{t+h,t}\tau_{\mathbf{x}} u(\mathbf{x}) - \tau_{\mathbf{x}} u(\mathbf{x}) = hF(A, p, \mathbf{x}, c, t) + o(h). \tag{21.11}$$

By translation invariance,

$$T_{t+h,t}\tau_{\mathbf{x}} u(\mathbf{x}) - \tau_{\mathbf{x}} u(\mathbf{x}) = \tau_{\mathbf{x}} T_{t+h,t} u(\mathbf{x}) - \tau_{\mathbf{x}} u(\mathbf{x}) = T_{t+h,t}u(0) - u(0).$$

Thus we see from (21.10) and (21.11) that $hF(A, p, 0, c, t) = hF(A, p, \mathbf{x}, c, t) + o(h)$. Divide both sides by h and let $h \to 0$ to see that

$$F(A, p, \mathbf{x}, c, t) = F(A, p, 0, c, t). \qquad \square$$

$\square$

This takes care of translations; rotations and reflections are combined in the group of linear isometries. If P is a linear isometry of $\mathbb{R}^N$, then the function Pu is defined by $Pu(\mathbf{x}) = u(P\mathbf{x})$.

Definition 21.15. *A pyramidal scale space $\{T_t\}$ is said to be* Euclidean invariant (or isotropic) *if*

$$PT_{t+h,t} = T_{t+h,t}P \tag{21.12}$$

for all linear isometries P of $\mathbb{R}^N$, all $t \geq 0$, and all $h \geq 0$.

We denote the group of linear isometries of $\mathbb{R}^N$ by $O_N = O(\mathbb{R}^N)$. Any transform $P \in O_N$ can be represented uniquely by an $N \times N$ orthogonal matrix P, assuming an orthonormal basis. We do not make a distinction between the operator P and the matrix P. If $P \in O_N$, then recall that its transpose $P' \in O_N$, and $PP' = I$. Recall also that given a symmetric matrix A there is always a $P \in O_N$ such that PAP' is diagonal. Euclidean invariance is illustrated in Figure 21.2.

Lemma 21.16. *If a translation invariant causal scale space $\{T_t\}$ is isotropic, then for every $R \in O_N$,*

$$F(RAR', Rp, t) = F(A, p, t), \tag{21.13}$$

where F is the function associated with $\{T_t\}$ by regularity.

Proof. Consider a quadratic function around 0, $u(\mathbf{y}) = (1/2)\langle A\mathbf{y}, \mathbf{y}\rangle + \langle p, \mathbf{y}\rangle$ and let F be the function associated with $\{T_t\}$. We know that the value of F is determined by the action of $T_{t+h,t}$ on u at t, that is,

$$T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x}) = hF(A, p, t) + o(h).$$

In particular,

$$\lim_{h\to 0} \frac{T_{t+h,t}u(0) - u(0)}{h} = F(A, p, t). \tag{21.14}$$

Let R be any element of O_N. Then

$$u(R\mathbf{y}) = \frac{1}{2}\langle AR\mathbf{y}, R\mathbf{y}\rangle + \langle p, R\mathbf{y}\rangle = \frac{1}{2}\langle R'AR\mathbf{y}, \mathbf{y}\rangle + \langle R'p, \mathbf{y}\rangle.$$

Thus we know immediately that

$$T_{t+h,t}(u \circ R)(\mathbf{x}) - u(R\mathbf{x}) = hF(R'AR, R'p, t) + o(h),$$

where $u \circ R$ denotes the function defined by $u \circ R(\mathbf{y}) = u(R\mathbf{y})$. The assumption that $\{T_t\}$ is isotropic means that $T_{t+h,t}(u \circ R)(\mathbf{x}) = T_{t+h,t}u(R\mathbf{x})$, so

$$T_{t+h,t}(u \circ R)(\mathbf{x}) - u(R\mathbf{x}) = T_{t+h,t}u(R\mathbf{x}) - u(R\mathbf{x}).$$

From this we conclude that

$$\lim_{h\to 0} \frac{T_{t+h,t}(u \circ R)(\mathbf{x}) - u(\mathbf{x})}{h} = \lim_{h\to 0} \frac{T_{t+h,t}u(R\mathbf{x}) - u(R\mathbf{x})}{h} = F(R'AR, R'p, t).$$

By letting $\mathbf{x} = 0$ in these limits, we conclude from (21.14) that

$$F(A, p, t) = F(R'AR, R'p, t).$$

Replacing R with R' completes the proof. □

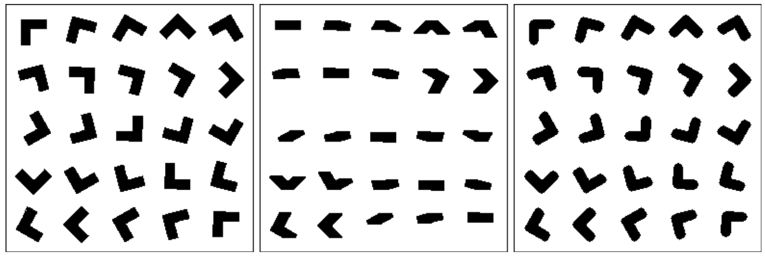

Figure 21.2: An isotropic filter and a nonisotropic filter. The left frame contains simple shapes that can be deduced from each other by rotations. The center image is the closing of the left image by a horizontal rectangle of size 6×2 pixels. This nonisotropic filter produces different results, depending on the shapes orientations. The right image is the closing of the left image by a circle of radius 4 pixels, which has the same area, up to the pixel precision, as the rectangle used in the center image. This filter is isotropic; thus, as one can see, the resulting shapes can be deduced from each other by rotations.

21.4.2 Contrast invariance

Definition 21.17. *A pyramidal scale space $\{T_t\}$ is said to be* contrast invariant *if*

$$g \circ T_{t+h,t} = T_{t+h,t} \circ g$$

for any nondecreasing continuous function $g : \mathbb{R} \to \mathbb{R}$.

An immediate consequence of this definition is that a contrast invariant scale space commutes with the addition of constants, that is, it satisfies Definition 21.10. To see this, just take $g(s) = 0$ and $g(s) = s + C$. Thus, in the next lemma, the function F does not depend on c.

Lemma 21.18. *If a translation invariant causal scale space $\{T_t\}$ is contrast invariant, then its associated function F satisfies the following condition:*

$$F(\mu A + \lambda p \otimes p, \mu p, t) = \mu F(A, p, t), \tag{21.15}$$

where A is any symmetric $N \times N$ matrix, p is any N-dimensional vector, λ is any real number, and μ is any real number greater than or equal to zero.

Proof. Recall that $p \otimes p$ denotes the $N \times N$ matrix whose entries are $p_i p_j$, $i, j \in \{1, 2, \dots, N\}$. Given C^2 functions u and g, we have these two applications of the chain rule:

$$D(g(u)) = g'(u)Du, \quad \text{and} \quad D^2(g(u)) = g'(u)D^2u + g''(u)Du \otimes Du. \tag{21.16}$$

Choose any quadratic function around 0 of the form

$$u(\mathbf{y}) = (1/2)\langle A\mathbf{y}, \mathbf{y}\rangle + \langle p, \mathbf{y}\rangle.$$

We know from the assumptions and equations (21.16), plus the relations $u(0) = 0$, $Du(0) = p$, and $D^2u(0) = A$, that

$$T_{t+h,t}g(u)(0) - g(0) = hF(g'(0)A + g''(0)p \otimes p, g'(0)p, t) + o(h).$$

Since $\{T_t\}$ is contrast invariant, $T_{t+h,t}g(u)(0) = g(T_{t+h,t}u(0))$, so

$$g(T_{t+h,t}u(0)) - g(0) = hF(g'(0)A + g''(0)p \otimes p, g'(0)p, t) + o(h).$$

From regularity, we have

$$T_{t+h,t}u(0) - u(0) = hF(A, p, t) + o(h),$$

so we can write $g(T_{t+h,t}u(0)) = g(hF(A, p, t) + o(h))$. Thus for small enough h,

$$g(T_{t+h,t}u(0)) = g(0) + g'(0)(hF(A, p, t) + o(h)) + o(h)$$

and

$$g'(0)hF(A, p, t) + g'(0)o(h)) + o(h) = hF(g'(0)A + g''(0)p \otimes p, g'(0)p, t) + o(h).$$

Dividing this by h and letting $h \to 0$ shows that

$$g'(0)F(A, p, t) = F(g'(0)A + g''(0)p \otimes p, g'(0)p, t),$$

or, since we can choose a C^2 contrast change g with arbitrary values for $g'(0) \geq 0$ and $g''(0) \in \mathbb{R}$,

$$\mu F(A, p, t) = F(\mu A + \lambda p \otimes p, \mu p, t).$$

$\square$

21.4.3 Scale and affine invariance

The main purpose of this section is to establish a normalized link between scale (t) and space ($\mathbf{x}$). If T_t is a causal scale space and $h : \mathbb{R}^+ \to \mathbb{R}^+$ is a C^2 increasing function, it is easily seen that $S_t = T_{h(t)}$ also is a causal scale space (see Exercise 21.1.) So there is no special link on the scale, unless we give a further specification. This specification will be given by a *scale invariance* axiom.

Scale invariance means intuitively that the result of applying a scale space $\{T_t\}$ must be independent of the size of the analyzed features. This is very important for analyzing natural images, since the same object can be captured at very different distances and therefore at very different scales (see Figure 21.3).

Scale invariance is the object of Definitions **??** and 21.10. The main result is Lemma 21.21 which gives a standard normalization: Scale can be taken proportional to space.

This result can be somewhat secluded from the rest of the invariance analysis, as we will prove it for arbitrary families of function operators $\{T_t\}$, $t \geq 0$, not even pyramidal. We shall just assume that the mapping $t : [0, \infty) \mapsto T_t$ is one-to-one.

The changes of scale on an image can be made by a zoom, in which case the zooming factor λ gives a scale parameter. In the case of an affine transform A, the square root of the determinant of A also will play the role of a scale parameter. By zoom we mean a map $\mathbf{x} \mapsto \lambda\mathbf{x}$, $\lambda > 0$, generating an image transform $H_\lambda u(\mathbf{x}) = u(\lambda\mathbf{x})$.

Definition 21.19. *A family of operators $\{T_t\}$ is said to be* scale invariant *if there exists a rescaling function $t' : (t,\lambda) \mapsto t'(t,\lambda)$, defined for all $\lambda > 0$ and $t \geq 0$, such that*

$$H_\lambda T_{t'} = T_t H_\lambda. \tag{21.17}$$

It $\{T_t\}$ is pyramidal, then

$$H_\lambda T_{t',s'} = T_{t,s} H_\lambda, \tag{21.18}$$

where $t' = t'(t,\lambda)$ and $s' = t'(s,\lambda)$. In addition, the function t' is assumed to be differentiable with respect to t and λ, and the function ϕ defined by $\phi(t) = (\partial t'/\partial\lambda)(t,1)$ is assumed to be continuous and positive for $t > 0$.

This definition implies, in particular, that t' is continuous in t and λ. Condition (21.18) implies (21.17). It will have the advantage of making our classification of scale-invariant scale spaces easier. Of course, we could not impose the condition $t' = t$, since the scale of smoothing and the scale of the image are covariant, as can be appreciated by considering the heat equation.

The assumption that $(\partial t'/\partial\lambda)(t,1) > 0$ can be interpreted by considering the relation $H_\lambda T_{t'} = T_t H_\lambda$ when the scale λ increases before the analysis by T_t, that is, when the size of the image is reduced before analysis. Then the corresponding scale before reduction is increased. Informally, we can say that the scale t of analysis increases with the size of the picture. It is easy to determine the function t' for several classical scale spaces (Exercise 21.5) and to check that it satisfies the previous requirements.

The next definition (axiom) introduces the scale space invariance under any orthographic projection of a planar shape. We write as usual $Au(\mathbf{x}) = u(A\mathbf{x})$.

Definition 21.20. *A family of operators $\{T_t\}$ is said to be* affine invariant *if it is scale invariant and if the following conditions hold: The associated function t' can be extended to a function $t' : (t,A) \mapsto t'(t,A)$, where $t \geq 0$ and A is any linear mapping $A : \mathbb{R}^N \to \mathbb{R}^N$ with $\det(A) \neq 0$, such that $t'(t,\lambda) = t'(t,\lambda I)$ and such that*

$$AT'_t = T_t A. \tag{21.19}$$

If $\{T_t\}$ is pyramidal, the transition operators $T_{t,s}$ satisfy the commutation relation

$$AT_{t',s'} = T_{t,s} A \tag{21.20}$$

for all $0 \leq s \leq t$, where $t' = t'(t,A)$ and $s' = t'(s,A)$.

This property means that the result of applying the scale space $\{T_t\}$ to a image is covariant with the distance and orientation in space of the analyzed planar image (see the introduction of Chapter **??**.) The fact that the function t' can be different for each scale space may seem mysterious. We will "fix" this in the next lemma by showing that we can, "up to a rescaling" assume that scale-invariant scale spaces have all the same scale-space function, namely $t' = \lambda t$.

Lemma 21.21. *[Scale normalization]Assume that the mapping $t \mapsto T_t$, $t \in [0,\infty)$, is one-to-one and that $T_0 = I$.*

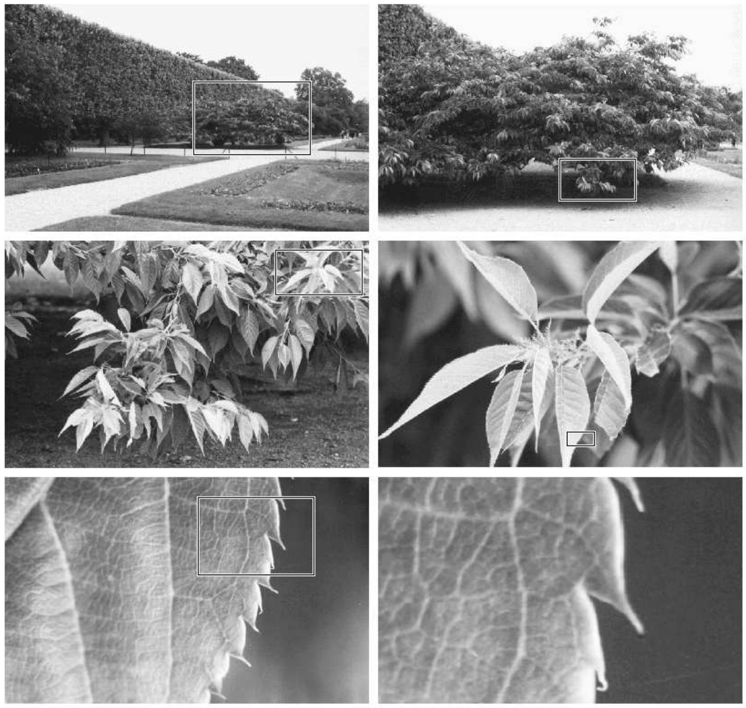

Figure 21.3: A multiscale world. This series of images is an experiment to show the relative perception of objects seen at different distances. Each photograph, after the first one, was taken by stepping forward to produce a snapshot from a distance closer than the one before. The rectangle in each image outlines the part of the object that appears in the next image. Clearly, as one gets closer to the subject, the visual aspect changes and new structures appear. Thus, computing primitives in an image is always a scale-dependent task, and it depends on the distance to objects. When we look at an object from a certain distance, we do not perceive the very fine structure: For instance, leaves cannot be seen in the two first photographs because we are too far from the trees. Nor do we see them in the last two, since we are now too close. Multiscale smoothing of a digital image tries to emulate and actually improve this and phenomenon, due to an optical blur, by defining a smoothing at different scales. The role of this multiscale smoothing is to eliminate the finer structures at a scale t, but minimally modify the image at scales above t.

(*i*) *If the family of operators* $\{T_t\}$ *is scale invariant, then there exists an increasing differentiable function* $\sigma : [0,\infty) \to [0,\infty)$ *such that* $t'(t,\lambda) = \sigma^{-1}(\sigma(t)\lambda)$. *If the operators* S_t *are defined by* $S_t = T_{\sigma^{-1}(t)}$, *then*

$$t'(t,\lambda) = t\lambda \tag{21.21}$$

for the rescaled analysis $\{S_t\}$.

(*ii*) *If the family* $\{T_t\}$ *is affine invariant, then the function* $t'; (t,B) \mapsto t'(t,B)$ *depends only on* t *and* $|\mathrm{det}B|$, *in particular,* $t'(t,B) = t'(t, |\mathrm{det}B|^{1/N})$, *and* t' *is increasing with respect to* t. *In addition, there exists an increas-*

ing differentiable function σ from $[0,\infty)$ to $[0,\infty)$ such that $t'(t,B) = \sigma^{-1}(\sigma(t)|\mathrm{det}B|^{1/N})$. If we set $S_t = T_{\sigma^{-1}(t)}$, then

$$t'(t,B) = t|\mathrm{det}B|^{1/N} \tag{21.22}$$

for the rescaled analysis $\{S_t\}$.

Proof. We will prove (ii) and then show how this proof can be reduced to a proof of (i).

Step 1: We prove that

$$t'(t,AB) = t'(t'(t,A),B) \tag{21.23}$$

for any linear transforms A and B with nonzero determinants. To see this, write

$$ABT_{t'(t,AB)} = T_tAB = AT_{t'(t,A)}B = ABT_{t'(t'(t,A),B)}.$$

Since the determinant of AB does not vanish, we have $T_{t'(t,AB)} = T_{t'(t'(t,A),B)}$, and since $t \mapsto T_t$ is one-to-one, we have (21.23).

Step 2: The function t' is increasing with respect to t.
We begin by proving that $t \mapsto t'(t,A)$ is one-to-one for any A with $\mathrm{det}A \neq 0$. If this were not the case, then there would be some A, $\mathrm{det}A \neq 0$, and some s and t, $s \neq t$, such that $t'(s,A) = t'(t,A)$. This implies that

$$T_sA = AT_{t'(s,A)} = AT_{t'(t,A)} = T_tA.$$

Since $\mathrm{det}A \neq 0$, this means that $T_s = T_t$, and since $t \mapsto T_t$ is one-to-one, we have $s = t$. Thus t' is one-to-one. By hypothesis, $T_0 = I$, and since $AT_{t'(0,A)} = T_0A = A$, we see that $t'(0,A) = 0$ for all A. By definition, t' is continuous and nonnegative. Since $t' \mapsto t'(t,A)$ is one-to-one and since $t'(0,A) = 0$, it is a homeomorphism of $[0,\infty)$ onto $[0,\infty)$ for every A with nonzero determinant. Thus t' is increasing in t.

Step 3: For every orthogonal matrix R,

$$t'(t,R) = t. \tag{21.24}$$

To prove that, define $t_1 = t'(t,R)$ and $t_{n+1} = t'(t_n,R)$. From (21.23), $t_n = t'(t,R^n)$. There are two cases to reject: (1) $t_1 < t$; (2) $t_1 > t$. In case (1), the fact that t' is strictly increasing in t implies that the sequence t_n is strictly decreasing. Similarly, in case (2), the sequence t_n is strictly increasing. Since the set of orthogonal matrices is compact, there is a subsequence n_k and an orthogonal matrix P such that $R^{n_k} \to P$ as $k \to \infty$. Let $m_k = n_{k+1} - n_k$. Then $R^{m_k} \to I$ as $k \to \infty$. Since t' is continuous, $\lim_{k\to\infty} t'(t,R^{m_k}) \to t'(t,I) = t$. In case (1), we have $t = \lim_{k\to\infty} t'(t,R^{m_k}) < t$, a contradiction. In case (2) we have $t = \lim_{k\to\infty} t'(t,R^{m_k}) > t$, a contradiction again.

Step 4: For all transforms B that have nonzero determinants,

$$t'(t,B) = t'(t,|\mathrm{det}B|^{1/N}). \tag{21.25}$$

This part of the proof is pure matrix theory. Let B be any $N \times N$ nonsingular matrix (linear transform of $\mathbb{R}^N$). The B can be written as $B = R_1DR_2$, where

R_1 and R_2 are orthogonal and D is diagonal. Furthermore, $d_{ii} = \lambda_i > 0$ and the λ_i are, up to a sign, the eigenvalues of B. As a consequence, using (21.23) and (21.24), we see that

$$t'(t, B) = t'(t, D).$$

The matrix D can be represented as $D = A(\lambda_1)R_2A(\lambda_2)R_2^{-1} \cdots R_N A(\lambda_N)R_N^{-1}$, where the mapping $A(\lambda_i)$ is defined by $(x_1, x_2, \ldots, x_N) \mapsto (\lambda_i x_1, x_2, \ldots, x_N)$ and R_j is the orthogonal mapping that interchanges x_1 and x_j. Repeated use of (21.23) and (21.24) and the fact that $A(\lambda_1)A(\lambda_2)\cdots A(\lambda_N) = A(\lambda_1\lambda_2\cdots\lambda_N)$ shows that

$$t'(t, D) = t'(t, A(\lambda_1\lambda_2\cdots\lambda_N)).$$

Now write $(\lambda_1\lambda_2\cdots\lambda_N)^{1/N} = \lambda$ and consider the matrix λI. As we have done above, we can write $\lambda I = A(\lambda)R_2A(\lambda)R_2^{-1} \cdots R_N A(\lambda)R_N^{-1}$. Then using (21.23) and (21.24) again, we see that

$$t'(t, \lambda I) = t'(t, A(\lambda^N)) = t'(t, A(\lambda_1\lambda_2\cdots\lambda_N)),$$

and we conclude that

$$t'(t, B) = t'(t, D) = t'(t, A(\lambda_1\lambda_2\cdots\lambda_N)) = t'(t, \lambda I).$$

By definition, $t'(t, \lambda I) = t'(t, \lambda)$, where we have used the same notation for the function $t' : (t, \lambda) \mapsto t'(t, \lambda)$ and its extension $t' : (t, \lambda I) \mapsto t'(t, \lambda I)$. So we obtain (21.25).

Step 5: There is an increasing differentiable function σ that satisfies the equation $t'(t, \lambda) = \sigma^{-1}(\sigma(t)\lambda)$, or equivalently, $\sigma(t'(t, \lambda)) = \sigma(t)\lambda$.
Differentiating the last equation with respect to λ and then setting $\lambda = 1$, shows that

$$\phi\sigma' = \sigma, \tag{21.26}$$

so it is reasonable to define σ by

$$\sigma(t) = \exp\left(\int_1^t \frac{\mathrm{d}s}{\phi(s)}\right).$$

Since by assumption ϕ is continuous and $\phi(s) > 0$ for $s > 0$, σ is clearly increasing and differentiable. It remains to show that $t'(t, \lambda) = \sigma^{-1}(\sigma(t)\lambda)$.

From equations (21.23) and (21.24), we know that $t'(t, \mu\nu) = t'(t'(t, \mu), \nu)$ and $t'(t, 1) = t$ for all positive μ and ν. Differentiating both sides of the first equation with respect to μ and then setting $\mu = 1$ and $\nu = \lambda$ shows that

$$\lambda\frac{\partial t'}{\partial\lambda}(t, \lambda) = \frac{\partial t'}{\partial t}(t, \lambda)\frac{\partial t'}{\partial\lambda}(t, 1). \tag{21.27}$$

By Definition 17.18, the function ϕ given by $\phi(t) = (\partial t'/\partial\lambda)(t, 1)$ is continuous and positive for $t > 0$. We have shown that t' is strictly increasing, thus the right-hand side of (21.27) is nonnegative. This implies that

$$\frac{\partial t'}{\partial\lambda}(t, \lambda) \geq 0. \tag{21.28}$$

If $t'(t, \lambda) = \sigma^{-1}(\sigma(t)\lambda)$ is going to be true, then it is also true if we replace λ with $\lambda/\sigma(t)$, which is the equation $t'(t, \lambda/\sigma(t)) = \sigma^{-1}(\lambda)$. This prompts us to examine the function g defined by $g(t,\lambda) = t'(t, \lambda/\sigma(t))$. When we differentiate g with respect to t, we will see that this derivative is zero:

$$\begin{aligned}
\frac{\partial g}{\partial t}(t,\lambda) &= \frac{\partial t'}{\partial t}(t,\lambda/\sigma(t)) - \lambda\frac{\sigma'(t)}{\sigma^2(t)}\frac{\partial t'}{\partial \lambda}(t,\lambda/\sigma(t)) \\
&= \frac{\partial t'}{\partial t}(t,\lambda/\sigma(t)) - \frac{\sigma'(t)}{\sigma(t)}\frac{\partial t'}{\partial t}(t,\lambda/\sigma(t))\phi(t) \quad \text{(using (21.27))} \\
&= 0 \quad (\text{using } \sigma'\phi = \sigma).
\end{aligned}$$

Thus, g does not depend on t, and we know from (21.28) that g in nondecreasing. Since g is also differentiable, we conclude that $g(t,\lambda) = \beta(\lambda)$, where β is differentiable and nondecreasing. By replacing λ with $\lambda\sigma(t)$, we have

$$t'(t,\lambda) = \beta(\lambda\sigma(t)). \tag{21.29}$$

By differentiating both sides of this equation with respect to λ and then letting $\lambda = 1$, we see that $\phi(t) = \sigma(t)\beta'(\sigma(t)) = \phi(t)\sigma'(t)\beta'(\sigma(t))$. Dividing both sides by $\phi(t)$ shows that

$$\frac{\partial\beta(\sigma(t))}{\partial t}(t) = 1.$$

Integrating this relation from zero to t yields the equation $\beta(\sigma(t)) = t + \beta(\sigma(0))$. Since $t'(0,\lambda) = 0$, $\beta(\sigma(0)) = 0$ by (21.29), and we conclude that $\beta = \sigma^{-1}$.

Step 6: To complete the proof of (*ii*), we must show that the operators S_t by $S_t = T_{\sigma^{-1}(t)}$ are affine invariant with $t'(t,\lambda) = \lambda t$. Thus let B be any nonsingular linear mapping and let $\lambda = |\det B|^{1/N}$. Then

$$S_tB = T_{\sigma^{-1}(t)}B = BT_{t'(\sigma^{-1}(t),\lambda)} = BT_{\sigma^{-1}(\lambda\sigma(\sigma^{-1}(t)))} = BT_{\sigma^{-1}(\lambda t)} = BS_{\lambda t}.$$

The proof of (*i*) is just the "image" of the proof of (*ii*) under the obvious mappings $B \mapsto |\det B|^{1/N}$ and $\lambda I \mapsto H_\lambda$, which entail $t'(t,B) \mapsto t'(t, |\det B|^{1/N})$, and so on. $\square$

Lemma 21.22. *If a translation invariant causal scale space $\{T_t\}$ is affine invariant, then, after the appropriate renormalization, its associated function satisfies the following condition:*

$$F(BAB', Bp, t) = |\det B|^{1/N} F(A, p, |\det B|^{1/N} t) \tag{21.30}$$

for any nonsingular linear map B. If a translation invariant causal scale space $\{T_t\}$ is scale invariant, then, after the appropriate renormalization, its associated function satisfies for any $\mu > 0$

$$F(\mu^2 A, \mu p, t) = \mu F(A, p, \mu t). \tag{21.31}$$

Proof. Recall that we have made the blanket assumption that causal spaces are invariant under the addition of constants. Thus, we assume that F does not depend on c or $\mathbf{x}$. Assume that B is a linear map and that $\lambda = |\det B|^{1/N}$. We also assume that the scale space $\{T_t\}$ is normalized so that $T_{t+h,t}B =$

$BT_{\lambda(t+h),\lambda t}$. Let u be a quadratic function around 0, $u(\mathbf{y}) = (1/2)\langle A\mathbf{y}, \mathbf{y}\rangle + \langle p, \mathbf{y}\rangle$. Then

$$T_{t+h,t}[u(B\mathbf{y})](0) = T_{\lambda(t+h),\lambda t}u(B0) = T_{\lambda(t+h),\lambda t}u(0).$$

Since $u(B'\mathbf{y}) = (1/2)\langle BAB'\mathbf{y}, \mathbf{y}\rangle + \langle Bp, \mathbf{y}\rangle$ around 0 and by the regularity of $\{T_t\}$,

$$T_{t+h,t}[u(B'\mathbf{y})](0) = hF(BAB', Bp, t) + o(h).$$

Also by regularity

$$T_{\lambda(t+h),\lambda t}u(0) = \lambda hF(A, p, \lambda t) + o(\lambda h).$$

Thus we have $hF(BAB', Bp, t) + o(h) = \lambda hF(A, p, \lambda t) + o(\lambda h)$. Dividing by h and letting $h \to 0$ proves the first part of the lemma. To prove the second part, just replace B with μI in the proof of the first part. □

21.5 Axiomatic approach to linear scale space

We are going to use previous results from this chapter, in particularly Theorem 21.9, to characterize the heat equation $\partial u/\partial t = \Delta u$ as the unique scale space that is both linear and isotropic. A consequence for image processing is that linear smoothing and contrast-invariance are incompatible. (Recall that we showed in Section 3.1.1 that the heat equation was not contrast invariant. This is illustrated numerically in Figure 21.4.) At some level, this explains the coexistence of at least two different schools of image processing: contrast-invariant mathematical morphology on the one hand, and classical linear scale space on the other, which is essentially convolution with the Gaussian (Theorem 2.3).

Theorem 21.23. *Let $\{T_t\}$ be a translation-invariant, causal, isotropic and linear scale space on $\mathcal{F}$. Then $F(D^2u, Du, t) = c(t)\Delta u$, where $c(t) \geq 0$. If, in addition, F is assumed to be continuous in t, then up to a rescaling $t' = h(t)$, the function $u(t, \mathbf{x}) = T_t u_0(\mathbf{x})$ is a viscosity solution of the heat equation*

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = \Delta u(t, \mathbf{x}), \quad (t, \mathbf{x}) \in [0, \infty) \times \mathbb{R}^N. \tag{21.32}$$

Proof. Since the scale space is translation invariant, F does not depend on $\mathbf{x}$ (Proposition 21.14), and since the scale space commutes with the addition of constants, F does not depend on c (Proposition 21.11). Thus, $F(A, p, \mathbf{x}, c, t) = F(A, p, t)$. We know from Theorem 21.9 that $T_{t+h,t}u(\mathbf{x}) - u(\mathbf{x}) = hF(D^2u(\mathbf{x}), Du(\mathbf{x}), t) + o(h)$ for any $u \in C^2(\mathbb{R}^N)$. Since $T_{t+h,t}$ is linear, we have $T_{t+h,t}(ru + sv) = rT_{t+h,t}u + sT_{t+h,t}v$ for any $u, v \in C^2(\mathbb{R}^N)$ and $r, s \in \mathbb{R}$. This and Theorem 17.8 imply that

$$F(D^2(ru + sv), D(ru + sv), t) = rF(D^2u, Du, t) + sF(D^2v, Dv, t),$$

which means that F is linear in the argument u. (In what follows, we keep t fixed, and for convenience we write $F(A, p)$ rather than $F(A, p, t)$.) We can choose any values for D^2u, D^2v, Du, and Dv. Thus we have

$$F(rA + sA', rp + sp') = rF(A, p) + sF(A', p')$$

where A and A' are arbitrary symmetric matrices and p and p' are arbitrary vectors. From this, we see that $F(A,p) = F(A,0) + F(0,p)$. Now define F_1 and F_2 by $F_1(p) = F(0,p)$ and $F_2(A) = F(A,0)$; F_1 and F_2 are clearly linear. Using the assumption that the operators are isotropic $T_{t,s}$, we see from Lemma **??** that

$$F_1(p) + F_2(A) = F_1(Rp) + F_2(RAR'),$$

where R is any linear isometry of $\mathbb{R}^N$. Taking $A = 0$, this implies that $F_1(Rp) = F_1(p)$ for any linear isometry R. Since F_1 is linear, this implies that F_1 is a constant. Since by Proposition 21.11 $F(0,0,t) = 0$, we conclude that $F_1(p) = 0$. This proves that $F(A) = F(RAR')$, where A is an arbitrary $N \times N$ symmetric matrix and R is an arbitrary linear isometry.

Given any symmetric matrix A, there is a linear isometry R such that RAR' is diagonal whenever the coordinate system is orthogonal. Furthermore, any two diagonalizations differ only in the arrangement of the diagonal entries, which are the N eigenvalues of A, and any arrangement of these entries can be achieved by some linear isometry. This means that the value of $F(A)$ depends only on some symmetric function f of the eigenvalues $\lambda_1, \lambda_2, \dots \lambda_N$ of A. That means that $F(A) = F(f(\lambda_1, \lambda_2, \dots \lambda_N))$, where f is a symmetric function of its arguments. Since F is also linear, we have

$$F(f(r\lambda_1, r\lambda_2, \dots r\lambda_N)) = rF(f(\lambda_1, \lambda_2, \dots \lambda_N)).$$

Since the only linear symmetric function of N variables is, up to a multiplicative constant, the linear function, we see that

$$F(A) = c\,\mathrm{trace}(A)$$

for some constant c. Since F is nondecreasing in A (Lemma 21.8), c is nonnegative. We conclude that $F(D^2u, Du) = c\Delta u$. Remember that this argument has been made with a fixed t that was not written. Thus, our real conclusion is that $F(D^2u, Du, t) = c(t)\Delta u$, where c is a nonnegative function of t.

If we assume that F is continuous in t, then $t \mapsto c(t)$ is continuous. Then by Theorem 21.12, $u(t,\mathbf{x}) = T_t u_0(\mathbf{x})$ is a viscosity solution of

$$\frac{\partial u}{\partial t}(t,\mathbf{x}) = c(t)\Delta u(t,\mathbf{x}).$$

Finally, if we rescale using the function $t \mapsto t'$ defined by $\partial t' \partial t(t) = c(t)$, we have the heat equation $\partial u/\partial t' = \Delta u$. □

21.6 Exercises

Exercise 21.1. Let T_t be a causal scale space (Definition 21.7) and $h : \mathbb{R}^+ \to \mathbb{R}^+$ a C^2 increasing function. Prove that that $S_t = T_{h(t)}$ also is a causal scale space. Assume that T_t is scale invariant and let $t'(t,\lambda)$ its rescaling function. Compute t' for the new scale space S_t. ■

Exercise 21.2. Consider the extrema killer T_t defined in section 7.4, where t denotes the area threshold. Show that the family $\{T_t\}$ is pyramidal and satisfies the global comparison principle, but that it does not satisfy the local comparison principle. Show that the family is, however, regular at $t = 0$ and, more precisely, that

$$F(A, p, 0) = 0 \quad \text{if } p \neq 0.$$

Check the other invariance properties of the extrema killer : prove in particular that it is affine invariant and compute $t'(t, \lambda)$ (Definition **??**). ■

Exercise 21.3. Let g be an integrable continuous function and for $u \in \mathcal{F}$, $Tu = g * u$. Prove that T is translation invariant and isotropic. ■

Exercise 21.4. Define $\{T_t\}$ by $T_t u_0 = g_t * u_0$, where $g_t(\mathbf{x}) = \frac{1}{t^2} g(\frac{\mathbf{x}}{t})$. Prove that $\{T_t\}$ is scale invariant (Definition **??**) and compute the function $t'(t, \lambda)$. Same questions if we set $g_t(\mathbf{x}) = \frac{1}{t} g(\frac{\mathbf{x}}{t^{\frac{1}{2}}})$. ■

Exercise 21.5. Check that Definition **??** is valid for the classical scale spaces we already know: For the morphological operators, dilation and erosion, show that $t'(t, \lambda) = \lambda t$, no matter what the structuring element B is. Prove that these operators are not affine invariant. For the heat equation and mean curvature motion, check that $t'(t, \lambda) = \lambda^2 t$. ■

21.7 Comments and references

The presentation in this chapter and the next one follows essentially the work of L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel in [11], [9], and [10]. Their stated objective was to "... describe all multiscale causal, local, stable and shape preserving smoothing operators. This classification contains the classical 'morphological' operators, and some new ones." This axiomatic approach is presented in several survey papers, with increasingly simple sets of axioms: Lions [184]; Alvarez and Morel [14]; Guichard, Lopez, and Morel [131]; and Guichard's doctoral thesis [130], which was an early version of this book.

Linear scale space. Scale space theory was founded (in the linear framework) by Witkin [289], Marr [198], and Koenderink. An earlier development of linear scale space has been traced to Japan in [282, 283]. Many works by Florack, ter Haar Romeny, Koenderink, and Viergever focus on the computation of partial derivatives of any order of an image and their use in image analysis [111, 112, 113, 115]. The concept of *causality*, used by all of these authors is crucial; it has been reinterpreted in this chapter as the combination of two requirements: a pyramidal structure and a comparison principle. De Giorgi founded his mathematical theory of barriers for geometric motions on similar principles [124]. There are many axiomatic characterizations of linear scale space in terms of causality, invariants, and conservation properties. We mention particularly the early work by Babaud, Witkin, Baudin, and Duda [28] and Hummel [150]. A slight relaxation of the initial axioms led Pauwels and others to discover other possible linear scale spaces, which, however, are less local [228]. There have also been several attempts to define nonlinear scale spaces, which are understood as nonlinear invariant families of smoothing operators. In mathematical morphology, we mention work by Chen [68], Toet [270, 271], and Jackway [158]; Jackway emphasized the scale space properties of multiscale erosions and dilations. After the publication of [11] by Alvarez, Guichard, Lions, and Morel, several different axiomatic approaches have been proposed for nonlinear scale spaces. Weickert insists on grey level conservation, which excludes all of the mathematical morphology operators, and proposes a line of conservative parabolic nonlinear PDEs [279, 280]. The axiomatic presentation of Olver, Sapiro, and Tannenbaum [220] deduces the various scale spaces as invariant heat flows. See also [217]; the book [267] contains miscellaneous contributions to geometric diffusion.

Extensions. Caselles, Coll, and Morel have questioned the very soundness of applying any of the proposed scale spaces to natural images [59]. They argue following the Kanisza psychophysical theory that occlusions generate T-junctions in images and that these T-junctions should be detected before any smoothing is applied (see Figure 21.6.) In [60], the same authors propose the set of level lines of the image, the so-called topographical map, as an alternative multiscale structure for describing images. In another direction, Geraets and others proposed a generalization of scale space to discrete point sets [121].

Contrast invariance. The Wertheimer principle, which states that human visual perception is independent of changes in illumination, was enunciated in 1923 [287]. Contrast invariance appears in mathematical morphology in the work of Serra [255]. Koenderink and van Doorn emphasized this requirement and introduced photometric invariants [173]. Florack and others studied contrast-invariant differential operators in [114]. Romeny and others construct third-order contrast-invariant operators to detect T-junctions [269]. See also [234]. The significance of contrast invariance for smoothing T-junctions is illustrated in Figure 21.6.

Rotation and scale invariance. One of the first discussions of rotation-invariant image operators was given by Beaudet in [38]. See also Lenz [181] for work on rotation-invariant operators. Scale-invariant shape representation is discussed by Baldwin, Geiger, and Hummel in [29]. Alvarez, Gousseau, and Morel use numerical experiments on natural images to confirm their scale invariance [7].

Affine invariance. Affine invariants are viewed as approximate projective invariants by Chang in [73]. The importance of affine invariance for three-dimensional object recognition is discussed in [29] and [181]. Work by Forsyth, Munday, and Zisserman has been fundamental and has launched wide-ranging discussion of this theme [116, 209, 210]. Further contributions to the use and computation of affine and projective differential invariants in image processing can be found in [41], [244], [274], and [285].

Figure 21.4: The heat equation 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 21.5: 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 a finite difference scheme. 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 this with the same experiment performed with the linear scale space, Figure 21.4.

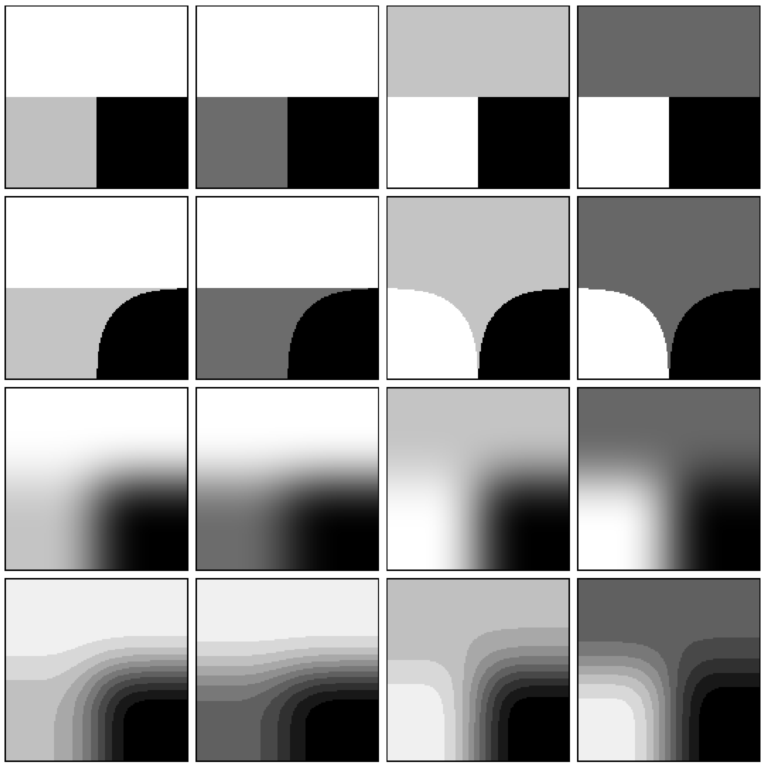

Figure 21.6: Same geometric figures, different evolutions under smoothing. 1st row: The four figures have T-junctions that differ in the way grey levels are distributed among the three regions. In the first two figures, the grey levels are monotone in, say, the clockwise direction. This means that they differ by a monotone contrast change. The same is true for the second two figures. However, the first and third figures differ by a nonmonotone contrast change.
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 of grey levels, then the geometric evolution is identical. If, instead, a nonmonotone contrast change has been applied, the evolutions are geometrically different.
3rd row: result of a smoothing by the linear scale space. All four T-junctions give different evolutions. The evolution depends on the gray-level values of the three level sets, rather than depending only on their order.
4rd row: quantization of the 3rd row to display the shapes of some level lines.

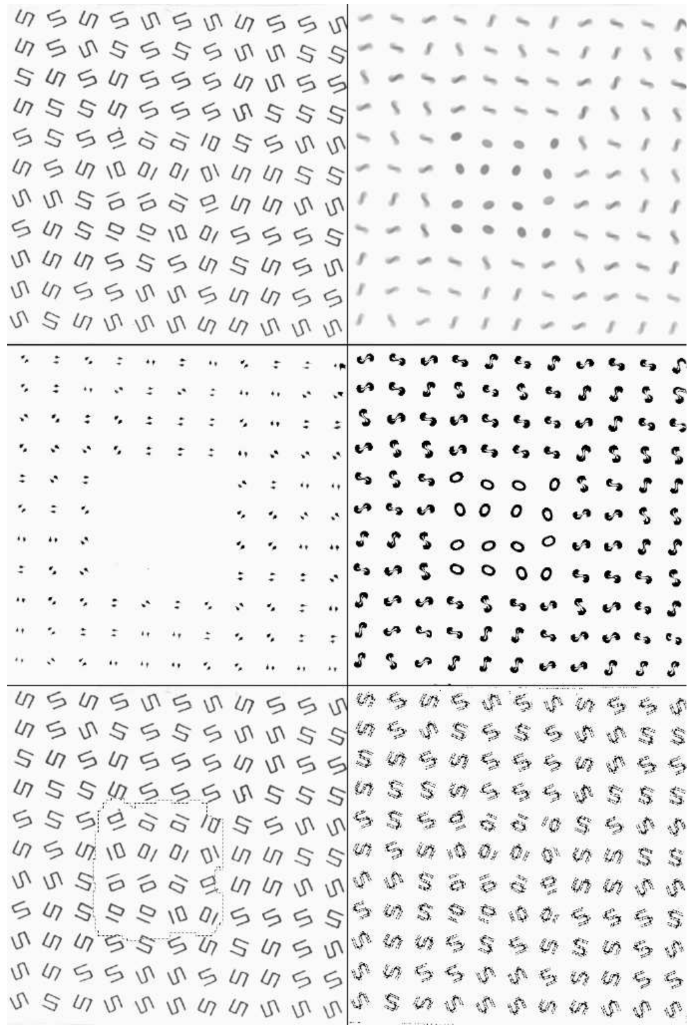

Figure 21.7: Hyperdiscrimination of textures by nonlinear scale space. According to the Julesz theory of textons, human perception can discriminate different textures if their average behavior in terms of "texton" density is different. As shown in its mathematical formalization, proposed by C. Lopez, 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 top to bottom: 1-an original texture pair that is preattentively undiscriminable. The central square of the image consists of rotated "10's" and the rest of the image of rotated "S's." These patterns are different, but have the same number of bars, angles, and so forth. 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 multi-image 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 easily discriminates between the two textures.

Chapter 22

The Contrast-Invariant and Affine-Invariant Scale Spaces

This chapter is a direct continuation of Chapter 21. We are going to characterize, up to a multiplicative constant, all of the contrast-invariant scale spaces as *curvature evolution equations*. In the interest of simplicity and clarity, we will first prove the result in two dimensions. The computations are more intuitive in this case, and they are easily displayed in detail. Then we shall obtain the (AMSS) equation as the unique contrast invariant, affine invariant self-dual scale space. This result is then generalized to any dimension in section **??**, where we find again a unique contrast and affine invariant self-dual scale space. This result also yields an impossibility : no further invariance requirement is possible. In particular, a causal, contrast and projective invariant scale-space is impossible.

22.1 The two-dimensional case

We will show that if a scale space $\{T_t\}$ is causal, isometric, and contrast invariant, then the associated PDE is of the form

$$\frac{\partial u}{\partial t} = |Du| G(\mathrm{curv}(u), t). \tag{22.1}$$

This does not tell us much about G, so the question is, What additional assumptions must be made to have a more specific characterization of G? One answer is this: If we assume that $\{T_t\}$ is affine invariant and that $T_t(-u) = -T_t u$ (which we call reverse contrast invariance or self-duality), then there is only one equation that satisfies all of these conditions, namely, the so-called affine morphological scale space (AMSS),

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

We are led by Theorem 21.9 to study scale spaces defined by PDEs of the form

$$\frac{\partial u}{\partial t} = F(D^2 u, Du, u, \mathbf{x}, t), \quad u(0) = u_0,$$

where u_0 is the original image, $u(t, \cdot)$ is the image smoothed at scale t, and $F(A, p, c, \mathbf{x}, t)$ is the function associated with $\{T_t\}$. In the two-dimensional case, A is a 2×2 symmetric matrix, p is a two-dimensional vector, c is a constant, $\mathbf{x}$ is a point in the plane, and $t \in \mathbb{R}^+$ is the scale. We will be using the following results from Chapter 21:

- If $\{T_t\}$ is translation invariant, then F does not depend on $\mathbf{x}$ (Proposition 21.14).
- It $\{T_t\}$ commutes with the addition of constants, then F does not depend on c (Proposition 21.11).
- If $\{T_t\}$ is isotropic, then $F(RAR', Rp, t) = F(A, p, t)$ for every $R \in O_2$ (Lemma **??**).
- It $\{T_t\}$ is contrast invariant, then $F(\mu A + \lambda(p \otimes p), \mu p, t) = \mu F(A, p, t)$, for any real numbers λ and μ, $\mu > 0$, any 2×2 symmetric matrix A, and any two-dimensional vector p (Lemma 21.18). Recall that the tensor product $p \otimes p$ is just the symmetric matrix $\{p_i p_j\}$, $i, j \in \{1, 2\}$:

$$p \otimes p = \begin{bmatrix} p_1^2 & p_1 p_2 \\ p_1 p_2 & p_2^2 \end{bmatrix}.$$

Relations $F(RAR', Rp, t) = F(A, p, t)$ and $F(\mu A + \lambda(p \otimes p), \mu p, t) = \mu F(A, p, t)$ will be used to show that F depends on two real functions $\tilde{a}_{12}$ and $\tilde{a}_{22}$ of A and p. These functions are defined by considering the rotation represented by the matrix

$$R_p = \frac{1}{|p|} \begin{bmatrix} p_1 & p_2 \\ -p_2 & p_1 \end{bmatrix}.$$

R_p has been chosen to map the unit vector $p/|p|$ onto the unit vector $e_1 = (1, 0)$:

$$R_p p = |p| e_1. \tag{22.2}$$

The functions $\tilde{a}_{ij}$, $i, j \in \{1, 2\}$, are defined by

$$\frac{1}{|p|} R_p A R_p' = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{bmatrix}. \tag{22.3}$$

A straightforward computation shows that

$$\tilde{a}_{12} = \frac{1}{|p|^3}((p_1^2 - p_2^2)a_{12} + p_1 p_2 (a_{22} - a_{11})) = \frac{A(p, p^\perp)}{|p|^3}, \tag{22.4}$$

$$\tilde{a}_{22} = \frac{1}{|p|^3}(a_{11} p_2^2 - 2a_{12} p_1 p_2 + a_{22} p_1^2) = \frac{A(p^\perp, p^\perp)}{|p|^3}. \tag{22.5}$$

We should keep in mind that A represents $D^2 u$ and p represents Du so these results and the calculations that follow, while purely algebraic, have interpretations in differential geometry. In particular,

$$\tilde{a}_{22}(D^2 u, Du) = \operatorname{div}\Big(\frac{Du}{|Du|}\Big) = \operatorname{curv}(u),$$

$$\tilde{a}_{12}(D^2 u, Du) = \operatorname{div}\Big(\frac{Du^\perp}{|Du|}\Big) = \operatorname{anticurv}(u).$$

Both differential operators are contrast invariant (see Exercise 22.2, while $\tilde{a}_{11}$ is not: We know that it is related to the Haralick edge detector. This gives the meaning of the next lemma.

Lemma 22.1. *If F satisfies the relations*

$$F(RAR', Rp, t) = F(A, p, t), \tag{22.6}$$
$$F(\mu A + \lambda(p \otimes p), \mu p, t) = \mu F(A, p, t), \tag{22.7}$$

then there is a function G of three real variables such that, for $p \neq 0$,

$$F(A, p, t) = |p| G(\tilde{a}_{12}, \tilde{a}_{22}, t). \tag{22.8}$$

Proof. We first use (22.7) with $\mu = 1/|p|$. Thus,

$$F(A, p, t) = |p| F\Big(\frac{A}{|p|} + \lambda(p \otimes p), \frac{p}{|p|}, t\Big),$$

where λ is any real number. Next, we apply (22.6) with $R = R_p$:

$$\begin{aligned} F(A, p, t) &= |p| F\Big(R_p\Big(\frac{A}{|p|} + \lambda(p \otimes p)\Big)R_p', R_p\frac{p}{|p|}, t\Big) \\ &= |p| F\Big(R_p\Big(\frac{A}{|p|}\Big)R_p' + \lambda R_p(p \otimes p)R_p', e_1, t\Big). \end{aligned}$$

It is easily checked that

$$R_p(p \otimes p)R_p' = \begin{bmatrix} |p|^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus, by (22.3),

$$F(A, p, t) = |p| F\left(\begin{bmatrix} \tilde{a}_{11} + \lambda |p|^2 & \tilde{a}_{12} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{bmatrix}, e_1, t \right).$$

Since λ is arbitrary and since $p \neq 0$, F depends only on $\tilde{a}_{12}$ and $\tilde{a}_{22}$. To finish, we can define G by

$$G(\tilde{a}_{12}, \tilde{a}_{22}, t) = F\left(\begin{bmatrix} 0 & \tilde{a}_{12} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{bmatrix}, e_1, t \right). \qquad \square$$

$\square$

Lemma 22.2. *The function G depends only on $\tilde{a}_{22}$ and t.*

Proof. We will use the fact established in Lemma 21.8 that F is nondecreasing with respect to its first argument and the assumption in Definition 21.6 that F is continuous in this argument. We will also use the result of Lemma 22.1. The intuitive argument here is that $\tilde{a}_{22}(A, p)$ is nondecreasing function of A, while $\tilde{a}_{12}(A, p)$ is not (see Exercise 22.2.)

Consider two symmetric matrices A and B such that $A \geq B$. In analogy with (22.3), we write

$$\frac{1}{|p|} R_P B R_p' = \begin{bmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{12} & \tilde{b}_{22} \end{bmatrix}. \tag{22.9}$$

Then

$$\frac{1}{|p|} R_p (A - B) R_p' = \begin{bmatrix} \tilde{a}_{11} - \tilde{b}_{11} & \tilde{a}_{12} - \tilde{b}_{12} \\ \tilde{a}_{12} - \tilde{b}_{12} & \tilde{a}_{22} - \tilde{b}_{22} \end{bmatrix}. \tag{22.10}$$

By assumption, the symmetric matrix $A - B$ is such that $A - B \geq 0$. This property is invariant under rotation, in particular, under the mapping $A - B \mapsto R_p(A-B)R_p'$. Hence, $A - B \geq 0$ if and only if $R_p(A-B)R_p' \geq 0$. This means that $A \geq B$ if and only if

$$(\tilde{a}_{11} - \tilde{b}_{11})x^2 + 2(\tilde{a}_{12} - \tilde{b}_{12})xy + (\tilde{a}_{22} - \tilde{b}_{22})y^2 \geq 0 \tag{22.11}$$

for all real numbers x and y, and this is true if and only if

$$(\tilde{a}_{11} - \tilde{b}_{11})(\tilde{a}_{22} - \tilde{b}_{22}) \geq (\tilde{a}_{12} - \tilde{b}_{12})^2. \tag{22.12}$$

Fix A and choose an arbitrary $\tilde{b}_{12}$, $\tilde{b}_{12} \neq \tilde{a}_{12}$. Choose any real number $\tilde{b}_{22}$ such that $\tilde{a}_{22} - \tilde{b}_{22} = \varepsilon > 0$. Now select $\tilde{b}_{11} = \tilde{b}_{11}(\varepsilon)$ so that (22.12) is satisfied. Then we have

$$F\left(\begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{bmatrix}, e_1, t \right) \geq F\left(\begin{bmatrix} \tilde{b}_{11}(\varepsilon) & \tilde{b}_{12} \\ \tilde{b}_{12} & \tilde{a}_{22} - \varepsilon \end{bmatrix}, e_1, t \right) = F\left(\begin{bmatrix} 0 & \tilde{b}_{12} \\ \tilde{b}_{12} & \tilde{a}_{22} - \varepsilon \end{bmatrix}, e_1, t \right).$$

Indeed, the value on the right-hand side of the inequality is independent of $\tilde{b}_{11}(\varepsilon)$ (Lemma 22.1). Now let ε tend to zero. By the continuity of F in its first argument, we conclude that

$$F\left(\begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{bmatrix}, e_1, t \right) \geq F\left(\begin{bmatrix} 0 & \tilde{b}_{12} \\ \tilde{b}_{12} & \tilde{a}_{22} \end{bmatrix}, e_1, t \right).$$

This shows that

$$G(\tilde{a}_{12}, \tilde{a}_{22}, t) \geq G(\tilde{b}_{12}, \tilde{a}_{22}, t).$$

A similar argument shows that

$$G(\tilde{b}_{12}, \tilde{a}_{22}, t) \geq G(\tilde{a}_{12}, \tilde{a}_{22}, t).$$

We conclude that $G(\tilde{a}_{12}, \tilde{a}_{22}, t)$ does not depend on $\tilde{a}_{12}$, and hence that G is a function of only $\tilde{a}_{22}$ and t. □

We summarize these last results in the following theorem.

Theorem 22.3. *If the two-dimensional scale space $\{T_t\}$ is causal, isometric, and contrast invariant, then its associated PDE has the form*

$$\frac{\partial u}{\partial t} = |Du| G(\mathrm{curv}(u), t), \tag{22.13}$$

where G is continuous and nondecreasing in its first variable.

We are now going to introduce scale and affine invariance. We are still working in two dimensions.

Theorem 22.4. *Assume that the scale space $\{T_t\}$ is causal, isometric, and contrast invariant. In addition, assume that it is scale invariant and that it is normalized according to Lemma 21.21. Then its associated PDE has the form*

$$\frac{\partial u}{\partial t} = |Du|\beta(t\mathrm{curv}(u)), \tag{22.14}$$

where β is continuous and nondecreasing.

If the scale space $\{T_t\}$ is affine invariant and normalized according to Lemma 21.21, then the associated PDE has the form

$$\frac{\partial u}{\partial t} = |Du|\beta(t\mathrm{curv}(u)), \tag{22.15}$$

where $\beta(s) = Cs^{1/3}$ if $s > 0$ and $\beta(s) = -D|s|^{1/3}$ if $s < 0$, for two nonnegative constants C and D. Conversely, this equation defines an affine-invariant scale space.

Proof. By Lemma **??**, if a causal scale space is affine invariant, then, after appropriate renormalization (Lemma 21.21), its associated function satisfies

$$F(BAB', Bp, t) = |\mathrm{det}B|^{1/2}F(A, p, |\mathrm{det}B|^{1/2}t) \tag{22.16}$$

for any linear map B. If we let $B = cI$, $c > 0$, then $F(c^2A, cp, t) = cF(A, p, ct)$. Since $F(A, p, t) = |p|G(\tilde{a}_{22}(A, p), t)$, this implies that

$$c|p|G(\tilde{a}_{22}(c^2A, cp), t) = c|p|G(\tilde{a}_{22}(A, p), ct),$$

and since

$$\tilde{a}_{22}(c^2A, cp) = \frac{c^2A(cp^\perp, cp^\perp)}{|cp|^3} = c\tilde{a}_{22}(A, p),$$

we see that

$$G(\tilde{a}_{22}(A, p), ct) = G(c\tilde{a}_{22}(A, p), t).$$

Since this equation is true for all A, $p \neq 0$, $c > 0$, and $t > 0$, $G(cs, t) = G(s, ct)$ for any s and any positive c and t. This implies that

$$G(s, t) = G(st, 1) = \beta(st),$$

where β is continuous and nondecreasing. This proves the first part of the theorem.

We now assume that the scale space is affine invariant. To identify the power $1/3$, we need to exploit the affine invariance. We shall do it by "stretching and shrinking" along the x and y axes, that is, by using the transformation represented by

$$B(\lambda) = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{bmatrix}.$$

First note that

$$BAB' = \begin{bmatrix} \lambda^2 a_{11} & a_{12} \\ a_{12} & \lambda^{-2}a_{22} \end{bmatrix} \quad \text{and} \quad Bp = (\lambda p_1, p_2/\lambda).$$

Then we see from (22.5) that

$$\tilde{a}_{22}(BAB', Bp) = \frac{a_{11}p_2^2 - 2a_{12}p_1p_2 + a_{22}p_1^2}{(\lambda^2 p_1^2 + \lambda^{-2}p_2^2)^{3/2}}. \tag{22.17}$$

We know that $F(A, p, t) = |p|\beta(t\tilde{a}_{22}(A, p, t))$. This and the affine-invariance relation (22.16) show that

$$|Bp|\beta(t\tilde{a}_{22}(BAB', Bp)) = |p|\beta(t\tilde{a}_{22}(A, p)). \tag{22.18}$$

If we let $p_1 = 1$, $p_2 = 0$, and $a_{22} = 1$, then from (22.17) and (22.18) it follows that

$$|\lambda|\beta(s/\lambda^3) = \beta(s).$$

The first thing to notice is that $\beta(0) = 0$, which is consistent with (and a consequence of) the fact that $F(0, 0, t) = 0$. On the other hand, nothing we have assumed precludes $\beta(s) = 0$ for all $s \geq 0$. However, in case $\beta(a) > 0$ for some $a > 0$, we can select $\lambda > 0$ so $\lambda^3 = s/a$, and we have

$$\beta(s) = a^{-1/3}\beta(a)s^{1/3} = Cs^{\frac{1}{3}}.$$

A similar argument shows that either $\beta(s) = 0$ for all $s \leq 0$, or

$$\beta(s) = |b|^{-1/3}\beta(b)|s|^{1/3} = -D|s|^{\frac{1}{3}}.$$

□

In general, $\beta(1) \neq -\beta(-1)$, that is, $C \neq D$. For example, if $D = 0$ and $C > 0$, then we have a pure affine erosion (shapes shrink); if $C = 0$, then we have a pure affine dilation (shapes expand).

Corollary 22.5. *If, in addition to the assumptions of Theorem 22.4, the scale space is reverse contrast invariant, which means that $T_{t+h,t} \circ g = g \circ T_{t+h,t}$ for continuous nonincreasing g, or that it is self-dual, $T(u) = -T(-u)$, then $D = C$.*

Proof. The reverse contrast invariance is equivalent to $T(u) = -T(u)$ plus the contrast invariance. The proof is an obvious adaptation of the proof of Lemma 21.18. If the scale space $\{T_t\}$ is reverse contrast invariant, then equation (17.16) is true for negative (as well as positive) μ, and we have $F(-A, -p, t) = -F(A, p, t)$. This then implies that $\beta(-1) = -\beta(1)$. □

22.2 Contrast-invariant scale space equations in N dimensions

We are going to extend the results of Theorem 22.4 to N dimensions, so we shall make the same assumptions about the scale space $\{T_t\}$: It is causal, it commutes with the addition of constants, and it is translation invariant. Our immediate aim is to deduce the general form of F in N dimensions from the assumption that $\{T_t\}$ is contrast invariant and isotropic. In the interest of notation, we will suppress t in the following discussion.

We know from Lemma 21.18 that the function F associated with a contrast-invariant scale space $\{T_t\}$ satisfies the relation

$$F(\mu A + \lambda(p \otimes p), \mu p) = \mu F(A, p) \tag{22.19}$$

for all $\lambda \in \mathbb{R}$, all $\mu \geq 0$, any $N \times N$ symmetric matrix $A \in S^N$, and any vector $p \in \mathbb{R}^N$. By taking $\lambda = 0$, this shows that F is positively homogeneous in (A, p):

$$F(\mu A, \mu p) = \mu F(A, p). \tag{22.20}$$

In particular, $F(0, 0) = 0$, which we also know from Proposition 21.11.

If we take $\mu = 1$, then (22.19) becomes $F(A + \lambda(p \otimes p), p) = F(A, p)$. If $N = 1$, this means that F depends only on $p \in \mathbb{R}$, and we conclude that

$$F(A, p) = \begin{cases} F(1)p & \text{if } p \geq 0, \\ -F(-1)p & \text{if } p \leq 0. \end{cases} \tag{22.21}$$

A more interesting situation occurs if $N \geq 2$, as we have already seen in case $N = 2$. From now on we assume that $N \geq 2$.

We need to introduce some notation. For $p \in \mathbb{R}^N$, $p \neq 0$, consider the linear operator defined by the matrix $Q_p = I_N - (1/|p|^2)(p \otimes p)$. It is easy to verify that Q_p is the projection of $\mathbb{R}^N$ onto the hyperplane $p^\perp$ (also denoted by $(\mathbb{R}p)^\perp$). Let A be an $N \times N$ symmetric matrix and consider the matrix Q_pAQ_p. Since Q_p is symmetric, it is clear that Q_pAQ_p is also symmetric. It is also clear that $q \in (\mathbb{R}p)^\perp$ implies that $Q_pAQ_pq \in (\mathbb{R}p)^\perp$ and that $q \in \mathbb{R}p$ implies that $Q_pAQ_pq = 0$. Since Q_pAQ_p is symmetric, it has N real eigenvalues, one of which we have just seen to be zero. Let $\mu_1, \mu_2 \dots, \mu_{N-1}$ denote the $N - 1$ other eigenvalues. These are the eigenvalues of Q_pAQ_p restricted to $(\mathbb{R}p)^\perp$. If $A = D^2u$ and $p = Du$, and if we define $\kappa_i = \mu_i/|p|$, $1 \leq i \leq N - 1$, then the κ_i are the principal curvatures of the level hypersurface of u (Definition 11.19). If $N = 2$, then by Definition 11.14, $\kappa_1 = (1/|p|)\operatorname{trace}(Q_pAQ_p) = \operatorname{curv}(u)$.

Theorem 22.6 (Giga, Goto []). *Let $\{T_t\}$ be a contrast-invariant scale space and assume that $N \geq 2$. Then the associated function F satisfies the following relation: For all $A \in S^N$ and all $p \in \mathbb{R}^N$, $p \neq 0$,*

$$F(A, p, t) = F(Q_pAQ_p, p, t). \tag{22.22}$$

Proof. We begin by fixing $p \in \mathbb{R}^N$, $p \neq 0$ and selecting an orthogonal coordinate system such that $p = |p|(0, \dots, 0, 1)$. Then $p \otimes p = |p|^2(\delta_{Ni}\delta_{Nj})$, $1 \leq i, j \leq N$. If we write $B = A + \lambda(p \otimes p)$, then $b_{ij} = a_{ij} + \lambda|p|^2(\delta_{Ni}\delta_{Nj})$. This means that $b_{ij} = a_{ij}$ except for $i = N$ and $j = N$, in which case $b_{NN} = a_{NN} + \lambda|p|^2$. Since $F(A + \lambda(p \otimes p), p) = F(A, p)$, this means that $F(A, p)$ does not depend on a_{NN}. We use this fact, combined with the assumption that F is nondecreasing in its first variable, to complete the proof.

Note that, with the coordinate system we have chosen, Q_pAQ_p is just the matrix A with the last column and last row replaced with zeros: If $C = Q_pAQ_p$, then $c_{ij} = a_{ij}$ for $1 \leq i, j \leq N - 1$ and $c_{ij} = 0$ if $i = N$ or if $j = N$. Now let $M = a^2_{N,1} + \dots + a^2_{N,N-1}$ and consider $I_\varepsilon = \varepsilon I + (M/\varepsilon - \varepsilon)(\delta_{Ni}\delta_{Nj})$,

$1 \le i, j \le N$. Thus, I_ε is an $N \times N$ diagonal matrix D, where $d_{ii} = \varepsilon$ for $1 \le i \le N-1$ and $d_{NN} = M/\varepsilon$. One can easily verify that $Q_pAQ_p \le A + I_\varepsilon$ and that $A \le Q_pAQ_p + I_\varepsilon$ for small $\varepsilon > 0$. Then we have

$$F(A,p) \le F(Q_pAQ_p + I_\varepsilon, p) \le F(A + 2I_\varepsilon, p).$$

As we let $\varepsilon \to 0$, the entries in the matrix $A + 2I_\varepsilon$ tend to a_{ij} except the entry $a_{NN} + 2M/\varepsilon$, which tends to $+\infty$. But F is independent of the value of its (N,N)-entry, and F is continuous in its first variable. Thus,

$$F(A + 2I_\varepsilon, p) \to F(A,p) \quad \text{and} \quad F(Q_pAQ_p + I_\varepsilon, p) \to F(Q_pAQ_p, p). \qquad \square$$

$\square$

Corollary 22.7. *Let $\{T_t\}$ be a contrast-invariant scale space and assume that $N \ge 2$. If $\{T_t\}$ is also isometric, then*

$$F(A,p,t) = |p| G(\kappa_1, \dots, \kappa_{N-1}, t) \tag{22.23}$$

for all $A \in S^N$, $p \in \mathbb{R}$, $p \ne 0$, where G is a continuous function on $\mathbb{R}^{N-1}$ that is symmetric in the $N-1$ variables $\kappa_1, \dots, \kappa_{N-1}$ and is nondecreasing with respect to each κ_i, $1 \le i \le N-1$.

Proof. By Lemma **??**, the function F associated with the scale space satisfies

$$F(RAR', Rp) = F(A,p) \tag{22.24}$$

for all $A \in S^N$, $p \in \mathbb{R}^N$, $p \ne 0$, and $R \in O_N$. (Recall that O_N denotes the group of linear isometries of $\mathbb{R}^N$ (section 17.4.1).)

Fix $p \ne 0$ and let R be any element of the subgroup O_N^p of O_N that leaves p fixed, that is, $Rp = p$. Since $Q_pAQ_p \in S^N$, we know from (22.24) that

$$F(RQ_pAQ_pR', Rp) = F(RQ_pAQ_pR', p) = F(Q_pAQ_p, p).$$

By Theorem 18.5, $F(Q_pAQ_p, p) = F(A,p)$. These two relations tell us that

$$F(RQ_pAQ_pR', p) = F(A,p)$$

for all $R \in O_N^p$. This means that the value of $F(A,p)$ depends only on p and the eigenvalues of Q_pAQ_p. Indeed, using the same coordinate system used in the proof of Theorem 18.5 based on $p = |p|(0, \dots, 1)$, there is always an $R \in O_N^p$ so that RQ_pAQ_pR' is diagonal with entries $\mu_1, \mu_2, \dots, \mu_{N-1}, 0$, where the μ_i are the eigenvalues of Q_pAQ_p restricted to $(\mathbb{R}p)^\perp$. Furthermore, we can choose R so that the μ_i appear in any order. Thus, there is a function G_1 such that

$$F(A,p) = G_1(\mu_1, \mu_2, \dots, \mu_{N-1}, p);$$

G_1 is continuous and symmetric in the μ_i, and G_1 is nondecreasing in each μ_i. (These last statements follow from the fact that F is continuous and nondecreasing in its first variable.)

Now take $R \in O_N$ and let $q = Rp$ (p is still fixed). Then $|p| = |q|$. Furthermore, given any $q \in \mathbb{R}^N$ such that $|p| = |q|$, there is an $R \in O_N$ such that

$Rp = q$. It is easy to verify that $R(p \otimes p) = q \otimes p$, and since $(q \otimes p)' = p \otimes q$, that $(p \otimes p)R' = p \otimes q$. Thus, $R(p \otimes p)R' = q \otimes q$, which implies that $RQ_p = Q_q R$. Consequently,

$$Q_q RAR' Q_q = RQ_p AQ_p R',$$

and this means that $Q_q RAR'Q_q$ and $Q_p AQ_p$ have the same eigenvalues. Using (22.24) again, we see that $F(RQ_pAQ_pR', Rp) = F(Q_pAQ_p, p)$, and since $Rp = q$,

$$F(RQ_pAQ_pR', q) = F(Q_pAQ_p, p).$$

Since $Q_qRAR'Q_q$ and Q_pAQ_p have the same eigenvalues, this implies that

$$F(A, p) = G_1(\mu_1, \mu_2, \ldots, \mu_{N-1}, p) = G_1(\mu_1, \mu_2, \ldots, \mu_{N-1}, q)$$

whenever $|p| = |q|$. This means that F depends only on the modulus of p, and therefore we can write $F(A, p) = G_1(\mu_1, \mu_2, \ldots, \mu_{N-1}, |p|)$. Since F is homogeneous,

$$G_1(\mu\mu_1, \mu\mu_2, \ldots, \mu\mu_{N-1}, \mu|p|) = \mu G_1(\mu_1, \mu_2, \ldots, \mu_{N-1}, |p|)$$

for $\mu \geq 0$. If we take $\mu = |p|^{-1}$, then

$$F(A, p) = |p| G_1(\mu_1/|p|, \mu_2/|p|, \ldots, \mu_{N-1}/|p|, 1).$$

Defining G by $G(\kappa_1, \kappa_2, \ldots, \kappa_{N-1}) = G_1(\mu_1/|p|, \mu_2/|p|, \ldots, \mu_{N-1}/|p|, 1)$, where $\kappa_i = \mu_i/|p|$, completes the proof. □

22.3 Affine-invariant scale spaces for $N \geq 2$

There is a function H_N in the following theorem that is defined on the set of N integers $\{-N+1+2k \mid 0 \leq k \leq N-1\}$. We will see in the proof of the theorem that H_N is nondecreasing and that it vanishes except at the points $-(N-1)$ and $N-1$. It will also be shown that $H_N(N-1) \geq 0$. There is not enough information to determine the value of $H_N(N-1)$; however, to avoid the trivial case $F \equiv 0$, we assume that $H_N(N-1) > 0$.

Theorem 22.8. *Assume that the scale space $\{T_t\}$ is contrast invariant and affine invariant. Assume also that $T_t(-u) = -T_t(u)$ and that $\{T_t\}$ has been normalized in accordance with Lemma 21.21. Then the PDE associated with $\{T_t\}$ is*

$$\frac{\partial u}{\partial t} = |Du| t^{\frac{N-1}{N+1}} \prod_{i=1}^{N-1} |\kappa_i|^{\frac{1}{N+1}} H_N\Big(\sum_{i=1}^{N-1} \operatorname{sgn}(\kappa_i)\Big), \tag{22.25}$$

where the κ_i are the principal curvatures of the level hypersurface of u, $\operatorname{sgn}(\kappa_i)$ denotes the sign of κ_i, and H_N is such that $H_N(N-1) > 0$, $H_N(N-1) = -H_N(-(N-1))$, and $H(n) = 0$ for all $-(N-1) < n < N-1$. In other words, H_N is equal to zero if all the κ_i do not have the same sign.

Proof. We begin with the result of Corollary 18.6: Since $\{T_t\}$ is contrast invariant and isometric, its associated function F is of the form

$$F(A, p, t) = |p| G(\kappa_1, \kappa_2, \ldots, \kappa_{N-1}, t),$$

and G is symmetric with respect to the κ_i. If $p \neq 0$, then $\kappa_i = \mu_i/|p|$, where the μ_i, $1 \leq i \leq N-1$ are the eigenvalues of A restricted to the hyperplane $p^\perp$ orthogonal to p. To simplify the proof, we prefer to use the more general form of F that appeared in the proof of Corollary 18.6, namely,

$$F(A, p, t) = G_1(\mu_1, \mu_2, \dots, \mu_{N-1}, |p|, t). \tag{22.26}$$

Since the restriction of A to the hyperplane $p^\perp$ is represented by a symmetric matrix, we can choose orthonormal vectors $e_1, \dots, e_{N-1}$ such that $\mu_i = A(e_i, e_i)$, $1 \leq i \leq N-1$. Each vector e_i is orthogonal to p, so we obtain an orthonormal basis for $\mathbb{R}^N$ by including the vector $e_N = p/|p|$. We now define some special linear affine transformations of $\mathbb{R}^N$: Let B_i, for $1 \leq i \leq N-1$, be the linear transform defined by

$$B_i(e_1, \dots, e_i, \dots, e_N) = (e_1, \dots, \beta e_i, \dots, \beta^{-1} e_N),$$

where $\beta \in \mathbb{R}$, $\beta \neq 0$. Clearly, $|B_i| = 1$. We are now going to apply the result of Lemma **??** with $B = B_1$. This says that $F(A, p, t) = F(B_1 A B_1', B_1 p, t)$, and in view of the representation (22.26), this means that

$$F(A, p, t) = G_1(\beta^2 \mu_1, \mu_2, \dots, \mu_{N-1}, \beta^{-1}|p|, t).$$

Assume for the moment that $\mu_1 \neq 0$ and take $\beta = |\mu_1|^{-1/2}$. Then

$$F(A, p, t) = G_1(\operatorname{sgn}(\mu_1), \mu_2, \dots, \mu_{N-1}, |\mu_1|^{1/2}|p|, t).$$

Repeat this argument with $B = B_i$ for $i = 2$ to $i = N-1$. The result is that

$$F(A, p, t) = G_1\Big(\operatorname{sgn}(\mu_1), \dots, \operatorname{sgn}(\mu_{N-1}), |p| \prod_{i=1}^{N-1} |\mu_i|^{1/2}, t\Big), \tag{22.27}$$

assuming that $\mu_i \neq 0$, $1 \leq i \leq N-1$. We return again to Lemma 17.21, which is based on the results of Lemma 17.20, and note that

$$F(\mu^2 A, \mu p, t) = \mu F(A, p, \mu t)$$

for $\mu \geq 0$ implies that

$$\begin{aligned} G_1\Big(\operatorname{sgn}(\mu_1), \dots, \operatorname{sgn}(\mu_{N-1}), \mu|p|\mu^{N-1} \prod_{i=1}^{N-1} |\mu_i|^{1/2}, t\Big) \\ = \mu G_1\Big(\operatorname{sgn}(\mu_1), \dots, \operatorname{sgn}(\mu_{N-1}), |p| \prod_{i=1}^{N-1} |\mu_i|^{1/2}, \mu t\Big). \end{aligned}$$

We deduce that

$$\begin{aligned} \mu^{-1} G_1\Big(\operatorname{sgn}(\mu_1), \dots, \operatorname{sgn}(\mu_{N-1}), \mu^N |p| \prod_{i=1}^{N-1} |\mu_i|^{1/2}, \mu^{-1} t\Big) \\ = G_1\Big(\operatorname{sgn}(\mu_1), \dots, \operatorname{sgn}(\mu_{N-1}), |p| \prod_{i=1}^{N-1} |\mu_i|^{1/2}, t\Big). \end{aligned}$$

By taking $\mu = t$, we see that

$$F(A,p,t) = t^{-1}G_1\Big(\mathrm{sgn}(\mu_1),\ldots,\mathrm{sgn}(\mu_{N-1}), t^N|p|\prod_{i=1}^{N-1}|\mu_i|^{1/2}, 1\Big),$$

and we write

$$F(A,p,t) = t^{-1}G_2\Big(t^N|p|\prod_{i=1}^{N-1}|\mu_i|^{1/2}, \mathrm{sgn}(\mu_1),\ldots,\mathrm{sgn}(\mu_{N-1})\Big).$$

At this point we use the fact that $\{T_t\}$ is contrast invariant, and so we have

$$F(\alpha^{-1}A, \alpha^{-1}p, t) = \alpha^{-1}F(A,p,t)$$

for all $\alpha > 0$. Thus,

$$tF(A,p,t) = \alpha G_2\Big(\alpha^{-1}t^N|p|\prod_{i=1}^{N-1}|\alpha^{-1}\mu_i|^{1/2}, \mathrm{sgn}(\mu_1),\ldots,\mathrm{sgn}(\mu_{N-1})\Big).$$

By taking $\alpha = (t^N|p|\prod_{i=1}^{N-1}|\mu_i|^{1/2})^{2/(N-1)}$, the function F is reduced to the following form:

$$\begin{aligned} F(A,p,t) &= t^{\frac{N-1}{N+1}}|p|^{\frac{2}{N+1}}\prod_{i=1}^{N-1}|\mu_i|^{\frac{1}{N+1}}G_2(1, \mathrm{sgn}(\mu_1),\ldots,\mathrm{sgn}(\mu_{N-1})) \\ &= t^{\frac{N-1}{N+1}}|p|^{\frac{2}{N+1}}\prod_{i=1}^{N-1}|\mu_i|^{\frac{1}{N+1}}H_1(\mathrm{sgn}(\mu_1),\ldots,\mathrm{sgn}(\mu_{N-1}). \end{aligned}$$

Since $\kappa_i = \mu_i/|p|$, we finally have

$$F(A,p,t) = |p|t^{\frac{N-1}{N+1}}\prod_{i=1}^{N-1}|\kappa_i|^{\frac{1}{N+1}}H_1(\mathrm{sgn}(\kappa_1),\ldots,\mathrm{sgn}(\kappa_{N-1}). \tag{22.28}$$

The derivation of this representation of F was based on the assumption that none of the eigenvalues μ_i were zero. If there were $\mu_i = 0$, we could have perturbed A by replacing these eigenvalues with $\varepsilon > 0$, done the derivation with positive eigenvalues, and then let ε tend to zero in (22.28).

We must now deal with the function H_1, or more precisely, the functions H_1, for there is a different function for each value of $N \geq 2$. First note that H_1 must be symmetric in its $N-1$ variables: F is invariant under any rotation that leaves p fixed, and we can always find an element of O_N^p that produces any given permutation of the μ_i, and thus of the κ_i. If some κ_i happens to be zero, we can set $H_1(\mathrm{sgn}(\kappa_1),\ldots,\mathrm{sgn}(\kappa_{N-1})) = 0$. Another way to do this is to define $\mathrm{sgn}(0) = 0$ and say that $H_1(\mathrm{sgn}(\kappa_1),\ldots,\mathrm{sgn}(\kappa_{N-1})) = 0$ if any of its $N-1$ variables is zero. Thus the only interesting situation is in case all of the κ_i are nonzero. Then by the invariance of H_1 under elements of O_N^p, it is clear that H_1 depends only on the number of variables equal to one and the number of variables equal to minus one. In other words, H_1 is a function of $\sum_{i=1}^{N-1}\mathrm{sgn}(\kappa_i)$:

$$H_1(\mathrm{sgn}(\kappa_1),\ldots,\mathrm{sgn}(\kappa_{N-1})) = H_N\Big(\sum_{i=1}^{N-1}\mathrm{sgn}(\kappa_i)\Big),$$

where we have written H_N to stress the fact that the functions depend on the dimension N. H_N is defined on the N integers $\{-N+1+2k \mid 0 \le k \le N-1\}$, but we no information about its range. We do, however, have one last invariant to call on, and that is the assumption that $T_i(u) = -T_i(-u)$. This translates into the relation $F(-A,-p,t) = -F(A,p,t)$, which in turn implies that $H_N(n) = -H_N(-n)$. Finally, we know that H_N must be nondecreasing based on the fact that F is nondecreasing in its first variable, A. In summary,

$$F(A,p,t) = |p|t^{\frac{N-1}{N+1}} \prod_{i=1}^{N-1} |\kappa_i|^{\frac{1}{N+1}} H_N\Big(\sum_{i=1}^{N-1} \operatorname{sgn}(\kappa_i)\Big), \tag{22.29}$$

where H_N is a nondecreasing function defined on the set $\{-N+1+2k \mid 0 \le k \le N-1\}$ and $H_N(n) = -H_N(-n)$. We are now going to consider three cases.

The case N= 2.

Here we have $N-1=1$, and (22.29) reads

$$F(A,p,t) = t^{1/3}|p||\kappa_1|^{1/3}\, H_2(\operatorname{sgn}(\kappa_1)).$$

The value of $H_2(1)$ is not determined, although it must be positive to avoid the trivial case $F \equiv 0$ and to ensure that $H_2(-1) = -H_2(1)$ and that H_2 is nondecreasing. Thus, up to a positive (multiplicative) constant $F(A,p,t) = t^{1/3}|p|\operatorname{sgn}(\kappa_1)|\kappa_1|^{1/3}$, which we have been writing as $F(A,p,t) = t^{1/3}|p|(\kappa_1)^{1/3}$, with the convention that $r^{1/3} = (r/|r|)|r|^{1/3}$. This can also be written as

$$F(A,p,t) = t^{1/3}|p|\left(\frac{1}{|p|}A\left(\frac{p^\perp}{|p|},\frac{p^\perp}{|p|}\right)\right)^{1/3} = t^{1/3}\left(A\left(p^\perp,p^\perp\right)\right)^{1/3}$$

If $A = D^2u$ and $p = Du$, then the associated scale space equation if

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

The case N= 3.

The PDE we obtain in this case is

$$\frac{\partial u}{\partial t} = t^{1/2}|Du||\kappa_1\kappa_2|^{1/4} H_3(\operatorname{sgn}(\kappa_1) + \operatorname{sgn}(\kappa_2)),$$

where κ_1 and κ_2 are the principal curvatures of the level surface of u. Their product is the Gaussian curvature of the level surface of u, and to highlight this we write the equation as

$$\frac{\partial u}{\partial t} = t^{1/2}|Du||G(u)|^{1/4} H_3(\operatorname{sgn}(\kappa_1) + \operatorname{sgn}(\kappa_2)). \tag{22.30}$$

Since $\operatorname{sgn}(\kappa_1) + \operatorname{sgn}(\kappa_2)$ takes only the values $-2, 0, 2$, and since $H_3(-2) = -H_3(-2)$, we are concerned with only two parameters: $H_3(2) = b$ and $H_3(0) = a$. We know that $H_3(-2) \le H_3(0) \le H_3(2)$, or $-b \le a \le b$. Hence, $b \ge 0$ and $|a| \le b$. We are now going to show that $a = 0$ by using the fact that F is increasing with respect to its first variable. We do this by choosing special values for the pair (κ_1,κ_2). For example, take $(-1, \alpha)$ and (α, α), $\alpha > 0$. The value of F for the first pair is less than or equal to the value of F for the second:

$$\alpha H_3(0) \leq \alpha^2 H_3(2).$$

Letting α tend to zero shows that $H_3(0) \leq 0$. Making a similar argument with the pairs $(-\alpha, \alpha)$ and (α, α), shows that $H_3(0) \geq 0$. Thus $H_3(0) = 0$, which means that the left-hand side of (22.30) is zero if the two principal curvatures have opposite signs. Consequently, up to a positive multiplicative constant, equation (22.30) is

$$\frac{\partial u}{\partial t} = \operatorname{sgn}(\kappa_1) t^{1/2} |Du| |G(u)^+|^{1/4}, \tag{22.31}$$

where x^+ stands for $\sup\{0, x\}$. This equation describes the unique multiscale analysis in three dimensions that is both affine invariant and contrast invariant and satisfies the condition $T_t(u) = -T_t(-u)$.

The case $N > 3$.

The only remaining task is to prove that H_N has the properties stated in the theorem. This is done by using arguments similar to those use for the three-dimensional case: By taking particular values for the κ_i and by using the fact that F is nondecreasing, one shows that $H_N(N-3) = 0$. Then since H_N is nondecreasing and since $H_N(n) = -H_N(-n)$, it follows that $H_N(n) = 0$ except for $n = N-1$ and $n = -(N-1)$. The details are left as an exercise. □

Exercise 22.1. Fill in the details for the last part of the proof. ■

22.4 Exercises

Exercise 22.2. The aim of the exercise is to give a geometric interpretation of anticurv(u) and to help interpreting the proofs of Lemmas 22.1 and lemma 22.2. We refer to the definitions of $\tilde{a}_{ij}$, $i = 1, 2$ given in Formulas (22.4)-(22.5).

1) Show that anticurv$(u)(\mathbf{x}) = \tilde{a}_{12}(D^2u(\mathbf{x}), Du(\mathbf{x}))$ is the curvature of the gradient line of u through the point $\mathbf{x}$. The gradient lines are the curves that are tangent to the gradient of u at every point. They form a system of curves that are orthogonal to the level lines of u.

2) Show that $\tilde{a}_{12}(D^2u, Du)$ and $\tilde{a}_{22}(D^2u, Du)$ are contrast invariant differential operators. More precisely, show that if u is C^2 and g a C^2 contrast change with $g' > 0$, then these operators are invariant when we replace u by $g(u)$. Prove that $\tilde{a}_{11}$ is not contrast invariant.

3) Question 2) explains why $\tilde{a}_{11}$ is ruled out by the contrast invariant requirement, but not why $\tilde{a}_{12}$ must also be ruled out for a causal scale space. Prove that $\tilde{a}_{22}(A, p)$ is a nondecreasing function of A, while $\tilde{a}_{12}(A, p)$ is not. ■

22.5 Comments and references

Axiomatics. In this chapter, we have followed an axiomatic presentation of scale spaces developed in [11], which is a simplified version of the original given in [130]. Other axioms for affine scale space have been proposed by Olver, Sapiro, and Tannenbaum [221, 222].

Curvature motion. The first complete mathematical study on the motion of a surface by its mean curvature is the book by Brakke [44]. The main question in this field is, How regular is a surface that has been smoothed by mean curvature? Huisken proved that a convex surface smoothed by mean curvature was transformed into a sphere before vanishing to a point [147]. This result generalizes to higher dimensions a result by Gage about curve evolution [119]. The question of whether or not a surface smoothed by curvature motion ended in a sphere was first introduced by Firey for Gaussian curvature motion [108]. Osher and Sethian developed the first numerical codes for mean curvature motion, where topological changes of the surface could be dealt with efficiently [224]. Yuille observed that the Koenderink–van Doorn dynamic shape algorithm could create singularities is a dumb-bell shaped surface [291]: The "handle" part ultimately evolved into a thin filament that broke, creating singularities. (See the figures in Chapter 2, particularly Figure 2.9.) This behavior contradicts causality, one of the main axioms of scales space: Smoothing a shape should not create new features. Koenderink comments on this creation of singularities in [172]. The corresponding mathematical study of this phenomenon is due to Grayson [129]. Bounds on the gradient for the mean curvature equation are given by Barles in [35]. A general survey of singularity formation by mean curvature motion is given by Angenent in [20]. Altschuler, Angenent, and Giga prove the smoothness of the evolution of rotationally symmetric hypersurfaces and estimate the number of singular points [6]. More about regularity and singularities related to mean curvature flow can be found in [96, 22, 148, 155]. Ishii and Souganidis developed a theory of viscosity solutions for general curvature equations, including any power function of the curvature or the Gaussian curvature [157]. Particular mention must be made about the work by Caselles and Sbert on the properties of scale spaces in three dimensions [66]. They prove that the dumb-bell is not "pinched off" by the affine scale space, but they exhibit examples of other surfaces where singularities may appear. Chow proved that in $\mathbb{R}^N$, a motion by the N-th root of the Gauss curvature deforms strictly convex surfaces into spheres [74, 75]. This result is analogue to the result by Huisken for mean curvature motion mentioned above.

Extensions of affine scale space in two dimensions. Several authors have attempted to extend the affine scale space in two dimensions to a projective-invariant scale space. The results of this chapter have clearly shown that a multiscale analysis that is both local and causal cannot be projective invariant. An affine-invariant scale space must have the form given in equation (22.25), which is completely determined up to a multiplicative constant. Thus, by requiring affine invariance, we have exhausted all degrees of freedom in the choice of the PDE. The way around this is to relax one or more of the other requirements, but not one of the invariants in the projective group. Faugeras and Keriven [105, 106, 107, 104] and Bruckstein and Shaked [48] give up the maximum principle. They then derive higher order PDEs that can hardly be considered smoothing operators. Establishing existence proofs and numerical simulations of this projective curve scale space are open problems. (See also Olver, Sapiro, and Tannenbaum [219].) Dibos does not give up locality or causality, and she is able to simulate her scale space numerically. This scale space no longer depends on a single scale parameter, but rather on two parameters. Geraets and others propose affine-invariant scale spaces for discrete sets with applications to object recognition [121, 122]. One of the first attempts to use the AMSS model for affine-invariant shape recognition was given by Cohignac, Lopez and Morel [76]. A more complete and sophisticated attempt, which performs image comparison by applying the affine scale space to all level lines of each image, is found in [186]. Alvarez and Morales used the affine scale space for corner and T-junction detection in digital images [13].

Chapter 23

Monotone image operators: "nonflat" morphology

23.1 General form of monotone operator.

Theorem 23.1. 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 $u \in I\!\!F$.)

$$\Leftrightarrow \forall\epsilon > 0,\ \sup_{v\in I\!\!F} \inf_{\mathbf{y}} u(\mathbf{y}) - \lambda + \epsilon - v(\mathbf{y}-\mathbf{x}) \geq 0$$

$$\Leftrightarrow \sup_{v\in I\!\!F} \inf_{\mathbf{y}} u(\mathbf{y}) - v(\mathbf{y}-\mathbf{x}) \geq \lambda$$

□

23.2 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 $I\!F$ and an operator T defined by

$$T(u)(\mathbf{x}) = \inf_{f \in I\!F} \sup_{\mathbf{y} \in I\!R^N} u(\mathbf{y} + \mathbf{x}) - f(\mathbf{y}).$$

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.

23.2.1 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

$$\begin{array}{lll} g_h(\mathbf{x}) & = \frac{1}{\pi h^2} & \text{for } |\mathbf{x}| \le h \\ & = 0 & \text{otherwise} \end{array}$$

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 $I\!F$ 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

$$T_h(u)(\mathbf{x}) = \inf_{f \in I\!F} \sup_{\mathbf{y} \in I\!R^N} u(\mathbf{x} + \mathbf{y}) - h^{\beta} f(\mathbf{y}/h). \tag{23.1}$$

23.2.2 Legendre Fenchel transform

Definition 23.2. 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.

23.2.3 Asymptotic theorem, first order case

Lemma 23.3. Let f be a function satisfying the following conditions:

$$\exists C > 0 \text{ and } \alpha > \max(\beta, 1) \text{ such that } \liminf_{|\mathbf{x}|\to\infty} \frac{f(\mathbf{x})}{|\mathbf{x}|^\alpha} \geq C \text{ and } f(0) \leq 0 \tag{23.2}$$

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 (23.2) 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 (23.2), 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

$$\sup_{\mathbf{y}\in \mathbb{R}^N} (u(h\mathbf{y}) - h^\beta f(\mathbf{y})) = h^\beta(\sup_{\mathbf{y}\in S_h} (h^{1-\beta} p.\mathbf{y} - f(\mathbf{y}) + O(h^{2(1-\frac{\beta-1}{\alpha-1})-\beta})))$$

$$= h^\beta(\sup_{\mathbf{y}\in \mathbb{R}^N} (h^{1-\beta} p.\mathbf{y} - f(\mathbf{y})) + O(h^{2(1-\frac{\beta-1}{\alpha-1})})) = h^\beta (f^*(h^{1-\beta}p)) + O(h^{2(1-\frac{\beta-1}{\alpha-1})})$$

It is easily checked that $O(h^{2(1-\frac{\beta-1}{\alpha-1})}) = o(h^\beta)$ for all $\beta < 2$. □

Theorem 23.4. Let $\mathbb{F}$ be a family of functions, all satisfying the condition (23.2) 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 $\mathbb{F}$ 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

$$H_1(p) = \inf_{f\in \mathbb{F}} f^*(p)$$

23.2.4 Second order case - some heuristics.

Theorem 23.4 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 $A = D^2u(0)/2$,

$$\sup_{\mathbf{y}\in \mathbb{R}^N} u(h\mathbf{y}) - h^\beta f(\mathbf{y}) = \sup_{\mathbf{y}\in \mathbb{R}^N} hp.\mathbf{y} + h^2 A\mathbf{y}.\mathbf{y} - h^\beta f(\mathbf{y}) + O(|h\mathbf{y}|^3)$$

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.

23.3 Application to image enhancement: Kramer's operators and the Rudin-Osher shock filter

In [176], 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.

23.3.1 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 [?] 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

$$(T_h u)(\mathbf{x}) = \begin{cases} (T_h^+ u)(\mathbf{x}) & \text{if } (T_h^+ u)(\mathbf{x}) - u(\mathbf{x}) < u(\mathbf{x}) - (T_h^- u)(\mathbf{x}) \\ \\ (T_h^- u)(\mathbf{x}) & \text{if } (T_h^+ u)(\mathbf{x}) - u(\mathbf{x}) > u(\mathbf{x}) - (T_h^- u)(\mathbf{x}) \\ \\ u(\mathbf{x}) & \text{otherwise} \end{cases} \tag{23.3}$$

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 23.4 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 23.1. 1. Check that $\forall u$ and $\forall \mathbf{x}$:

$$T_h^- u(\mathbf{x}) \le u(\mathbf{x}) \le T_h^+ u(\mathbf{x})$$

2. Using Lemma 23.3 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 23.5. Let T_h be the "Kramer" operator (given by 23.3), 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 23.1: 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 23.1, one has to push the asymptotic of T_h^+ and T_h^- to the second order. We have

$$T_h^+(u)(\mathbf{x}) = \sup_{\mathbf{y}\in \mathbb{R}^N} u(\mathbf{y}) - \frac{(\mathbf{x}-\mathbf{y})^2}{2h} \qquad \text{and } T_h^-(u)(\mathbf{x}) = \inf_{\mathbf{y}\in \mathbb{R}^N} u(\mathbf{y}) + \frac{(\mathbf{x}-\mathbf{y})^2}{2h}$$

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

$$T_h^+(u)(0) - u(0) = \frac{h}{2}|\mathbf{p}|^2 + h^2 A(\mathbf{p},\mathbf{p}) + o(h^2) \tag{23.4}$$

Similarly,

$$T_h^-(u)(0) - u(0) = \frac{h}{2}|\mathbf{p}|^2 - h^2 A(\mathbf{p},\mathbf{p}) + o(h^2) \tag{23.5}$$

From these two last equalities we deduce that

$$((T_h^+ u)(\mathbf{x}) - u(\mathbf{x})) - (u(\mathbf{x}) - (T_h^- u)(\mathbf{x})) = h^2(D^2u(\mathbf{x}))(Du(\mathbf{x}), Du(\mathbf{x})) + o(h^2) \tag{23.6}$$

We therefore have

$$T_h(u)(\mathbf{x}) - u(\mathbf{x}) = |Du(\mathbf{x})|^2 \; sgn(\; D^2u(\mathbf{x}) \; (Du(\mathbf{x}), Du(\mathbf{x})) \;) + o(h)$$

□

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.

23.3.2 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 23.2. Prove that

$$\lim_{h\to 0} T_h u - u = sgn(\Delta u)|Du|$$

23.4 Can we approximate a parabolic PDE by the iterations of a monotone image operator ?

23.4.1 Approximation of first order equation.

Let us address the converse of theorem 23.4: 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 23.6. Let G be a convex function, such that G^* satisfies condition 23.2, then choosing $I\!F_h = \{hG^*(\mathbf{x}/h)\}$ one has for the operator T_h associated to $I\!F_h$ 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 23.3 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 23.7. 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 23.2, then choosing $I\!F_h = \{hg_q^*(\mathbf{x}/h)\}$ one has for the operator T_h associated to $I\!F_h$ 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 23.7 is a straighforward consequence of Theorem 23.4. □

Examples of functions G that fit the hypothesis of the theorem 23.7 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,

$$g_q^*(p) = \begin{cases} pq - G(q) & if|p| \le K \\ +\infty & otherwise \end{cases}$$

So that $g_q^*(p)$ satisfies the condition 23.2.

Remark 23.1. *However, the hypotheses of Theorem 23.7 do not permit to construct any function G. The main issue is in fact the condition 23.2, which localizes the filter when $h > 0$ tends to 0, in the theorem 23.4.*

Frédéric Cao proposes in [53] 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 $G(p) = \inf_{q\in\mathbb{R}^N} g_q(p)$. He then set $f_q(\mathbf{x}) = (g_q^*)(\mathbf{x}) = -G(q) + q\mathbf{x}$ and $\mathbb{F}_h = \{f_{q,h}, q \in \mathbb{R}^N\}$ where, for a $\alpha \in]1/2, 1[$,

$$f_{q,h}(\mathbf{x}) = \begin{cases} -hG(q) + q\mathbf{x} & \textit{if } \mathbf{x} \in B(0, h^\alpha) \\ +\infty & \textit{elsewhere} \end{cases}$$

The familly $\mathbb{F}_h$ is not a rescaling of the familly $\mathbb{F}_1$. There is indeed, two scales: the explicit one h, and an implicit one, h^α since the functions of $\mathbb{F}_h$ are truncated outside a ball of radius h^α. This truncature localizes the corresponding operator T_h and makes the result of theorem 23.4 true, even if the functions of $\mathbb{F}_h$ do not satisfy the condition 23.2.

23.4.2 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

$$q = Du(0) \Leftrightarrow \sup_{\mathbf{x}\in\mathbb{R}^N} u(\mathbf{x}) - f_q(\mathbf{x}) = O(h^2)$$

$$q \ne Du(0) \Leftrightarrow \sup_{\mathbf{x}\in\mathbb{R}^N} u(\mathbf{x}) - f_q(\mathbf{x}) > C(q, u)h$$

This indicates that a way to get second order operator is to choose the familly of functions $\mathbb{F}$ so that $\forall f \in \mathbb{F}$ and $\forall q \in \mathbb{R}^N$ one has $f + q\mathbf{x} \in \mathbb{F}$.

The Heat Equation as the asymptotic of a non-flat morphological operator.

Lemma 23.8. Let A be in $SM(\mathbb{R}^N)$ (set of the $N \times N$ symmetric matrices). Then,

$$Tr(A) = N \inf_{Q\in SM(\mathbb{R}^N), Tr(Q)=0} \sup_{\mathbf{x}, |\mathbf{x}|=1} (A - Q)(\mathbf{x}, \mathbf{x}) \tag{23.7}$$

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 $\forall Q \in SM(\mathbb{R}^N)$, $N\sup_{\mathbf{x},|\mathbf{x}|=1}(A-Q)(\mathbf{x},\mathbf{x}) \geq Tr(A-Q) = Tr(A)$. Thus

$$N \inf_{Q\in SM(\mathbb{R}^2),Tr(Q)=0} \sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x}) \geq Tr(A).$$

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

$$\inf_{Q\in SM(\mathbb{R}^2),Tr(Q)=0} \sup_{\mathbf{x},|\mathbf{x}|=1} (A-Q)(\mathbf{x},\mathbf{x})$$

$$\leq \inf_{\{q_1,...,q_N\},q_1+...+q_N=0} \max\{\lambda_1+q_1, ..., \lambda_N+q_N\} = (\lambda_1+...+\lambda_N)/N$$

□

Lemma 23.9. We set for $p \in \mathbb{R}^N$, $Q \in SM(\mathbb{R}^N)$, and $h > 0$,

$$\begin{aligned} f_{p,Q,h}(\mathbf{x}) &= p\mathbf{x} + Q(\mathbf{x},\mathbf{x}) && \text{if } \mathbf{x} \in B(0,h) \\ &= -\infty && \text{otherwise} \end{aligned}$$

We then set $\mathbb{F}_h = \{f_{p,Q,h}; \text{with } Q \in SM(\mathbb{R}^N); Tr(Q) = 0 \text{ and } p \in \mathbb{R}^N\}$ which is to say that $\mathbb{F}_h$ is made of the truncature around zero of all quadratic forms whose trace is zero. With $T_h(u)(\mathbf{x}) = \inf_{f\in\mathbb{F}_h} \sup_{\mathbf{y}\in\mathbb{R}^N} u(\mathbf{x}+\mathbf{y}) - f(\mathbf{y})$, 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 $A = \frac{1}{2}D^2u(0)$. We have

$$T_h(u)(0) - u(0) = \inf_{p\in\mathbb{R}^N,Q\in SM(\mathbb{R}^N);Tr(Q)=0} \sup_{\mathbf{y}\in B(0,h)} u(\mathbf{y}) - u(0) - p\mathbf{y} - Q(\mathbf{y},\mathbf{y})$$

$$= \inf_{p,Q} \sup_{\mathbf{y}\in B(0,1)} u(h\mathbf{y}) - u(0) - hp\mathbf{y} - h^2Q(\mathbf{y},\mathbf{y})$$

$$= \inf_{p,Q} \sup_{\mathbf{y}\in B(0,1)} h(Du(0)-p)\mathbf{y} - h^2(A-Q)(\mathbf{y},\mathbf{y}) + o(h^2)$$

$$= h^2 \inf_{Q\in SM(\mathbb{R}^N);Tr(Q)=0} \sup_{\mathbf{y}\in B(0,1)} (A-Q)(\mathbf{y},\mathbf{y}) = \frac{1}{N}h^2Tr(A)$$

□

Chapter 24

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 $I\!R^3$, 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 **??**), **Translation invariant** (Definition **??**) and **invariant by grey level translation** (Definition **??**). Therefore, as shown in Chapter **??**, 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 **??**, 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).

24.1 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 **??** 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 $I\!R^3$, 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 $I\!R^2 * \{0\}$. That is, when we apply the same affine transform on each image of the movie.

Definition 24.1. 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 24.2. 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 24.1 and 24.2. 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 24.3. For any e in $I\!R^+$ 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 24.1 and 24.2.

Now, we want to state the scale invariance, as done in chapter **??**. 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 $I\!R^3$. 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 **??** 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 **??**, we deduce that we can normalize the relation between t, B and t' so that

$$t' = (det(B))^{\frac{1}{2}} t \tag{24.1}$$

Thus the affine invariance is reduced to the property :

$$F(BA^tB, Bp, t) = |det(B)|^{\frac{1}{2}} F(A, p, t|det(B)|^{\frac{1}{2}}) \tag{24.2}$$

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

$$F(S_eAS_e, S_ep, t) = eF(A, p, et) \tag{24.3}$$

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 24.4. 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 24.1 or 24.2 and 24.3).

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 24.5. 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

$$B_v(T_{t^*}u) = T_t(B_vu), \text{ and } B_v(T_{t^*(t,v), t^*(s,v)}u) = T_{t,s}(B_vu)$$

$t^*(t, B_{-v}) = t^*(t, B_v)$, 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 **??**(i), we have

$$t^*(t^*(t, B_v), B_v) = t^*(t^*(t, B_v), B_{-v}) = t^*(t, B_vB_{-v}) = t^*(t, Id) = t.$$

Repeating the argument of the step (ii) of the proof of the Lemma **??**, 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)

$$B_v(T_tu) = T_t(B_vu) \Leftrightarrow F({}^tB_vAB_v, {}^tB_v\mathbf{p}, t) = F(A, \mathbf{p}, t) \qquad \forall A \text{ in } S^3, \mathbf{p} \in I\!R^3 \tag{24.4}$$

Finally, we state the morphological property, (as in definition **??**):

Definition 24.6. 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 **??** that this implies

$$F(\mu A + \lambda \mathbf{p} \otimes \mathbf{p}, \mu \mathbf{p}, t) = \mu F(A, \mathbf{p}, t), \tag{24.5}$$

for every real values λ, μ, every symmetric matrix A and every three-dimensional vector $\mathbf{p}$.

24.2 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 $I\!R^2$), 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 ([**?**, **?**, **?**, **?**, **?**],...)

$$u(\mathbf{x} + \mathbf{v}(\mathbf{x})\Delta\theta) = u(\mathbf{x}) + o(\Delta\theta) \tag{24.6}$$

$$\Leftrightarrow Du.\mathbf{v} = 0 \tag{24.7}$$

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 :

$$(G_t * u)(\mathbf{x} + \mathbf{v}(\mathbf{x})\Delta\theta) = (G_t * u)(\mathbf{x}) \tag{24.8}$$

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 (24.8), $\mathbf{v}_2$ must satisfy

$$(G_t * u(A.))(\mathbf{x} + \mathbf{v}_2 \Delta\theta) = (G_t * u(A.))(\mathbf{x}) \tag{24.9}$$

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 (24.8), $\mathbf{v}_2 = 2\mathbf{v}_1$. Indeed, in the two cases, the smoothing is not done in the same way : because this linear smoothing <u>is not Galilean invariant</u>. Therefore a such smoothing implies some perturbation into the estimation of the velocities.

Adelson and Bergen [3], and Heeger [**?**] 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 (24.6) 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 $I\!R$ into $I\!R$, then the equation (24.6) 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 (24.7) 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 [**?**], Snyder [**?**], Nagel [**?**], 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.

24.3 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 $I\!R^3$. 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 24.7. Let a multiscale analysis T_t be causal (as defined in theorem **??**), translation, Euclidean, Galilean, and constrast invariant. Then, there exists a function F such that T_t is governed by the equation

$$\frac{\partial u}{\partial t} = |\nabla u| \; F(curv(u), accel(u), t) \qquad (24.10)$$

(for the exact meaning of "governed by", we refer to the theorem **??**.)

If in addition, T_t is affine, time-scale and time invariant then the only possible scale-space equations are

$$(\mathrm{AMG}) \qquad \frac{\partial u}{\partial t} = |\nabla u| \; curv(u)^{\frac{1-q}{3}} \; (\mathrm{sgn}(curv(u))accel(u)^q)^+ \qquad (24.11)$$

for some $q \in]0, 1[$, or

$$(q = 0) \qquad \frac{\partial u}{\partial t} = |\nabla u| curv(u)^{\frac{1}{3}} \qquad (24.12)$$

$$(q=0) \qquad \frac{\partial u}{\partial t} = |\nabla u| curv(u)^{\frac{1}{3}} (\text{sgn}(accel(u)curv(u))^{+} \tag{24.13}$$

$$(q=1) \qquad \frac{\partial u}{\partial t} = |\nabla u| \text{sgn}(curv(u))(\text{sgn}(curv(u))accel(u))^{+} \tag{24.14}$$

In the above formulae, we use the convention that the power preserves the sign, that is $a^q = |a|^q \text{sgn}(a)$. 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 (24.11) 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 (24.11) 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 (24.11) 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 24.8. 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 (24.11) satifies $A(T_\tau u) = T_t(Au)$ with

$$\tau(A, t) = (p^{4(\frac{1-q}{3})} e^{2q})t \tag{24.15}$$

We see in relation (24.15), 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 (24.14). 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

$$\frac{\partial u}{\partial t} = |Du|(\mathrm{G}(u)^{+})^{\frac{1}{4}} \tag{24.16}$$

which is the unique contrast and 3D affine invariant scale-space as described in chapter **??**.

Let us before begining the proof of the theorem give a hint on the kind of smoothing the equation (24.11) 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 24.7 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 [11]. $\square$

.

24.4 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 (24.11) 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

$$\mathcal{W} = \{ \mathbf{v} = (\vec{v}, 1) \text{ for all } \vec{v} \text{ in } \mathbb{R}^2 \} \tag{24.17}$$

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 $\mathcal{V}$:

$$\mathcal{V} = \{ \mathbf{v}_\mu = \frac{|Du|}{|\nabla u|}(\mu \mathbf{e}^\perp - \mathbf{e}^\pm), \text{ for all } \mu \in \mathbb{R} \} \tag{24.18}$$

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 24.9. Definition of the "velocity vector". When ∇u and $curv(u) \neq 0$, we define the "velocity vector": $\mathbf{V}$by

$$\mathbf{V} = \frac{|Du|}{|\nabla u|}(\frac{\Gamma_2}{\Gamma_1}\mathbf{e}^\perp - \mathbf{e}^\pm) \tag{24.19}$$

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:

$$v_1 = -\frac{u_\theta}{|\nabla u|} \qquad v_2 = \frac{|Du|}{|\nabla u|}\frac{\Gamma_2}{\Gamma_1} \tag{24.20}$$

Proposition 24.10. 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 (24.9) 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 24.9 is to make the choice of a travelling direction on the space-time level surface. With the definition 24.9, we choose 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.

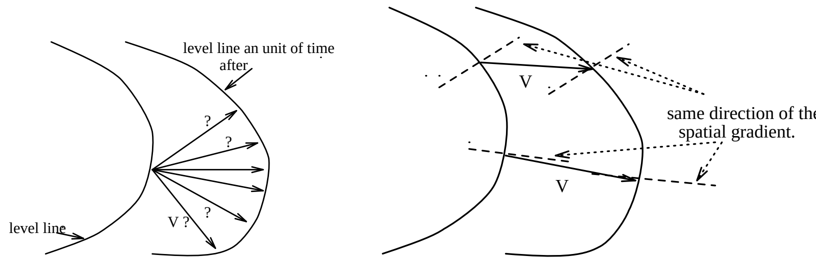

Figure 24.1: According to the optical flow definition, all above drawn velocity vectors are possible, since they allow the moving point to remain on the same level surface. One possibility to get rid of this ambiguity is choose as velocity the direction which does not change the orientation of the spatial gradient.

Proposition 24.11. 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 (24.9), and v_1 its component in the direction of the spatial gradient.

$$accel(u) = -\frac{\mathrm{D}v_1}{\mathrm{D}\theta} = -(v^x \frac{\partial v_1}{\partial x} + v^y \frac{\partial v_1}{\partial y} + \frac{\partial v_1}{\partial \theta})$$

$$= -((Dv_1).\mathbf{V}) = -(D(\mathbf{V}.\nabla u).\mathbf{V}) \tag{24.21}$$

This formula[1] shows that $accel(u)$ is the acceleration in the direction of $-\nabla u$. As v_1 the component 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 24.19, then

$$accel(u) = \frac{(D^2u)(\mathbf{V}, \mathbf{V})}{|\nabla u|} \tag{24.22}$$

Proof of proposition 24.11 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 24.12. Whenever the spatial gradient ∇u and the spatial curvature $curv(u)$ are not equal to zero,

$$|\nabla u|(sgn(curv(u))\ accel(u))^+ = min_{\mathbf{v}\in\mathcal{V}}|(D^2u)(\mathbf{v}, \mathbf{v})| \tag{24.23}$$

[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.

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 24.11. 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 24.13. Let ∇u and $curv(u)$ be not equal to zero, and u be C^2, then the expression

$$min_{\mathbf{v}\in\mathcal{W}}\left(\frac{1}{\Delta\theta^2}(|u(\mathbf{x}-\mathbf{v}\Delta\theta) - u(\mathbf{x})| + |u(\mathbf{x}+\mathbf{v}\Delta\theta) - u(\mathbf{x})|)\right) \tag{24.24}$$

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

$$F(\mathbf{v},h) = |-\frac{\mathbf{p}.\mathbf{v}}{h} + \frac{1}{2}A(\mathbf{v},\mathbf{v})| + |\frac{\mathbf{p}.\mathbf{v}}{h} + \frac{1}{2}A(\mathbf{v},\mathbf{v})| \tag{24.25}$$

Let $\mathbf{w} \in \mathcal{V}$ be a vector which minimizes the min in (24.23), $\mathbf{w} \in \mathcal{V}$ then $\mathbf{w}.\mathbf{p} = 0$), thus (24.25) becomes

$$F(\mathbf{w},h) = |A(\mathbf{w},\mathbf{w})|$$

Therefore

$$\lim_{h\to 0}(min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h)) \le F(\mathbf{w},h) = |\nabla u|(sgn(curv(u))\ accel(u))^+ \tag{24.26}$$

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 $F(\mathbf{v}_h,h) \ge 2|(\mathbf{p}.\mathbf{v}_h)/h|$, we necessarily have

$$|(\mathbf{p}.\mathbf{v}_h)| = O(h) \tag{24.27}$$

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 (24.27) leads that $|\mathbf{v}_h^{\pm}|$ is bounded when h tends to zero. As before, we have

$$F(\mathbf{v}_h,h) \ge |A(\mathbf{v}_h,\mathbf{v}_h)| \ge |A((\mathbf{v}_h^{\perp}+\mathbf{v}_h^{\pm}),(\mathbf{v}_h^{\perp}+\mathbf{v}_h^{\pm}))| \ge$$

$$|A(\mathbf{v}_h^{\perp},\mathbf{v}_h^{\perp}) + 2hA(\mathbf{v}_h^{\perp},\mathbf{v}_h^{\pm}) + h^2A(\mathbf{v}_h^{\pm},a\mathbf{v}_h^{\pm})|$$

Since $|\mathbf{v}_h^{\pm}|$ is bounded, we get $\lim_{h\to 0}F(\mathbf{v}_h,h) \ge |A(\mathbf{v}_h^{\perp},\mathbf{v}_h^{\perp})|$ Now, $\mathbf{v}_h^{\perp}$ is in $\mathcal{V}$ then $|A(\mathbf{v}_h^{\perp},\mathbf{v}_h^{\perp})| \ge min_{\mathbf{v}\in\mathcal{V}}|A(\mathbf{v},\mathbf{v})|$, so

$$\lim_{h\to 0}(min_{\mathbf{v}\in\mathcal{W}}F(\mathbf{v},h)) = \lim_{h\to 0}F(\mathbf{v}_h,h)$$

$$\ge min_{\mathbf{v}\in\mathcal{V}}|A(\mathbf{v},\mathbf{v})| = |\nabla u|(sgn(curv(u))\ accel(u))^+ \tag{24.28}$$

(24.26) and (24.28) 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 24.14. Let u be C^2, ∇u and $curv(u)$ not zero, then

$$min_{\mathbf{v}\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}\Delta\theta)-u(\mathbf{x})| + |u(\mathbf{x}+\mathbf{v}\Delta\theta)-u(\mathbf{x})|) = \tag{24.29}$$

$$min_{\mathbf{v}_b,\mathbf{v}_a\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}_b\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}_a\Delta\theta)-u(\mathbf{x})|+\Delta\theta|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|)+o(\Delta\theta^2)$$

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.

$$\begin{aligned}
&(|u(\mathbf{x}-\mathbf{v}_bh)-u(\mathbf{x})| + |u(\mathbf{x}+\mathbf{v}_ah)-u(\mathbf{x})| + h|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|) \\
= \quad & |-h(Du.\mathbf{v}_b) + \frac{h^2}{2}(D^2u)(\mathbf{v}_b,\mathbf{v}_b)| + |h(Du.\mathbf{v}_a) + \frac{h^2}{2}(D^2u)(\mathbf{v}_a,\mathbf{v}_a)| \\
& +h|Du.(\mathbf{v}_b-\mathbf{v}_a)|) + o(h^2) \\
\ge \quad & \frac{h^2}{2}(|(D^2u)(\mathbf{v}_b,\mathbf{v}_b)| + |(D^2u)(\mathbf{v}_a,\mathbf{v}_a)|) + o(h^2) \\
\ge \quad & min_{\mathbf{v}\in\mathcal{W}}(|(D^2u)(\mathbf{v},\mathbf{v})|) + o(h^2) \\
= \quad & min_{\mathbf{v}\in\mathcal{W}}(|u(\mathbf{x}-\mathbf{v}h)-u(\mathbf{x})| + |u(\mathbf{x}+\mathbf{v}h)-u(\mathbf{x})|) + o(h^2)
\end{aligned}$$

by Proposition 24.13. □

Interpretation. We deduce from all of these propositions an explicit formula for the apparent acceleration

$$|\nabla u|(sgn(curv(u))\ accel(u))^+ = \tag{24.30}$$

$$min_{\mathbf{v}_b,\mathbf{v}_a\in\mathcal{W}} \frac{1}{\Delta\theta^2}(|u(\mathbf{x}-\mathbf{v}_b\Delta\theta)-u(\mathbf{x})|+|u(\mathbf{x}+\mathbf{v}_a\Delta\theta)-u(\mathbf{x})|+\Delta\theta|\nabla u.(\mathbf{v}_b-\mathbf{v}_a)|)+o(1)$$

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 $|\nabla u|$.

24.5 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 $curv(u)^{\frac{1-q}{3}}$ $(\mathrm{sgn}(curv(u))accel(u)^q)^+$. (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 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 24.5, 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 24.5, 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 24.5).

We conclude that the assumptions we made for our model are incompatible with the notion of smoothing trajectories. Indeed non-straight trajectories are

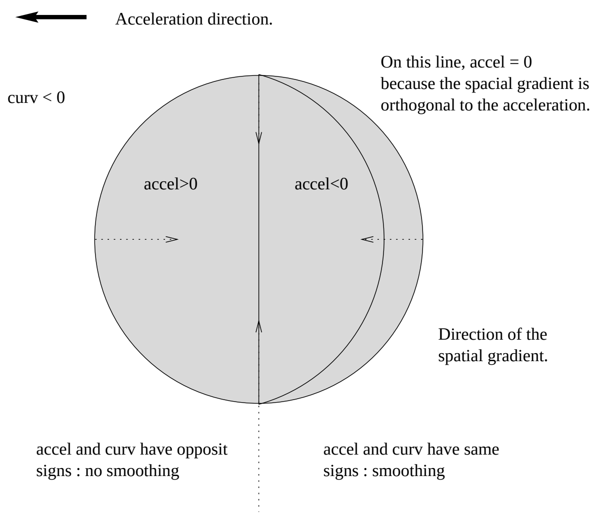

Figure 24.2: The AMG model erodes a circle in acceleration only on one side. Indeed, when the curvature and the acceleration have opposite signs, the evolution in scale is zero. (see the AMG equation).

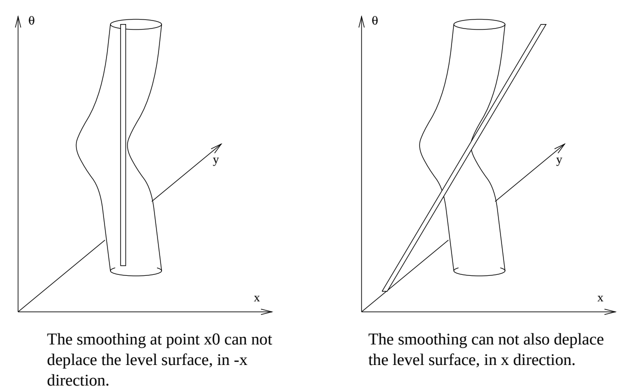

Figure 24.3: Saddle points of level surfaces remain steady by the AMG model. Indeed, our scale-space can be seen as a motion in scale of gray level-surfaces (isophotes). The level-surfaces that are straight in time correspond to a uniform translation and are not changed by the smoothing. Therefore, the 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.

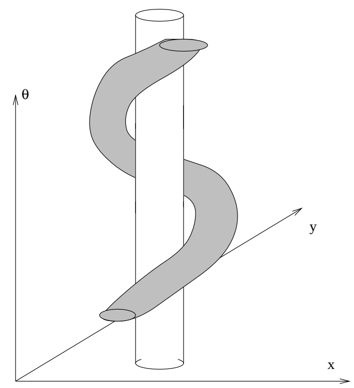

Figure 24.4: The level-surface in gray cannot become straight : it would have to cross the white level-surface which is invariant by the scale space. Now, during the smoothing process, the level-surface in gray will be eroded on its convex part, and will eventually disappear at a fixed scale : it cannot 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.

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.

24.6 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 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.

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: [**?**, **?**, **?**, 3, **?**, **?**, **?**, 17, **?**, **?**]... 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

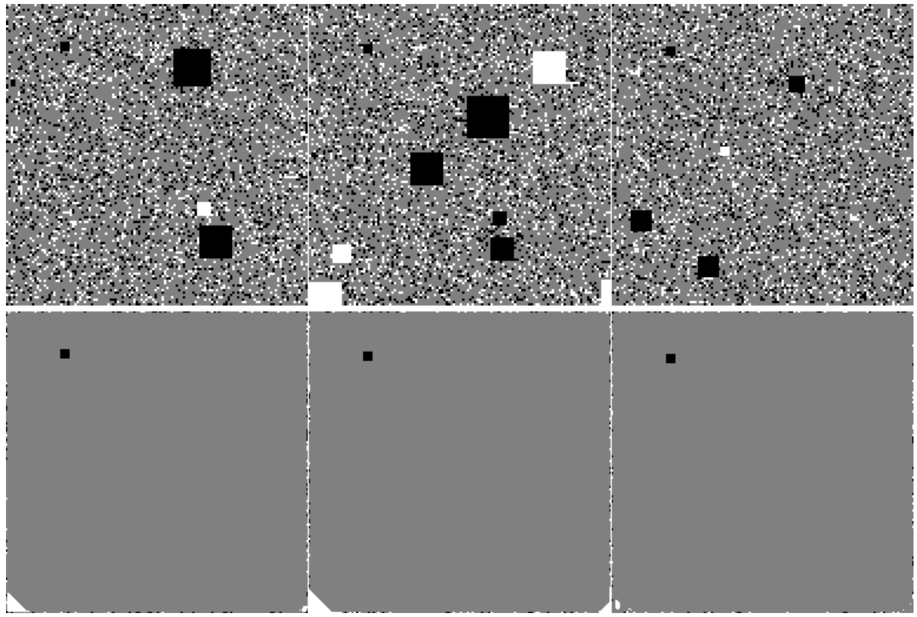

Figure 24.5: The affine, morphological, galilean (AMG) model used for image sequence restoration (extraction of coherent trajectories). Above : three successive images extracted from a synthetic sequence, made of salt and pepper noise, plus some squares placed at random locations. In addition, a little 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.

smoothness constraint on the flow it self, see e.g. [**?**, **?**, **?**, **?**] or in some cases by an implicit smoothing of image sequences see e.g. [3]...

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 [3], 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 [11] 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 direction of the (unknown) optical flow [11, **?**]... Several other works have introduced other smoothings depending on its aim and where the time plays a specific rule. In [**?**] the author formalizes a smoothing compatible with the aim of estimating depth from an image sequence. In [**?**] and one could find adaptations of the 2D linear smoothing theory to an anisotropic diffusion in the direction of an estimated optical

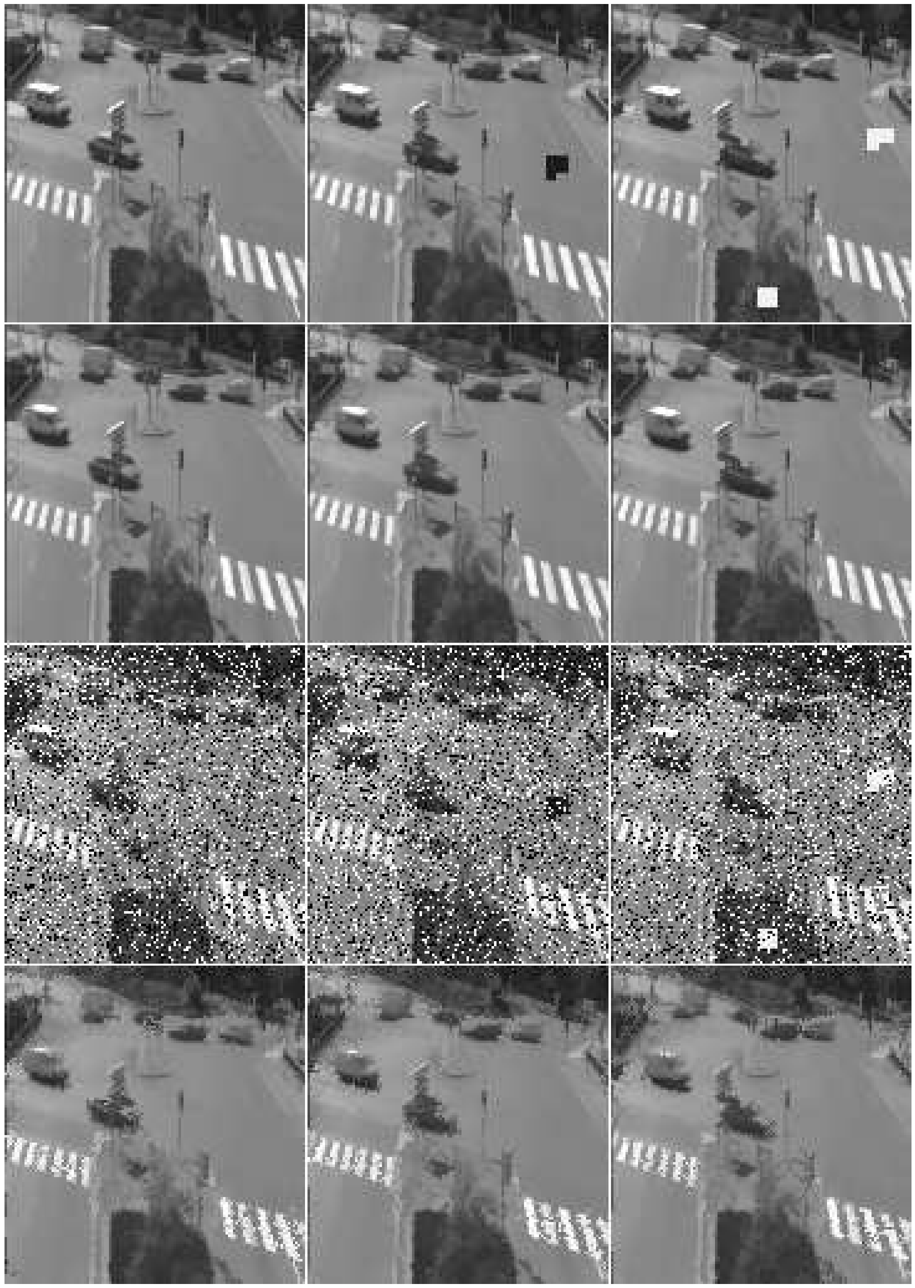

Figure 24.6: AMG model (Affine, Morphological, Galilean) used for image sequence “denoising”. Above : three successive images extracted from a sequence. Second row : resulting images at calibrated scale 100 pixels (scale at which a spatiotemporal sphere of 100 pixels disapears). Third row : Some noise has been added to the original sequence (25% of the pixels are corrupted). Bottom : resulting images at scale 100.

flow.

Chapter 25

A snake from A to Z...

25.1 An active contour model

Boundaries of objects perceived against a different background induce some discontinuities in the gray level, resulting in high gradient. Let us call *contour* a closed Jordan curve located mostly at high gradient points. The aim of the active contour methods is to find such contours, starting from an initial curve, usually sketched by hand. The curve moves numerically from its original location until it reaches a position where it maximizes the image contrast. We do not call as usual the image u_0 because our main focus is not on the "analyzed" image but on the "analyzing image" describing the snake motion. This last image will be called $u(t, \mathbf{x})$ and the analyzed image will be called $I(\mathbf{x})$. The boundary detection problem can be formulated as an optimization problem. We shall treat it in 2D but the 3D case has exactly the same formalism. Of course, then, "curve" has to be replaced by "surface". Let us choose a function g from $I\!R^2$ into $I\!R$ 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. We shall set for example

$$g(\mathbf{x}) = \sigma^2/(\sigma^2 + |DI(\mathbf{x})|^2) \tag{25.1}$$

on the image domain, where σ^2 is the estimated variance of the noise and texture around the object. For convenience we shall extend I and g outside this domain to $\mathbb{R}^2$, but assume that $g(\mathbf{x})$ is zero for $\mathbf{x} \geq R$ large enough. Thus, given an image I, and an initial curve $C_0 = (\mathbf{x}_0(s)), s \in [0, L(C_0)]$, we want to find a curve $C = (\mathbf{x}(s)), s \in [0, L(C)]$ that minimizes the energy

$$E(C) = \int_0^L g(\mathbf{x}(s))ds \tag{25.2}$$

around $\mathbf{x}_0$, where s is an arc length parameter and $L = L(C)$ the length of the curve C.

In the following, we will assume g to be twice differentiable with respect to $\mathbf{x}$. In order to achieve this in practice, I is previously smoothed by the heat equation and DI is therefore replaced in (25.1) by $G * DI = D(G * I)$ where G is a gaussian with small variance. In that way g becomes C^∞ (see Proposition 1.5.)

Let us refresh some differential notation on curves. We denote by $\mathbf{x}(z)$ a parameterization of a curve C on a fixed interval $[0,1]$. Recall that $\boldsymbol{\tau}(z) = \frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}$ is the tangent unit vector to the curve and $\mathbf{n}(z) = \boldsymbol{\tau}(z)^{\perp}$ the unit normal. Notice that if $\mathbf{v}$ is a vector, then we can decompose it on the mobile frame $(\boldsymbol{\tau}(z), \mathbf{n}(z))$ as

$$\mathbf{v} = (\mathbf{v}.\boldsymbol{\tau})\boldsymbol{\tau} + (\mathbf{v}.\mathbf{n})\mathbf{n}. \tag{25.3}$$

Calling $s(z)$ a length parameter on the curve, defined up to a constant by $s'(z) = |\mathbf{x}'(z)|$, one has $\boldsymbol{\tau}(s) = \frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}$ and

$$\frac{\partial \boldsymbol{\tau}}{\partial s} = \boldsymbol{\kappa}(\mathbf{x}(z)), \tag{25.4}$$

which is the curvature vector. Thus, differentiating the tangent vector with respect to z yields

$$\left(\frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}\right)' = \frac{\partial \boldsymbol{\tau}}{\partial z} = \frac{\partial \boldsymbol{\tau}}{\partial s}\frac{\partial s}{\partial z} = \boldsymbol{\kappa}(\mathbf{x}(z))|\mathbf{x}'(z)|. \tag{25.5}$$

Proposition 25.1. *Let $C(t) = \mathbf{x}(t,s)$ be a curve resulting from the gradient descent of the energy (25.2), starting from $C(0)$. Assume that $C(t)$ is C^2. Then $C(t)$ satisfies the following equation*

$$\frac{\partial \mathbf{x}}{\partial t} = -(Dg(\mathbf{x}).\mathbf{n})\mathbf{n}) + g(\mathbf{x})\boldsymbol{\kappa}(\mathbf{x}) \tag{25.6}$$

and $\mathbf{x}(0,s) = \mathbf{x}_0(s)$ *that is* $(C(0) = C_0)$

Proof. We shall first change the parameterization of the curve C so that its length is no longer a parameter of the energy. We parameterize the curve with $z \in [0,1]$. We have $ds = |\mathbf{x}'(z)|dz$, where $'$ denotes the derivative with respect to z. Thus

$$E(C) = \int_0^1 g(\mathbf{x}(z))|\mathbf{x}'(z)|dz$$

Consider any C^2 perturbation of the curve $\mathbf{x}(z)$, which we call dC and denote its parameterization by $d\mathbf{x}(z)$. By $|dC|$ we mean the C^2 sup-norm of $d\mathbf{x}(z)$ on $[0,1]$. By an easy differentiation,

$$\begin{aligned} E(C+dC) - E(C) &= \int_0^1 Dg(\mathbf{x}(z)).d\mathbf{x}(z)|\mathbf{x}'(z)|dz \\ &+ \int_0^1 g(\mathbf{x}(z))\frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}.d\mathbf{x}'(z)dz + o(|dC|). \end{aligned}$$

Integrating by parts the last integral, we therefore have by (25.5),

$$\begin{aligned} E(C+dC) - E(C) &= \int_0^1 Dg(\mathbf{x}(z)).d\mathbf{x}(z)|\mathbf{x}'(z)|dz \\ &- \int_0^1 (Dg(\mathbf{x}(z)).\mathbf{x}'(z))\frac{\mathbf{x}'(z)}{|\mathbf{x}'(z)|}.d\mathbf{x}(z)dz \\ &- \int_0^1 g(\mathbf{x}(z))\boldsymbol{\kappa}(\mathbf{x}(z)).d\mathbf{x}(z)|\mathbf{x}'(z)|dz + o(|dC|). \end{aligned}$$

Using (25.3) with $\mathbf{v} = Dg(\mathbf{x}(z))$, the two first integrals can be merged and we obtain

$$E(C+dC) - E(C) =$$
$$\int_0^1 (Dg(\mathbf{x}(z)).\mathbf{n})\mathbf{n}.d\mathbf{x}(z)|\mathbf{x}'(z)|dz - \int_0^1 g(\mathbf{x}(z))\boldsymbol{\kappa}(\mathbf{x}(z)).d\mathbf{x}(z)|\mathbf{x}'(z)|dz + o(|dC|).$$

Let us denote the intrinsic scalar product between two vectorial functions f and h defined on the curve $\mathbf{x}(z)$ by

$$\langle f.g\rangle = \int_0^1 f(\mathbf{x}(z))g(\mathbf{x}(z))|\mathbf{x}'(z)|dz$$

Thus,

$$E(C+dC) - E(C) = \langle dC, ((Dg(\mathbf{x}(z)).\mathbf{n})\mathbf{n} - g(\mathbf{x}(z))\boldsymbol{\kappa}\mathbf{x}(z)))\rangle + o(|dC|)$$

We therefore have

$$\nabla E(C) = (Dg(\mathbf{x}(z)).\mathbf{n})\mathbf{n} - g(\mathbf{x}(z))\boldsymbol{\kappa}(\mathbf{x}(z)).$$

As a consequence the gradient descent for a curve $C = (\mathbf{x}(z))$ following the steepest gradient descent can be described by the equation $\frac{\partial C}{\partial t} = -\nabla E(C)$, that is

$$\frac{\partial \mathbf{x}(t,z)}{\partial t} = -(Dg(\mathbf{x}(t,z)).\mathbf{n})\mathbf{n} + g(\mathbf{x}(z))\boldsymbol{\kappa}(\mathbf{x}(z)).$$

From this equation we can deduce the normal motion

$$\frac{\partial \mathbf{x}(t,z)}{\partial t} = -(Dg(\mathbf{x}(t,z)).\mathbf{n})\mathbf{n} + g(\mathbf{x}(z))(\boldsymbol{\kappa}(\mathbf{x}(z)).\mathbf{n})\mathbf{n}.$$

This last evolution is obtained from the former one by projecting $\frac{\partial \mathbf{x}(t,z)}{\partial t}$ on the normal line $\mathbb{R}\mathbf{n}(t,z)$. Indeed, in order to describe the geometric evolution of a curve we only need to give the motion of each one of its points in the direction normal to its tangent. A last simplification is now obtained when for each t, we choose to reparametrize the curve by an arc-length parameter s. In such a case $\boldsymbol{\kappa}(\mathbf{x}(t,s)) = \frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s)$ is normal to the curve. Thus we can simply write

$$\frac{\partial \mathbf{x}(t,s)}{\partial t} = -(Dg(\mathbf{x}(t,s)).\mathbf{n}(t,s))\mathbf{n}(t,s) + g(\mathbf{x}(t,s))\boldsymbol{\kappa}(\mathbf{x}(t,s)) =$$
$$-(Dg(\mathbf{x}(t,s)).\mathbf{n}(t,s))\mathbf{n}(t,s) + g(\mathbf{x}(t,s))\frac{\partial^2 \mathbf{x}}{\partial s^2}(t,s),$$

or, if we omit the variables:

$$\frac{\partial \mathbf{x}}{\partial t} = -(Dg(\mathbf{x}).\mathbf{n})\mathbf{n} + g(\mathbf{x})\boldsymbol{\kappa}(\mathbf{x}) = -(Dg(\mathbf{x}).\mathbf{n})\mathbf{n} + g(\mathbf{x})\frac{\partial^2 \mathbf{x}}{\partial s^2}$$

□

Unfortunately, we cannot be sure that such an evolution yields a regular curve for all t. In fact, it is in general false, since topological changes for the curve can occur, which imply the appearance of infinite curvatures (see Figure 25.4.)

By a straightforward adaptation of the proof of Proposition 12.7 one immediately obtains a formal link between the snake curve motion and an image motion.

Proposition 25.2. *Assume that a function $(t, \mathbf{x}) \mapsto u(t, \mathbf{x})$ is C^2 in a neighborhood of $(t_0, \mathbf{x}_0)$ and that $Du(t_0, \mathbf{x}_0) \neq 0$. Then u satisfies the snake equation*

$$\frac{\partial u}{\partial t}(t, \mathbf{x}) = g(\mathbf{x})\mathrm{curv}(u)(t, \mathbf{x})|Du|(t, \mathbf{x}) + Dg(\mathbf{x}).Du(t, \mathbf{x}) \tag{25.7}$$

in a neighborhood of $(t_0, \mathbf{x}_0)$ if and only if the normal flow (Definition 12.6) of the level lines of u passing in this neighborhood satisfies the intrinsic snake equation,

$$\frac{\partial \mathbf{x}}{\partial t}(t, \mathbf{y}) = g(\mathbf{y})\boldsymbol{\kappa}(t, \mathbf{y}) - (Dg(\mathbf{y}).\mathbf{n}(t, \mathbf{y}))\mathbf{n}(t, \mathbf{y})), \tag{25.8}$$

where $\boldsymbol{\kappa}(t, \mathbf{x}(t, \mathbf{y}))$ denotes the curvature vector of the level line of $u(t)$ passing by $\mathbf{x}(t, \mathbf{y})$ and $\mathbf{n}(t, \mathbf{y})$ one of its unit normals.

Exercise 25.1. By imitating the proof of Corollary 12.7, prove Proposition 25.2.
■

25.2 Study of the snake equation

We study in this section the equation (25.7), which we can abbreviate as

$$\frac{\partial u}{\partial t} = g|Du|curv(u) + Dg.Du \tag{25.9}$$

and shall call the *snake equation.*

Admissibility of the equation and uniqueness of solutions. Let us set

$$F(A, p, \mathbf{x}) = g(\mathbf{x})A(p^\perp, p^\perp) + Dg(\mathbf{x}).p$$

Equation (25.9) can be obviously written as

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

It is easy checked that F is admissible (see Definition 19.1). As a consequence Theorem 19.17 ensures uniqueness of viscosity solutions of the equation (25.9) for any Lipschitz initial condition u_0.

Exercise 25.2. Check that $F(A, p, \mathbf{x})$ is admissible. ■

Existence of solutions by approximation. Let us now construct an approximation scheme to the solution of Equation (25.9). It is possible to construct a family of structuring elements having as asymptotic behavior the right hand term of the equation (25.9). However, this term being a sum of two simple operators, it is simpler to associate with each one of these operators a simple family of structuring elements and to alternate their corresponding filters. Note that, due to the presence of $g(\mathbf{x})$ and $Dg(\mathbf{x})$, the equation is not invariant by translation. As a consequence, the families of structuring elements will depend on $\mathbf{x}$. This situation is new. We will need conditions ensuring that inf-sup operators with space-varying sets of structuring elements preserve Lipschitz constants.

Lemma 25.3. *Let $B \subset B(0,1)$ and $B(\mathbf{x}) = g(\mathbf{x})B$ be a space varying structuring element such that $g(\mathbf{x})$ is M-Lipschitz. Consider the associated space-varying dilation $Tu(\mathbf{x}) = \sup_{\mathbf{z}\in B(\mathbf{x})} u(\mathbf{x}+\mathbf{z})$. Then if $u(\mathbf{x})$ is a L-Lipschitz function, Tu is a LM-Lipschitz function. The same result works with an* inf *instead of a* sup.

Proof. For every $\mathbf{z} \in B$, $\mathbf{x}$ and $\mathbf{y}$ one has

$$u(\mathbf{x}+g(\mathbf{x})\mathbf{z}) \le u(\mathbf{x}+g(\mathbf{y})\mathbf{z}) + L|(g(\mathbf{x})-g(\mathbf{y}))\mathbf{z}| \le u(\mathbf{x}+g(\mathbf{y})\mathbf{z}) + LM|\mathbf{x}-\mathbf{y}|.$$

Taking the sup on B on both sides,

$$Tu(\mathbf{x}) = \sup_{\mathbf{z}\in B} u(\mathbf{x}+g(\mathbf{x})\mathbf{z}) \le Tu(\mathbf{y}) + LM|\mathbf{x}-\mathbf{y}|.$$

Using this and the analogous inequality interchanging $\mathbf{x}$ and $\mathbf{y}$ one gets the announced result. □

Corollary 25.4. *Let $\mathbb{B}$ be a family of structuring elements B such that $B \subset B(0,1)$ and $g(\mathbf{x})$ a M-Lipschitz function. Let*

$$Tu(\mathbf{x}) = \inf_{B\in\mathbb{B}} \sup_{\mathbf{y}\in g(\mathbf{x})B} u(\mathbf{x}+\mathbf{y}).$$

Then if u is L-Lipschitz, Tu is LM-Lipschitz. The same result is true with a sup inf *instead of an* inf sup.

Proof. This follows from Lemma 25.3 and the fact that an arbitrary infimum of L-Lipschitz functions also is Lipschitz with the same constant (see exercise 25.3.) □

Exercise 25.3. Prove that if $(u_i)_{i\in I}$ is a family of L-Lipschitz functions such that $|u_i(0)| \le C$ is bounded independently of i, then then $u(\mathbf{x}) = \inf_{i\in I} u_i(\mathbf{x})$ and $v(\mathbf{x}) = \sup_{i\in I} v(\mathbf{x})$ also are L-Lipschitz.
■

Let us now define the space-varying structuring elements naturally associated with the snake equation.

Approximation of $-DgDu$. We consider the family made of a single element:

$$\mathbb{B}_h(\mathbf{x}) = \{\{hDg(\mathbf{x})\}\}$$

By Taylor formula we then have, for each point where u is C^2:

$$(S_h u)(\mathbf{x}) = \inf_{B\in\mathbb{B}_h(\mathbf{x})} \sup_{\mathbf{y}\in B} u(\mathbf{x}+\mathbf{y}) = u(\mathbf{x}+hDg)$$

$$= u(\mathbf{x}) + hDg(\mathbf{x}).Du(\mathbf{x}) + O(\mathbf{x},h^2),$$

where $O(\mathbf{x},h^2)$ converges uniformly on every compact set K where u is C^2. We can rewrite the last relation

$$(S_h u)(\mathbf{x}) - u(\mathbf{x}) = hDg(\mathbf{x}).Du(\mathbf{x}) + O(\mathbf{x},h^2). \tag{25.10}$$

Approximation of $g|Du|curv(u)$. We consider the structuring elements of the median filter (See Chapter 10):

$$\mathbb{B}'_h(\mathbf{x}) = \{B \mid B \subset B(0, \sqrt{6g(\mathbf{x})h}) \text{ and } meas(B) \geq 3\pi g(\mathbf{x})h)\}.$$

Set

$$(S'_h u)(\mathbf{x}) = \sup_{B \in \mathbb{B}'_h(\mathbf{x})} \inf_{\mathbf{y} \in B} u(\mathbf{x}+\mathbf{y}).$$

Thanks to Theorem 14.7, we have for any $u : \mathbb{R}^2 \to \mathbb{R}$ which is C^2:

(i) On every compact set $K \subset \{\mathbf{x} \mid Du(\mathbf{x}) \neq 0\}$,

$$S'_h u(\mathbf{x}) = u(\mathbf{x}) + hg(\mathbf{x})|Du(\mathbf{x})|\text{curv}(u)(\mathbf{x}) + O(\mathbf{x}, h^{\frac{3}{2}}), \tag{25.11}$$

where $|O(\mathbf{x}, h^{\frac{3}{2}})| \leq C_K h^{\frac{3}{2}}$ for some constant C_K that depends only on u and K.

(ii) On every compact set K in $\mathbb{R}^2$,

$$|S'_h u(\mathbf{x}) - u(\mathbf{x})| \leq C_K h \tag{25.12}$$

where the constant C_K depends only on u and K.

Alternating the two filters. We now consider $T_h = S_h S'_h$, the alternate filter whose iteration should mimic the snake equation. Using (25.10), (25.11) and (25.12), Lemma 18.5 ensures that for any compact set K where $|Du(\mathbf{x})| \neq 0$,

$$(T_h u)(\mathbf{x}) - u(\mathbf{x}) = h(g(\mathbf{x})Du(\mathbf{x})curv(u)(\mathbf{x}) + Dg(\mathbf{x})Du(\mathbf{x})) + O(\mathbf{x}, h^{\frac{3}{2}})$$

and for $\mathbf{x}$ in any compact set K:

$$(T_h u)(\mathbf{x}) - u(\mathbf{x}) = O(\mathbf{x}, h),$$

where in both cases the convergence of $O(\mathbf{x}, h)$ is uniform on K. As a consequence the filter T_h is uniformly consistent (see Definition 19.15) with the PDE (25.9).

Exercise 25.4. Check carefully that in the above argument, Lemma 18.5 applies. ■

Construction of the approximate solutions. We consider a L-Lipschitz initial function u_0. We then define $u_h(t, \mathbf{x})$ for every $h > 0$ by

$$\forall n \in \mathbb{N}, \quad u_h(nh, \mathbf{x}) = (T_h^n u_0)(\mathbf{x}).$$

From now on we shall assume that

$$Dg(\mathbf{x}) \text{ and } \sqrt{6g(\mathbf{x})} \text{ are 1-Lipschitz and bounded by 1.} \tag{25.13}$$

By Corollary 25.4, the 1-Lipschitz assumptions on g and $\sqrt{6g(\mathbf{x})}$ ensure that $(T_h)^n u_0$ is L-Lipschitz for every n. All of these bounds can be achieved by simple scaling, namely multiplying g by a small enough constant.

Uniform continuity in t of the approximate solutions.

Lemma 25.5. *The approximate solutions of the snake equation $u_h(nh, \mathbf{x}) = (T_h^n u_0)(\mathbf{x})$ are uniformly continuous in t. More precisely:*

$$\forall t, \forall n \mid nh \leq t, \qquad -(L\sqrt{2t} + Lt) \leq T_h^n u - u \leq +(L\sqrt{2t} + Lt).$$

Proof. Let us bound the operator T_h by two isotropic and translation invariant operators. For every L-Lipschitz function u, one has

$$u(\mathbf{x}) - Lh|Dg| \leq (S_h u)(\mathbf{x}) = u(\mathbf{x} + hDg) \leq u(\mathbf{x}) + Lh|Dg|$$

$$u(\mathbf{x}) - Lh \leq (S_h u)(\mathbf{x}) \leq u(\mathbf{x}) + Lh$$

Then due to the fact that $S_h'(u+c) = S_h'(u) + c$ for any constant c, we also have

$$(S_h' u) - Lh \leq T_h u = S_h' S_h u \leq (S_h' u) + Lh \tag{25.14}$$

Let us consider $v(\mathbf{y}) = L|\mathbf{x} - \mathbf{y}|$. The family $\mathbb{B}'_{\sqrt{6g(\mathbf{x})h}}$ of structuring elements of the filter S_h' is made of the subsets of the disk of center 0 and radius $\sqrt{6g(\mathbf{x})h} \leq \sqrt{h}$. It is easy to check that for any $B \in \mathbb{B}'_{\sqrt{6g(\mathbf{x})h}}$ satisfying $|B| \geq 3\pi g(\mathbf{x})h$, there exists $B' \in \mathbb{B}'_{\sqrt{h}}$ satisfying $|B| \geq \frac{\pi h}{2}$, such that

$$\inf_{\mathbf{y} \in B} v(\mathbf{x} - \mathbf{y}) = \inf_{\mathbf{y} \in B'} v(\mathbf{x} - \mathbf{y}).$$

Thus,

$$(S_h' v)(\mathbf{x}) = \sup_{B \in \mathbb{B}'_{\sqrt{6g(\mathbf{X})h}}} \inf_{\mathbf{y} \in B} v(\mathbf{x} - \mathbf{y}) \leq \sup_{B \in \mathbb{B}'_{\sqrt{h}}} \inf_{\mathbf{y} \in B'} v(\mathbf{x} - \mathbf{y}).$$

This yields

$$\forall \mathbf{x},\ (S_h' v)(\mathbf{x}) \leq (M_h v)(\mathbf{x}) \tag{25.15}$$

where M_h denotes the median filter on the ball $B(0, \sqrt{h})$, as defined in Chapter 10. Similarly, for $w(\mathbf{y}) = -L|\mathbf{x} - \mathbf{y}|$, we have

$$\forall \mathbf{x},\ (M_h w)(\mathbf{x}) \leq (S_h' w)(\mathbf{x}) \tag{25.16}$$

We deduce from (25.15), (25.16) and (25.14) the inequalities

$$(T_h v)(\mathbf{x}) \leq (M_{6h} v)(\mathbf{x}) + Lh \qquad (M_{6h} w)(\mathbf{x}) - Lh \leq (T_h w)(\mathbf{x}) \tag{25.17}$$

By monotonicity of T_h and of the median operator, we thus have for all $n \in \mathbb{N}$, and for all $\mathbf{x} \in \mathbb{R}^2$,

$$(T_h^n v)(\mathbf{x}) \leq (M_h^n v)(\mathbf{x}) + nLh \qquad (M_h^n w)(\mathbf{x}) - nLh \leq (T_h^n w)(\mathbf{x}) \tag{25.18}$$

Now, since u_0 is L-Lipschitz, one has

$$u(\mathbf{x}) - L|\mathbf{x} - \mathbf{y}| \leq u(\mathbf{y}) \leq u(\mathbf{x}) - L|\mathbf{x} - \mathbf{y}|.$$

Thus

$$u(\mathbf{x}) + (T_h^n w)(\mathbf{x}) \leq (T_h^n u)(\mathbf{x}) \leq u(\mathbf{x}) + (T_h^n v)(\mathbf{x})$$

Using (25.18), we obtain

$$(M_h^n w)(\mathbf{x}) - nLh \leq (T_h^n u)(\mathbf{x}) - u(\mathbf{x}) \leq (M_h^n v)(\mathbf{x}) + nLh$$

Lemma 20.5 tells us that for h small enough and for $nh \leq t$, one has

$$(M_h^n v)(\mathbf{x}) \leq L\sqrt{2t}.$$

An analogous inequality obviously holds for w, so that for h small enough

$$\forall t, \forall n; nh \leq t \qquad -(L\sqrt{2t} + Lt) \leq T_h^n u - u \leq +(L\sqrt{2t} + Lt)$$

□

Exercise 25.5. Give all details for the proof of the property used in the above proof: For any $B \in I\!B'_{\sqrt{6g(\mathbf{x})h}}$ satisfying $|B| \geq 3\pi g(\mathbf{x})h$, there exists $B' \in I\!B'_{\sqrt{h}}$ satisfying $|B| \geq \frac{\pi h}{2}$, such that $\inf_{\mathbf{y}\in B} v(\mathbf{x}-\mathbf{y}) = \inf_{\mathbf{y}\in B'} v(\mathbf{x}-\mathbf{y})$. ■

Convergence of the approximate solutions

Theorem 25.6. *Let g be a C^2 function which is zero outside a ball $B(0,R)$ and satisfies the bounds (25.13). Then for every Lipschitz function $u_0 \in \mathcal{F}$, there exists a unique viscosity solution $u(t,\mathbf{x})$ of the snake equation*

$$\frac{\partial u}{\partial t} = F(D^2u, Du, \mathbf{x}) = g|Du|curv(u) + Dg.Du \qquad u(0,\mathbf{x}) = u_0(\mathbf{x}).$$

In addition, $u(t,\mathbf{x})$ is Lipschitz in $\mathbf{x}$ and holderian in t and when h tends to 0 and $nh \to t$, $(T_h^n u_0)(\mathbf{x})$ converges towards $u(t,\mathbf{x})$ uniformly on compact sets of $\mathbb{R}^+ \times \mathbb{R}^2$.

Proof. The operator T_h is monotone and local (and therefore satisfies the uniform local comparison principle.) It is uniformly consistent with the PDE (25.9) and commutes with the addition of constants. By Corollary 25.4 and Lemma 25.5, its associated approximate solutions $h \to u_h(t,\mathbf{x})$ are L-Lipschitz in $\mathbf{x}$ and uniformly Holderian in t for any initial Lipschitz function u_0. Thus, using Ascoli-Arzela theorem, there is a sub-sequence of the sequence $h \to u_h$ which is uniformly converging on every compact set towards a function $u(t,\mathbf{x})$. By Proposition (19.14), this implies that u is a viscosity solution of (25.9). In other words, we get the existence of a viscosity solution for any initial Lipschitz function u_0. Since this solution is unique, all subsequences of u_h converge to the same function u and therefore the whole sequence u_h converges to u. □

25.3 Back to shape evolution

We now consider the operator T_t which associates to any Lipschitz function u_0 in $\mathcal{F}$ the unique viscosity solution $u(t,.)$ of (25.9) with initial condition u_0.

T_t is clearly a monotone operator as limit of monotone operators. According to Proposition 20.2, T_t is also contrast invariant. Let us check that it also is standard monotone. We have assumed that g is zero outside a ball $B(0, R)$. It is a straightforward deduction that $T_h^n u_0(\mathbf{x}) = u_0(\mathbf{x})$ for every $\mathbf{x}$ with norm larger than R. Since $T_h^n u_0$ converges uniformly to $T_t u_0$ on compact sets as $nh \to t$ and $n \to \infty$, we still have $T_t u_0(\mathbf{x}) = u_0(\mathbf{x})$ outside $B(0, R)$ and therefore $T_t u_0(\infty) = u_0(\infty)$. Thus $T_t u_0$ belongs to $\mathcal{F}$ and T_t is standard monotone.

Exercise 25.6. Check in detail the above two statements, that T_t is monotone and contrast invariant. The second statement can be proven as indicated by using Proposition 20.2, but also directly by using the contrast invariance of the iterated operators T_h^n which converge to T_t as $nh \to t$. ■

Proposition 25.7. *By direct application of the level set extension theorem 7.19, the monotone and contrast invariant image operator T_t, defined for any initial Lipschitz function u_0 in $\mathcal{F}$, defines a unique set operator $\mathcal{T}_t$ on defined on the set $\mathcal{L}$ of the compact sets of S_N. Then $\mathcal{T}_t$ is monotone, T_t and $\mathcal{T}_t$ satisfy the commutation with thresholds $\mathcal{T}_t(\mathcal{X}_\lambda u) = \mathcal{X}_\lambda(T_t u)$ for all $\lambda \in \mathbb{R}$, T_t is the stack filter associated with $\mathcal{T}_t$ and $\mathcal{T}_t$ is upper semicontinuous on $\mathcal{L}$. In addition, since T_t is standard, so is $\mathcal{T}_t$.*

Exercise 25.7. Theorem 7.19 applies to an operator T defined on $\mathcal{F}$. Now, we have defined T_t on the Lipschitz functions of $\mathcal{F}$ only. In order to show that this is not a problem, prove first that any function u_0 in $\mathcal{F}$ can be approximated uniformly by a sequence of functions u_n which are C^1 and Lipschitz. Then show that that $T_t u_n$ is a Cauchy sequence for the uniform convergence and conclude that T_t can be extended into a contrast invariant standard monotone operator on all of $\mathcal{F}$. ■

The snake algorithm

Let us now see how we can the above results to define a curve evolution. Consider a closed curve $C = \mathbf{x}(s)$ surrounding a compact set K of $I\!R^2$. We define the generalized "curve" evolution of C by the following algorithm:

Step 1 Construct a Lipschitz function u_0 so that:

- $\mathcal{X}_0 u_0 = K$
- u_0 is Lipschitz.

Such a function u can be obtained by considering the signed distance function to the set K defined by $u(\mathbf{x}) = dist(\mathbf{x}, K^c)$ if $\mathbf{x} \in K$ and $u(\mathbf{x}) = \max(-dist(\mathbf{x}, K), -1)$ if $\mathbf{x} \in K^c$.

Step 2 Compute the viscosity solution $u(t, \mathbf{x})$ of equation (25.9) with initial condition u_0. We know one way to do it, by computing $u_h(nh, \mathbf{x}) = (T_h^n u_0)(\mathbf{x})$.

Step 3 Set $K(t) = \mathcal{X}_0 u(t, .)$ and $C(t) = K(t) \cap \bar{K(t)^c}$. ($C(t)$ is the topological boundary of $K(t)$.)

According to all preceding considerations, this algorithm defines for any curve C and for any $t \geq 0$ a unique set of points $C(t)$. The evolution of $C(t)$ is independent from the choice of the initial function u_0 and corresponds to a generalization of the curve evolution PDE(25.6). The preceding algorithm satisfies the *shape inclusion principle*: If initially C_1 and C_2 are two curves such that C_1 surrounds C_2, which means $K_2 \subset K_1$ then $C_1(t)$ surrounds $C_2(t)$: $K_2(t) \subset K_1(t)$.

However, $C(t)$ is not necessarily a curve of $I\!R^2$. It simply is the boundary of a set $K(t)$. It is therefore difficult to check if the initial geodesic snake energy estimated on $C(t)$ is decreasing. This problem is open! We have defined a very robust and weak evolution of curves and the initial C needs not be even a curve. Any set of curves or more generally any closed set can be taken as initial datum. The very good point of this generalized curve evolution is that it allows $C(t)$ to break into pieces surrounding separated shapes, as illustrated in Figure 25.4.

25.4 Implementation by finite difference scheme

As usual, the approximation of the process by iterated inf sup filter is not quite satisfactory, because these filters fail to be consistent with the equation at small scales, as pointed out for the median filter. Before proposing another way to simulate the snake equation, some heuristic comments of the snake equation behavior will be useful.

$$\frac{\partial u}{\partial t} = g|Du|curv(u) + Dg.Du$$

The first term is the well known mean curvature motion. As we have seen, it tends to shrink the level lines towards points. The speed of this motion is related to the amplitude of g. On an edge, g is small, but not zero. Thus, the motion is slowed down, but does not stop.

The second term is the erosion term. It tends to move the level lines of u downwards for g, that is, towards the edges of I (see figure 25.1), creating therefore shocks for u around edges since level lines of u converge to them on both sides. Contrarily to the first term, this term is not active on flat regions for g. Even worse, due to noise, little gradients for I will induce a non negligible variation of amplitude of g, resulting in non negligible Dg term with random direction. In others words, on flat regions, one can expect to observe random perturbations of the shape of the evolving contour.

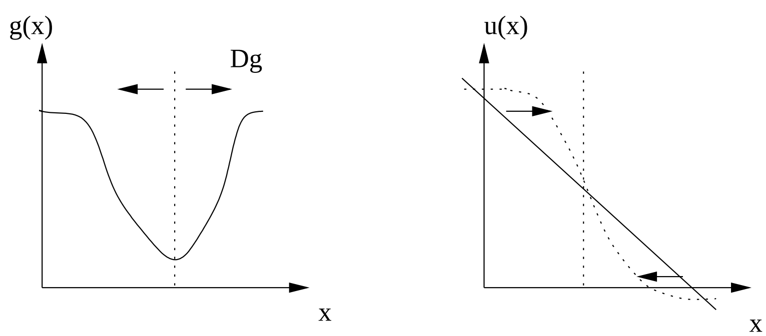

Figure 25.1: Convection term of the active contour equation. The convection term of the active contour equation tends to create around minima of g. Indeed, the level lines of u are moved in the direction opposite to the gradient of g.

More precisely, we have:

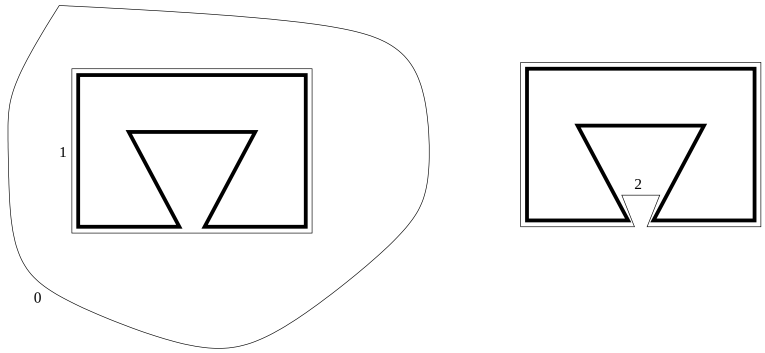

Figure 25.2: A difficulty : the local minima of the active contour energy. Left image: assume that g is null on the shape (drawn in bold) and that the initial contour is the line #0. From the initial line to the contour of the shape, the intermediate state #1 consisting of the convex hull of the polygon shows a smaller energy than the intermediate state #2 (drawn on the right). This illustrates the difficulties arising with the snake equation when we wish to land the active contour onto concave parts of the desired contour.

- **near an edge:** There is a risk that the curvature term pushes the level line over the edge, since it tends to shrink the curve. The second term instead moves the level line towards the edge. Thus the effects of both terms can be opposed. Modifications should be made in the equation to ensure that the second term always wins.
- **far from an edge:** The first term moves the level line fast, since $g(\mathbf{x})$ is high. The second term attracts the level line towards tiny edges of the image, thus creating little shocks. Here again, modifications should be made in the equation to let the second term win.

Even if the equation does not show any parameter weighting the two terms, a weight between them is in fact hidden in the choice of the function g. Finding a function g that makes the correct equilibrium both near an edge and far from an edge is somehow complex and case sensitive.

Assuming such an equilibrium could be found, relying on the single (weighted) mean curvature motion to shrink the level line is not a good idea. Indeed, a weighted mean curvature motion will never help transforming a convex level line into a general non-convex one. If we start with a circle, it is impossible to recover (e.g.) a star, as illustrated in Figure 25.2.

To cope with this problem, we shall add an extra term to the equation. This term is a classical erosion with the same weight as the curvature motion. In the experiments we therefore considered the modified snake equation

$$\frac{\partial u}{\partial t} = g|Du|(curv(u) - 1) + Dg.Du. \tag{25.19}$$

This new term is added under the assumption that the initial contour $C(0)$ has been drawn on the outside of the shape and that the initial function u_0 is positive inside the contour. Then the term -1 tends to erode the contour. If instead the initial contour was drawn inside the shape, then the new term should be +1. By an obvious adaptation of the proof of proposition **??** one sees that the associated curve evolution is

$$\frac{\partial \mathbf{x}}{\partial t}(t,s) = g(\mathbf{x}(t,s))\boldsymbol{\kappa}(\mathbf{x}(t,s)) - (Dg(\mathbf{x}(t,s)).\mathbf{n}(\mathbf{x}(t,s))).\mathbf{n}(\mathbf{x}(t,s))). \quad (25.20)$$

Returning to the snake equation, our finite difference scheme will the following:

$$\begin{aligned} u^{n+1}(\mathbf{x}) = \quad & u^n(\mathbf{x}) + dt(g(\mathbf{x})(E_1(u^n)(\mathbf{x}) + M_{\sqrt{6}}(u^n)(\mathbf{x}) - 2u^n(\mathbf{x}))) \\ & + dt(u^n(\mathbf{x} + Dg(\mathbf{x})) - u^n(\mathbf{x})), \end{aligned}$$

where $E_1 u(\mathbf{x}) = \inf_{\mathbf{y} \in B(0,1)} u(\mathbf{x}+\mathbf{y})$ denotes the erosion by the unit ball $B(0,1)$ and $M_{\sqrt{6}}$ is the median filter on the ball $B(0, \sqrt{6}$. The scheme is not contrast invariant, but is maximum decreasing and minimum increasing, provided $|2g(\mathbf{x}) + 1| \leq 1$.

Exercise 25.8. Prove the last statement, that the above scheme is maximum decreasing and minimum increasing, provided $|2g(\mathbf{x}) + 1| \leq 1$. Prove that the scheme is consistent, in the sense that if the pixel size h tends to zero then the second member of the equation tends to the second member of the modified snake equation 25.19. ■

Figure 25.3 illustrates the extraction of the bird shape on a textured background. This experiment illustrates well the complexity of the figure-background problem: the shape of the bird body has a quickly changing color from the white head to its dark tail. The background being uniform grey, there is no unique level line surrounding the whole shape. In fact, the gradient of the contour of the bird vanishes at many points. The fact that we "see" this contour is a classical illusion, called *subjective contour*. To uncover the illusion, the reader should scan small parts of the shape contour by using a white sheet with a small hole. Then he or she will realize that the contour seen globally has no complete numerical local evidence. This observation implies that no classical edge detection device would give out the whole contour. This can be checked by applying a Canny edge detector to the shape. To some extent, the snake method manages instead to surround the body shape. All the same, there is a risk that, because of the erosion term, the active contour goes through the subjective contour. If instead the curvature term is too strong, it can stop the contour before it reaches a concave corner of the shape. These facts explain the obvious inaccuracy and irregularity of the found contour.

25.5 Exercises

Exercise 25.9. In the whole exercise $u(t,\mathbf{x})$ and $\mathbf{x}(t)$ are supposed as smooth as needed to make the computations. Our aim is to interpret the equation

$$\frac{\partial u}{\partial t} = Dg.Du \quad (25.21)$$

as a motion of the level lines of u towards the minima of g. Let us consider a point $\mathbf{x}(t)$ on a level line of $u(t,\mathbf{x})$ with level λ and denote by $\mathbf{x}'(t)$ the motion vector of $\mathbf{x}(t)$ in the direction normal to the level line. We know that $\mathbf{x}(t)$ obeys the normal flow equation (12.4). Deduce from this equation and (25.21) that $\mathbf{x}(t)$ moves downwards in the landscape given by g, that is $g(\mathbf{x}(t))$ is a non-increasing function of t. ■

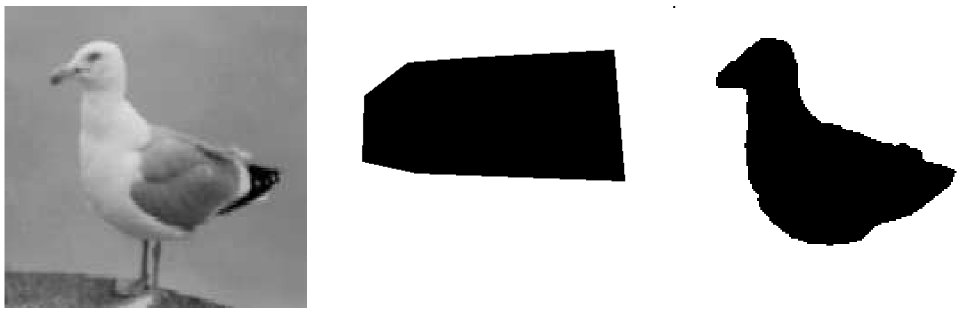

Figure 25.3: Silhouette of a bird by active contour. Left: original image, middle: initial contour, right: final contour (steady state of the snake equation). The contours of the bird body are partly subjective. The snake evolution manages to some extent to find them, but tends to indent the subjective contours and to round to the concave corners of the shape.

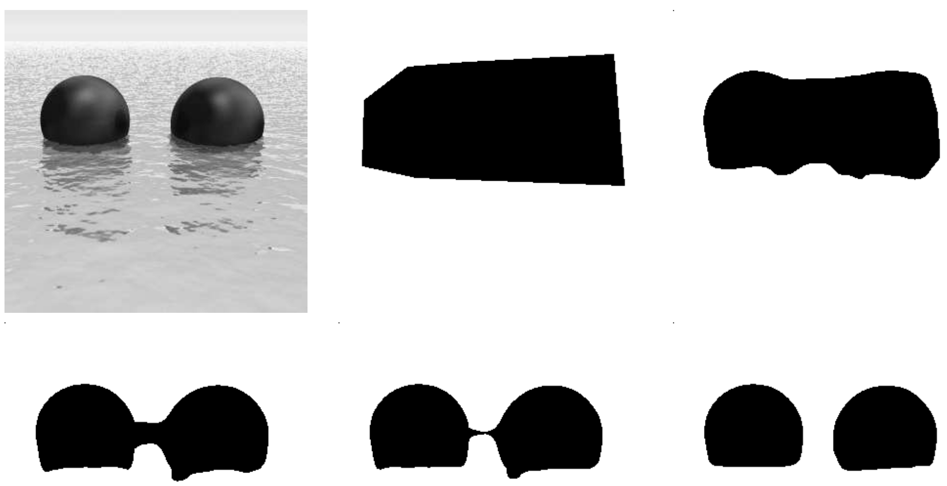

Figure 25.4: Active contour with topological change. Top, left: original image, middle: initial contour, right: intermediate state. Down, left and middle: successive intermediate states, down-right: final contour (steady state). This experiment shows that the level lines of u which bound the evolving contour cannot be classical solutions of the modified curve evolution equation (25.20). Indeed, the motion generates singularities when the contour splits. The generalized evolution provided by the viscosity solution of the snake equation yields more flexibility and allows an evolving curve to split. Original image is "Vue d'esprit 3", by courtesy of *e-on software*.

Exercise 25.10. Construction of another inf-sup scheme converging towards the viscosity solution of the equation (25.9). Consider the family of structuring elements

$$\mathbb{B}_h(\mathbf{x}) = \{B \mid B \subset B(\mathbf{x} + hDg(\mathbf{x}), \sqrt{6g(\mathbf{x})h}) \text{ and } meas(B) \geq 3\pi g(\mathbf{x})h\}.$$

and the operator

$$T_h u(\mathbf{x}) = \inf_{B \in \mathbb{B}_h} \sup_{\mathbf{y} \in B} u(\mathbf{x} + \mathbf{y})$$

1. Interpret the operator T_h as a shifted median filter.
2. Show that T_h is uniformly consistent with equation (25.9).
3. Show that the iteration of T_h converges towards a viscosity solution of (25.9).

■

Bibliography

[1] *Parallel implementations of AOS schemes: A fast way of nonlinear diffusion filtering.*

[2] S. Acton, *Multigrid anisotropic diffusion*, Image Processing, IEEE Transactions on, 7 (1998), pp. 280–291.

[3] E. H. Adelson and J. R. Bergen, *Spatiotemporal energy models for the perception of motion*, J. Opt. Soc. Amer. A, 2 (1985), pp. 284–299.

[4] G. J. Agin, *Representation and Description of Curved Objects*, PhD thesis, Stanford U., October 1972.

[5] F. Almgren, J. E. Taylor, and L. Wang, *Curvature-driven flows: A variational approach*, SIAM J. Control Optimization, 31 (1993), pp. 387–438.

[6] S. Altschuler, S. B. Angenent, and Y. Giga, *Mean curvature flow through singularities for surfaces of rotation*, J. Geom. Anal., 5 (1995), pp. 293–358.

[7] L. Alvarez, Y. Gousseau, and J.-M. Morel, *Scales in natural images and a consequence on their bounded variation norm*, in Scale-Space Theories in Computer Vision, M. Nielsen, P. Johansen, O. F. Olson, and J. Weickert, eds., Lecture Notes in Computer Science 1682, Springer, 1999, pp. 247–258.

[8] L. Alvarez, Y. Gousseau, and J.-M. Morel, *The size of objects in natural and artificial images*, in Advances in Imaging and Electron Physics, vol. 111, Academic Press, 1999, pp. 167–242.

[9] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, *Axiomatisation et nouveaux opérateurs de la morphologie mathématique.*, C. R. Acad. Sci. Paris Sér. I, 315 (1992), pp. 265–268.

[10] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, *Axiomes et équations fondamentales du traitement d'images. (Analyse multiéchelle et E.D.P.)*, C. R. Acad. Sci. Paris Sér. I, 315 (1992), pp. 135–138.

[11] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal., 16 (1993), pp. 200–257.

[12] L. Alvarez, P.-L. Lions, and J.-M. Morel, *Image selective smoothing and edge detection by nonlinear diffusion (ii)*, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.

[13] L. Alvarez and F. Morales, *Affine morphological multiscale analysis of corners and multiple junctions*, Int. J. Comput. Vision, 25 (1997), pp. 95–107.

[14] L. Alvarez and J.-M. Morel, *Formalization and computational aspects of image analysis*, Acta Numerica, (1999), pp. 1–59.

[15] L. Ambrosio, V. Caselles, S. Masnou, and J.-M. Morel, *The connected components of caccioppoli sets and applications to image processing*, J. Eur. Math. Soc., 3 (2001), pp. 213–266.

[16] L. Ambrosio, V. Caselles, S. Masnou, and J.-M. Morel, *Connected components of sets of finite perimeter and applications to image processing*, J. Eur. Math. Soc., 3 (2001), pp. 39–92.

[17] P. Anandan, *A computational framework and an algorithm for the measurement of visual motion*, Int. J. Comput. Vision, 2 (1989), pp. 283–310.

[18] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazon, *Minimizing total variation flow*, C. R. Acad. Sci. Paris Sér. I, 331 (2000), pp. 867–872.

[19] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazon, *Minimizing total variation flow*, Differential and Integral Equations, 4 (2001), pp. 321–360.

[20] S. B. Angenent, *Some recent results on mean curvature flow*, in Recent advances in partial differential equations, M. A. H. et al., ed., 1994.

[21] S. B. Angenent, G. Sapiro, and A. Tannenbaum, *On the affine heat flow for nonconvex curves*, J. Amer. Math. Soc., 11 (1998), pp. 601–634.

[22] S. B. Angenent and J. J. L. Velazquez, *Degenerate neckpinches in mean curvature flow*, J. Reine Angew. Math., 482 (1997), pp. 15–66.

[23] G. R. Arce and N. Gallagher, *Stochastic analysis for the recursive median filter process*, IEEE Trans. Information Theory, 34 (1988), pp. 669–679.

[24] G. R. Arce and M. P. McLoughlin, *Theoretical analysis of the max/median filter*, 35 (1987), pp. 60–69.

[25] H. Asada and M. Brady, *The curvature primal sketch*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986), pp. 2–14.

[26] J. Astola, P. Heinonen, and Y. Neuvo, *On root structures of median and median-type filters*, IEEE Trans. Acoust. Speech Signal Process., 35 (1987), pp. 1199–1201.

[27] F. ATTNEAVE, *Some informational aspects of visual perception.*, Psychol Rev, 61 (1954), pp. 183–93.

[28] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, *Uniqueness of the Gaussian kernel for scale-space filtering*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986).

[29] B. Baldwin, D. Geiger, and R. A. Hummel, *Resolution-appropriate shape representation*, in International Conference on Computer Vision, 1998, pp. 460–465.

[30] C. Ballester, V. Caselles, and P. Monasse, *The tree of shapes of an image*, tech. rep., CMLA, Ecole Normale Supérieure-Cachan, France, 2001. Available at `http://www.cmla.ens-cachan.fr/Cmla/index.html`.

[31] C. Ballester, E. Cubero-Castan, M. González, and J.-M. Morel, *Fast contrast invariant image intersection*. Preprint CEREMADE, 1998.

[32] J. A. Bangham, P. Ling, and R. Young, *Multiscale recursive medians, scale-space, and transforms with applications to image processing*, IEEE Trans. Image Process., 5 (1996).

[33] G. Barles.

[34] G. Barles, *Remarks on a Flame Propagation Model*, Institut national de recherche en informatique et en automatique, 1985.

[35] G. Barles, *Interior gradient bounds for the mean curvature equation by viscosity solutions methods*, Differential and Integral Equations, 4 (1991).

[36] G. Barles and C. Georgelin, *A simple proof of convergence for an approximation scheme for computing motions by mean curvature*, SIAM J. Numer. Anal., 32 (1995), pp. 484–500.

[37] G. Barles and P. M. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Analysis, 4 (1991), pp. 271–283.

[38] P. R. Beaudet, *Rotationally invariant image operators*, in Proc. International Conference on Pattern Recognition, 1978, pp. 579–583.

[39] G. Bellettini and M. Novaga, *Comparison results between minimal barriers and viscosity solutions for geometric evolutions*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. IV, 26 (1998), pp. 97–131.

[40] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger, *Robust anisotropic diffusion*, IEEE Trans. Image Process., 7 (1998), pp. 421–432.

[41] J. Blom, B. M. te Haar Romeny, and J. J. Koenderink, *Affine invariant corner detection*, tech. rep., 3D Computer Vison Research Group, Utrecht University, NL, 1992.

[42] J. Blom, B. T. te Haar Romeny, A. Bel, and J. J. Koenderink, *Spatial derivatives and the propagation of noise in Gaussian scale-space*, Tech. Rep. 3DCV 91-03, U. of Utrecht, Dept. of Medical and Physiological Physics, 1991.

[43] S. K. Bose, K. K. Biswas, and S. K. Gupta, *Model-based object recognition: The role of affine invariants*, AIEng, 10 (1996), pp. 227–234.

[44] K. A. Brakke, *The motion of a surface by its mean curvature*, Mathematical Notes, 20, Princeton University Press, Princeton, NJ, 1978.

[45] L. Breiman, *Probability*, Addison-Wesley, 1968.

[46] H. Brezis, *Analyse Fonctionnelle, Théorie et Applications*, Masson, Paris, 1983.

[47] D. R. K. Brownrigg, *The weighted median filter*, Communications of the ACM, 27 (1984), pp. 807–818.

[48] A. M. Bruckstein and D. Shaked, *On projective invariant smoothing and evolutions of planar curves and polygons*, Journal of Mathematical Imaging and Vision, 7 (1997), pp. 225–240.

[49] P. J. Burt and E. H. Adelson, *The laplacian pyramid as a compact image code*, IEEE Trans. Comm., 31 (1983), pp. 532–540.

[50] A. R. Butz, *A class of rank order smoothers*, IEEE Trans. Acoust. Speech Signal Process., 34 (1986), pp. 157–165.

[51] J. F. Canny, *A computational approach to edge detection*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986), pp. 679–698.

[52] F. Cao, *Absolutely minimizing lipschitz extension with discontinuous boundary data*, C. R. Acad. Sci. Paris Sér. I, 327 (1998).

[53] F. Cao, *Partial differential equation and mathemetical morphology*, J. Math. Pures Appl., 77 (1998), pp. 909–941.

[54] F. Cao, J. Delon, A. Desolneux, P. Musé, and F. Sur, *A unified framework for detecting groups and application to shape recognition*, Rapport interne, 1746, pp. 1166–8687.

[55] F. Cao, Y. Gousseau, P. Muse, F. Sur, and J. Morel, *Accurate estimates of false alarm number in shape recognition*, tech. rep., Technical report, Cachan, France (http://www. cmla. enscachan. fr/Cmla/), 2004. 4, 5.

[56] R. A. Carmona and S. F. Zhong, *Adaptive smoothing respecting feature directions*, IEEE Trans. Image Process., 7 (1998), pp. 353–358.

[57] V. Caselles, F. Catté, T. Coll, and F. Dibos, *A geometric model for active contours*, Numerische Mathematik, 66 (1993), pp. 1–31.

[58] V. Caselles, B. Coll, and J.-M. Morel, *Topographic maps and local contrast changes in natural images*, International Journal of Computer Vision, volume=.

[59] V. Caselles, 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, 1996.

[60] V. Caselles, B. Coll, and J.-M. Morel, *Scale space versus topographic map for natural images*, in Scale Space, 1997.

[61] V. Caselles, B. Coll, and J.-M. Morel, *Topographic maps and local contrast changes in natural images*, Int. J. Comput. Vision, 33 (1999), pp. 5–27.

[62] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, Int. J. Comput. Vision, 1 (1997), pp. 61–79.

[63] V. Caselles, J. L. Lisani, J.-M. Morel, and G. Sapiro, *Shape preserving local histogram modification*, 1999.

[64] V. Caselles, J.-M. Morel, and C. Sbert, *An axiomatic approach to image interpolation*, IEEE Trans. Image Process., 7 (1998), pp. 376–386.

[65] V. Caselles, G. Sapiro, and D. H. Chung, *Vector median filters, vector morphology, and coupled PDE's: Theoretical connections*, J. Math. Imaging Vision, 12 (2000), pp. 109–120.

[66] V. Caselles and C. Sbert, *What is the best causal scale space for 3D images?*, SIAM, 56 (1996).

[67] 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 J. Numer. Anal., volume=.

[68] M. H. Chen and P. F. Yan, *A multiscaling approach based on morphological filtering*, IEEE Trans. Patt. Anal. Mach. Intell., 11 (1989), pp. 6794–700.

[69] R. C. Chen and P. T. Yu, *On the optimal design of rational rank selection filters for image restoration*, in Proceedings International Conference on Image Processing, 1996, p. 16A9.

[70] Y. G. Chen, Y. Giga, and S. Goto, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, Proc. Japan Acad. Ser. A Math. Sci., 65 (1989), pp. 207–210.

[71] Y. G. Chen, Y. Giga, and S. Goto, *Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations*, J. Differential Geom., 33 (1991), pp. 749–786.

[72] F. Cheng and A. N. Venetsanopoulos, *An adaptive morphological filter for image processing*, IEEE Trans. Image Process., 1 (1992), pp. 533–539.

[73] Y. Cheng, *Analysis of affine invariants as approximate perspective invariants*, Computer Vision and Image Understanding, 63 (1996), pp. 197–207.

[74] B. Chow, *Deforming convex hypersurfaces by the nth root of the gaussian curvature*, J. Differential Geom.

[75] B. Chow, *Deforming convex hypersurfaces by the square root of the scalar curvature*, Invent. Math., 87 (1987), pp. 63–82.

[76] T. Cohignac, C. Lopez, and J.-M. Morel, *Integral and local affine invariant parameter and application to shape recognition*, in Proceedings International Conference on Pattern Recognition, year=.

[77] R. Courant and F. John, *Introduction to Calculus and Analysis*, Springer, 2000. Reprint of the 1989 edition.

[78] I. Cox, S. Roy, and S. Hingorani, *Dynamic histogram warping of image pairs for constant imagebrightness*, Image Processing, 1995. Proceedings., International Conference on, 2 (1995).

[79] E. Coyle, N. Gallagher, S. Bass, J. Fitch, and R. Harber, *Digital filtering by threshold decomposition*, 1987. U.S. Patent 4,868,773.

[80] E. J. Coyle, *Rank order operators and the mean absolute error criterion*, IEEE Trans. Acoust. Speech Signal Process., 36 (1988), pp. 63–76.

[81] M. Crandall and P. Lions, *Convergent difference schemes for nonlinear parabolic equations and mean curvature motion*, Numerische Mathematik, 75 (1996), pp. 17–41.

[82] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27 (1992), pp. 1–67.

[83] J. Damon, *Generic structure of two dimensional images under Gaussian blurring*, SIAM J. Appl. Math., 59 (1998), pp. 97–138.

[84] P. E. Danielsson, *Getting the median faster*, Computer Graphics Image Processing, 17 (1981), pp. 71–78.

[85] J. G. Daugman, *Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by 2D visual cortical filters*, Journal of the Optical Society of America, 2 (1985), pp. 1160–1169.

[86] E. R. Davies, *On the noise suppression and image enhancement characteristics of the median, truncated median and mode filters*, Pattern Recognition Letters, 7 (1988), pp. 87–97.

[87] J. Delon, *Midway Image Equalization*, Journal of Mathematical Imaging and Vision, 21 (2004), pp. 119–134.

[88] J. Delon, *Movie and video scale-time equalization application to flicker reduction*, Image Processing, IEEE Transactions on, 15 (2006), pp. 241–248.

[89] R. Deriche, *Using canny's criteria to derive a recursively implemented optimal edge detector*, Int. J. Comput. Vision, 1 (1987), pp. 167–187.

[90] R. Deriche and G. Giraudon, *A computational approach for corner and vertex detection*, International Journal of Computer Vision, 10 (1993), pp. 101–124.

[91] F. Dibos, *Affine length and affine dimension of a 1-set of the plane*, Proceedings of the Royal Society of Edinburgh, 126 (1996), pp. 985–993.

[92] G. Dudek, *Shape Representation from Curvature*, PhD thesis, U. Toronto, 1990.

[93] G. Dudek and J. K. Tsotsos, *Robustly recognizing curves using curvature-tuned smoothing*, in IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 35–41.

[94] C. L. Epstein and M. Gage, *The curve shortening flow*, in Wave motion: theory, modelling, and computation, Springer, NY, 1987, pp. 15–59. Proc. Conf. Hon. 60th Birthday P. D. Lax (Berkeley, CA, 1986) Math. Sci. Res. Inst. Publ. 7.

[95] L. C. Evans, *Nonlinear semigroup theory and viscosity solutions of hamilton-jacobi pde*, in Nonlinear Semigroups, Partial Differential Equations and Attractors year=.

[96] L. C. Evans, *Regularity for fully nonlinear elliptic equations and motion by mean curvature*, in Viscosity Solutions and Applications, I. C. D. et al., ed., Lect. Notes Math. 1660, springer, 1987, pp. 97–133. Lectures given at the 2nd session of the Centro Internazionale Matematico Estivo, Montecatini Terme, Italy, June 12-20, 1995.

[97] L. C. Evans, *Convergence of an algorithm for mean curvature motion*, Indiana Univ. Math. J., 42 (1993), pp. 553–557.

[98] L. C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, 19, American Mathematical Society, Providence, RI, 1998.

[99] L. C. Evans and J. Spruck, *Motion of level sets by mean curvature. I.*, J. Differential Geom., 33 (1991), pp. 635–681.

[100] L. C. Evans and J. Spruck, *Motion of level sets by mean curvature. II.*, Trans. Am. Math. Soc., 330 (1992), pp. 321–332.

[101] L. C. Evans and J. Spruck, *Motion of level sets by mean curvature. III.*, J. Geom. Anal., 2 (1992), pp. 121–150.

[102] L. C. Evans and J. Spruck, *Motion of level sets by mean curvature. IV.*, J. Geom. Anal., 5 (1995), pp. 79–116.

[103] F. Falzon and G. Giraudon, *Singularity analysis and derivative scale-space*, in IEEE Computer Vision and Pattern Recognition, 1994, pp. 245–250.

[104] O. 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, J. L. Mundy, A. Zisserman, and D. Forsyth, eds., Lecture Notes in Computer Science 825, Springer-Verlag, 1994, pp. 11–46.

[105] O. Faugeras and R. Keriven, *Affine curvature from affine scale-space*, in International Conferenc on Computer Vision, 1995. Submitted.

[106] O. Faugeras and R. Keriven, *Scale-spaces and affine curvature*, in Proceedings Europe-China Workshop on Geometrical modelling and Invariants for Computer Vision, vol. 1, 1995, pp. 17–24.

[107] O. Faugeras and R. Keriven, *Some recent results on the projective evolution of 2-D curves*, in International Conference in Image Processing, vol. 3, 1995, pp. 13–16.

[108] W. J. Firey, *Shapes of worn stones*, Mathematika, 21 (1974), pp. 1–11.

[109] J. P. Fitch, E. J. Coyle, and N. C. Gallagher, Jr., *Median filtering by threshold decomposition*, IEEE Trans. Acoust. Speech Signal Process., 32 (1984), pp. 1183–1188.

[110] L. M. J. Florack, *Image Structure*, vol. 10 of Computational Imaging and Vision, Kluwer Academic Publishers, Dordrecht, NL, 1997.

[111] 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, 1992, pp. 19–23.

[112] 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 (1992), pp. 376–388.

[113] 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, Utrecht, The Netherlands, 1993, pp. 21–24.

[114] 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 (1994), pp. 171–187.

[115] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, *Linear scale-space*, J. Math. Imaging Vision, 4 (1994), pp. 325–351.

[116] D. A. Forsyth, J. L. Mundy, and A. Zisserman, *Transformational invariance: A primer*, Image and Vision Computing, 10 (1992), pp. 39–45.

[117] D. Gabor, *Information theory in electron microscopy*, Journal of Laboratory Investigation, 14 (1965), pp. 801–807. The Journal of Laboratory Investigation was published by the International Academy of Pathology.

[118] M. Gage, *An isometric inequality with applications to curve shortening*, Duke Math. J., 50 (1983), pp. 1225–1229.

[119] M. Gage, *Curve shortening makes convex curves circular*, Invent. Math., 76 (1984), pp. 357–364.

[120] M. Gage and R. S. Hamilton, *The heat equation shrinking convex plane curves*, J. Differential Geom., 23 (1986), pp. 69–96.

[121] R. Geraets, A. H. r, B. M. ter Haar Romeny, and M. A. Viergever, *Affine scale-space for discrete pointsets*, in Proc. Soc. for Neural Networks, Nijmegen, NL, 1995.

[122] 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, Utrecht, the Netherlands, 1995, SION, pp. 86–97.

[123] P. J. Giblin and G. Sapiro, *Affine-invariant symmetry sets*, in Foundations of Computational Mathematics Conference, Instituto Nacional de Matemtica Pura e Aplicada, 1997.

[124] E. D. Giorgi, *New conjectures on flow by mean curvature*, in Nonlinear variational problems and partial differential equations. Proceedings of the 3rd conference held in the autumn of 1990 at Isola d'Elba, Italy. publisher=.

[125] E. D. Giorgi and L. Ambrosio, *Un nuovo tipo di funzionale del calcolo delle variazioni*, Atti. Accad. Naz. Lincei, s.8 (1988), pp. 199–210.

[126] M. J. E. Golay, *Smoothing data by least squares procedures and by filtering*, IEEE Trans. Computer, 21 (1972), pp. 299–301.

[127] R. Gonzalez and R. Woods, *Digital image processing*, Addison-Wesley Reading, Mass, 1987.

[128] M. A. Grayson, *The heat equation shrinks embedded plane curves to round points*, J. Differential Geom., 26 (1987), pp. 285–314.

[129] M. A. Grayson, *A short note on the evolution of a surface by its mean curvature*, Duke Math. J., 58 (1989), pp. 555–558.

[130] F. Guichard, *Axiomatisation des analyses multi-échelles d'images et de films - Axiomatization of images and movies scale-space.*, PhD thesis, Univ. Paris IX-Dauphine, 1994.

[131] F. Guichard, C. Lopez, and J.-M. Morel, *Une axiomatization de la vision préattentive*, Images des Mathématiques, (1995), pp. 48–57.

[132] F. Guichard and J. Morel, *Mathematical morphology almost everywhere*, Proceedings of ISMM, pp. 293–303.

[133] F. Guichard and J.-M. Morel, *Partial Differential Equations and Image Iterative Filtering*, 1995. Washington D.C.

[134] F. Guichard and J.-M. Morel, *Partial differential equations and image iterative filtering*, in The State of the Art in Numerical Analysis, I. S. D. et al., ed., Inst. Math. Appl. Conf. Ser., New Ser. 63, Institute of Mathematics and its Applications (IMA), Oxford Univ. Press, 1997, pp. 525–562.

[135] R. M. Haralick, *Digital step edges from zero-crossing of second directional derivatives*, IEEE Trans. Patt. Anal. Mach. Intell., 6 (1984), pp. 58–68.

[136] R. M. Haralick, S. R. Sternberg, and X. Zhuang, *Image analysis using mathematical morphology*, IEEE Trans. Patt. Anal. Mach. Intell., 9 (1987), pp. 532–550.

[137] R. C. Hardie and K. E. Barner, *Rank conditioned rank selection filters for signal restoration*, IEEE Trans. Image Process., 3 (1994), pp. 192–206.

[138] H. J. A. M. Heijmans, *Self-dual morphological operators and filters*, J. Math. Imaging Vision, 6 (1996), pp. 15–36.

[139] H. J. A. M. Heijmans and C. Ronse, *The algebraic basis of mathematical morphology: I. dilations and erosions*, Computer Vision, Graphics, and Image Processing, 50 (1990), pp. 245–295.

[140] C. W. Helstrom, *Image restoration by the method of least squares*, J. Opt. Soc. Amer., 57 (1967), pp. 297–303.

[141] J. M. Hereford and W. T. Rhodes, *Nonlinear optical image filtering by time-sequential threshold decomposition*, Opt. Eng., 27 (1988), pp. 274–279.

[142] H.Ishii, *A generalization of the bence, merriman and osher algorithm for motion by mean curvature*, in Proceedings of the international conference on curvature flows and related topics held in Levico, Italy, A. D. et al., ed., vol. 5 of GAKUTO Int. Ser. Math. Sci. Appl., Tokyo, 1994, Gakkotosho.

[143] B. K. P. Horn and E. J. Weldon, Jr., *Filtering closed curves*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986), pp. 665–668.

[144] C. T. Huang and O. R. Mitchell, *A Euclidean distance transform using grayscale morphology decomposition*, IEEE Trans. Patt. Anal. Mach. Intell., 16 (1994), pp. 443–448.

[145] T. S. Huang, G. J. Yang, and G. Y. Tang, *A fast two-dimensional median filtering algorithm*, IEEE Trans. Acoust. Speech Signal Process., 27 (1979), pp. 13–18.

[146] D. H. Hubel and T. N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, J. Physiol., 160 (1962), pp. 106–154.

[147] G. Huisken, *Flow by mean curvature of convex surfaces into spheres*, J. Differential Geom., 20 (1984), pp. 237–266.

[148] G. Huisken, *Singularities of the mean curvature flow*, in Tsing Hua lectures on geometry and analysis, S.-T. Yau, ed., Cambridge, MA, 1997, International Press, pp. 125–130. Lecture notes taken by Shun–Cheng Chang, Taiwan, 1990-91.

[149] R. A. Hummel, *Representations based on zero-crossing in scale-space*, in IEEE Computer Vision and Pattern Recognition, 1986, pp. 204–209.

[150] R. A. Hummel, *The scale-space formulation of pyramid data structures*, in PCV88, 1988, pp. 107–123.

[151] R. A. Hummel, B. Kimia, and S. W. Zucker, *Gaussian blur and the heat equation: Forward and inverse solutions*, in IEEE Computer Vision and Pattern Recognition, 1983, pp. 416–421.

[152] R. A. Hummel, B. Kimia, and S. W. Zucker, *Deblurring gaussian blur*, Computer Vision, Graphics, and Image Processing, 38 (1987), pp. 66–80.

[153] R. A. Hummel and R. Moniot, *Reconstructions from zero-crossings in scale-space*, IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 2111–2130.

[154] R. A. Hummel and H. J. Wolfson, *Affine invariant matching*, in DARPA88, 1988, pp. 351–364.

[155] T. Ilmanen, P. Sternberg, and W. P. Ziemer, *Equilibrium solutions to generalized motion by mean curvature*, J. Geom. Anal., 8 (1998), pp. 845–858.

[156] H. H. S. Ip, D. J. Potter, and D. S. Lebedev, *Impulse noise cleaning by iterative threshold median filtering*, Pattern Recognition Letters, 2 (1983), pp. 89–93.

[157] H. Ishii and P. Souganidis, *Generalized motion of noncompact hypersurfaces with velocity having arbitrary growth on the curvature tensor*, Tohoku Math. J. II, 47 (1995), pp. 227–250.

[158] P. T. Jackway and M. Deriche, *Scale-space properties of the multiscale morphological dilation erosion*, IEEE Trans. Patt. Anal. Mach. Intell., 18 (1996), pp. 38–51.

[159] S. Jaffard, Y. Meyer, and R. D. Ryan, *Wavelets: Tools for Science and Technology*, Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[160] R. Jensen, *The maximum principle for viscosity solutions of fully nonlinear second order partial differential equations*, Arch. Rational Mech. Anal., 101 (1988).

[161] B. Julesz, *A method of coding TV signals based on edge detection*, Bell System Tech. J. 38 (1959), pp. 1001–1020.

[162] J. Kacur and K. Mikula, *Solution of nonlinear diffusion appearing in image smoothing and edge detection*, Applied Numerical Mathematics, 17 (1995), pp. 47–59.

[163] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, Int. J. Comput. Vision, 1 (1998), pp. 321–331.

[164] P. Kempenaers, L. VanGool, and A. Oosterlinck, *Shape recognition under affine distortions*, in VF91, 1991, pp. 323–332.

[165] S. Kichenassamy, *The Perona–Malik paradox*, SIAM J. Appl. Math., 57 (1997), pp. 1328–1342.

[166] V. Kim and L. Yaroslavskii, *Rank algorithms for picture processing*, Computer Vision, Graphics, and Image Processing, 35 (1986), pp. 234–258.

[167] B. Kimia, A. Tannenbaum, and S. W. Zucker, *On the evolution of curves via a function of curvature. I. The classical case*, J. Math. Anal. Appl., 163 (1992), pp. 438–458.

[168] B. Kimia, A. R. Tannenbaum, and S. W. Zucker, *Shapes, shocks, and deformations I: The components of 2-dimensional shape and the reaction-diffusion space*, Int. J. Comput. Vision, 15 (1995), pp. 189–224.

[169] R. Kimmel and A. M. Bruckstein, *Shape offsets via level sets.*, Computer Aided Design, 5 (1993), pp. 154–162.

[170] R. Kimmel, D. Shaked, N. Kiryati, and A. M. Bruckstein, *Skeletonization via distance maps and level sets*, Computer Vision and Image Understanding, 3 (1995), pp. 382–391.

[171] J. J. Koenderink, *The structure of images*, Biological Cybernetics, 50 (1984), pp. 363–370.

[172] J. J. Koenderink, *A hitherto unnoticed singularity of scale-space*, IEEE Trans. Patt. Anal. Mach. Intell., 11 (1989), pp. 1222–1224.

[173] J. J. Koenderink and A. J. van Doorn, *Photometric invariants related to solid shape*, Optica Acta, 27 (1980), pp. 981–996.

[174] J. J. Koenderink and A. J. van Doorn, *Dynamic shape*, Biological Cybernetics, 53 (1986), pp. 383–396.

[175] L. S. G. Kovasznay and H. M. Joseph, *Image processing*, Proc. IRE, 43 (1955), p. 560.

[176] H. P. Kramer and J. B. Bruckner, *Iterations of a non-linear transformation for enhancement of digital images*, Pattern Recognition, 7 (1975).

[177] S. G. Krantz and H. R. Parks, *The Geometry of Domains in Space*, Birkäuser, 1999.

[178] A. S. Kronrod, *On functions of two variables*, Uspehi Matem. Nauk (N. S.), 5 (1950), pp. 24–134.

[179] P. D. Lax, *Numerical solution of partial differential equations*, Amer. Math. Monthly, 72 (1965), pp. 74–85.

[180] Y. H. Lee and S. A. Kassam, *Generalized median filtering and related nonlinear filtering techniques*, IEEE Trans. Acoust. Speech Signal Process., 33 (1985), pp. 672–683.

[181] R. Lenz, *Rotation-invariant operators*, in Proc. International Conference on Pattern Recognition, 1986, pp. 1130–1132.

[182] T. Lindeberg, *Scale-Space Theory in Computer Vision*, vol. 256 of Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht, NL, 1993.

[183] M. Lindenbaum, M. Fischer, and A. M. Bruckstein, *On gabor's contribution to image enhancement*, Pattern Recognition, 27 (1994), pp. 1–8.

[184] P.-L. Lions, *Axiomatic derivation of image processing models*, Math. Models Methods Appl. Sci., 4 (1994).

[185] J. Lisani, P. Monasse, and L. Rudin, *Fast Shape Extraction and Applications*, to appear, prepublication of CMLA ENS Cachan, 16 (2001).

[186] 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 the International Symposium of Mathematical Morphology, Palo Alto, CA, 2000. Available at `http://pascal.monasse.free.fr/index.html`.

[187] C. Liu, Z. Huan, and M. Tong, *Uniqueness of the viscosity solutions for the initial value problem of one type of second order partial differential equation*, J. Beijing Inst. Tech., Engl. Ed., 7 (1998), pp. 6–11.

[188] A. Mackworth and F. Mokhtarian, *Scale based description and recognition of planar curves and two-dimensional shapes*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986), pp. 34–43.

[189] A. K. Mackworth and F. Mokhtarian.

[190] J. Maeda, *A characteristic property of space curves of constant first affine curvature*, Tohoku Math. J., 48 (1941), pp. 148–151.

[191] R. Malladi, J. A. Sethian, and B. C. Vemuri, *Shape modeling with front propagation: A level set approach*, IEEE Trans. Patt. Anal. Mach. Intell., 17 (1995), pp. 158–175.

[192] P. Maragos, *A representation theory for morphological image and signal processing*, IEEE Trans. Patt. Anal. Mach. Intell., 11 (1989), pp. 586–599.

[193] P. Maragos, *Differential morphology and image-processing*, IEEE Trans. Image Process., 5 (1996), pp. 922–937.

[194] P. Maragos, *Partial differential equations in image analysis: Continuous modeling, discrete processing*, in Eusipco, 1998.

[195] P. Maragos and R. W. Schafer, *Morphological filters. Part I: Their set-theoretic analysis and relations to linear shift-invariant filters*, IEEE Trans. Acoust. Speech Signal Process., 35 (1987), pp. 1153–1169.

[196] P. Maragos and R. W. Schafer, *Morphological filters. Part II: Their relations to median, order-statistic, and stack filters*, IEEE Trans. Acoust. Speech Signal Process., 35 (1987), pp. 1170–1184.

[197] P. Maragos and R. D. Ziff, *Threshold superposition in morphological image analysis systems*, IEEE Trans. Patt. Anal. Mach. Intell., 12 (1990), pp. 498–504.

[198] D. Marr, *Vision*, W. H. Freeman and Company, New York, NY, 1982.

[199] D. Marr and E. Hildreth, *Theory of edge detection*, Proc. Roy. Soc. London Ser. B, 207 (1980), pp. 187–217.

[200] Mascarenhas, *Diffusion generated motion by mean curvature*, tech. rep., CAM Report 92–33, UCLA, 1992.

[201] S. Masnou, *Image restoration involving connectedness*, in Proceedings of the 6^{th} International Workshop on Digital Image Processing and Computer Graphics, 1998.

[202] G. Matheron, *Random Sets and Integral Geometry*, John Wiley & Sons, New York, NY, 1975.

[203] B. Merriman, J. Bence, and S. Osher, *Diffusion generated motion by mean curvature*, in Computational Crystal Growers Workshop. Proceedings of the Geometry Center Workshop held in Minneapolis, Minnesota, February 22–28, 1992., J. E. Taylor, ed., Selected Lectures in Mathematics, American Mathematical Society, 1992, pp. 73–83.

[204] B. Merriman, J. Bence, and S. Osher, *Motion of multiple junctions: a level set approach*, Journal of Computational Physics, 112 (1994), pp. 334–363.

[205] L. Moisan, *Affine plane curve evolution: A fully consistent scheme*, IEEE Trans. Image Process., 7 (1998), pp. 411–420.

[206] P. Monasse, *Représentation morphologique d'images numériques et application au recalage*, PhD thesis, Ecole Normale Supérieure-Cachan, France, 2000. French and English versions available at `http://pascal.monasse.free.fr`.

[207] P. Monasse and F. Guichard, *Scale-space from a level lines tree*, Journal of Visual Communication and Image Representation, 11 (2000), pp. 224–236. Available at `http://pascal.monasse.free.fr`.

[208] J. L. Mundy, *Object recognition: The search for representation*, in ORCV95, 1995, pp. 19–50.

[209] J. L. Mundy and A. Zisserman, eds., *Geometric Invariance in Computer Vision*, The MIT Press, Cambridge, MA, 1992.

[210] J. L. Mundy, A. Zisserman, and D. Forsyth, eds., *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science 825, Springer-Verlag, 1994.

[211] P. Muse, F. Sur, F. Cao, and Y. Gousseau, *Unsupervised thresholds for shape matching*, Image Processing, 2003. Proceedings. 2003 International Conference on, 2 (2003).

[212] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J. Morel, *An A Contrario Decision Method for Shape Element Recognition*, International Journal of Computer Vision, 69 (2006), pp. 295–315.

[213] M. Nagao and T. Matsuyama, *Edge preserving smoothing*, Computer Graphics Image Processing, 9 (1979), pp. 394–407.

[214] Y. Nakagawa and A. Rosenfeld, *A note on the use of local min and max operations in digital image processing*, IEEE Trans. Systems, Man and Cybernetics, 8 (1978), pp. 632–635.

[215] P. M. Narendra, *Noise reduction by median filtering*, in IEEE Conf. Pattern Recogn. Image Processing, 1978.

[216] W. Niessen, B. ter Haar Romeny, and M. Viergever, *Numerical Analysis of Geometry-Driven Diffusion Equations*, Geometry Driven Diffusion in Computer Vision, pp. 393–410.

[217] 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 (1997), pp. 187–205.

[218] E. Ochoa, J. P. Allebach, and D. W. Sweeney, *Optical median filtering by threshold decomposition*, Appl. Opt., 26 (1987), pp. 252–260.

[219] P. J. Olver, G. Sapiro, and A. Tannenbaum, *Classification and uniqueness of invariant geometric flows*, C. R. Acad. Sci. Paris Sér. I, 319 (1994), pp. 339–344.

[220] P. J. Olver, G. Sapiro, and A. Tannenbaum, *Differential invariant signatures and flows in computer vision: A symmetry group approach*, in Geometry Driven Diffusion in Computer Vision, Kluwer, 1994.

[221] P. J. Olver, G. Sapiro, and A. Tannenbaum, *Affine invariant gradient flows*, in 12th international conference on analysis and optimization of Systems:Images, Wavelets and PDE's, Springer-Verlag, 1996.

[222] P. J. Olver, G. Sapiro, and A. Tannenbaum, *Invariant geometric evolutions of surfaces and volumetric smoothing*, SIAM J. Appl. Math., 57 (1997), pp. 176–194.

[223] S. Osher and L. I. Rudin, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., 27 (1990), pp. 919–940.

[224] S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on the Hamilton–Jacobi formulation*, J. Comput. Phys., 79 (1988), pp. 12–49.

[225] H. C. Park and R. T. Chin, *Decomposition of arbitrarily-shaped morphological structuring elements*, IEEE Trans. Patt. Anal. Mach. Intell., 17 (1995), pp. 2–15.

[226] D. Pasquignon, *Computation of skeleton by PDE*, in International Conference in Image Processing, IEEE, ed., 1995.

[227] D. Pasquignon, *Approximation of viscosity solution by morphological filters*, in ESAIM COCV, vol. 4, 1999.

[228] E. J. Pauwels, L. J. VanGool, P. Fiddelaers, and T. Moons, *An extended class of scale-invariant and recursive scale space filters*, IEEE Trans. Patt. Anal. Mach. Intell., 17 (1995), pp. 691–701.

[229] S. C. Pei, C. L. Lai, and F. Y. Shih, *An efficient class of alternating sequential filters in morphology*, Graphical Models and Image Processing, 59 (1997), pp. 109–116.

[230] S. S. Perlman, S. Eisenhandler, P. W. Lyons, and M. J. Shumila, *Adaptive median filtering for impulse noise elimination in real-time TV signals*, IEEE Trans. Communications, 35 (1987), pp. 646–652.

[231] P. Perona and J. Malik, *Scale space and edge detection using anisotropic diffusion*, in Proc. Workshop on Computer Vision, Representation and Control, IEEE Computer Society, IEEE Press, Piscataway, NJ, 1987, pp. 16–22.

[232] P. Perona and J. Malik, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Patt. Anal. Mach. Intell., 12 (1990), pp. 629–639.

[233] F. Pollick and G. Sapiro, *Constant affine velocity predicts the 1/3 power law of planar motion perception and generation*, Vision Research, 37 (1997), pp. 347–353.

[234] M. Rahmati and L. G. Hassebrook, *Intensity-invariant and distortion-invariant pattern-recognition with complex linear morphology*, Pattern Recognition, 27 (1994), pp. 549–568.

[235] L. Rawtani, *Image smoothing by median filtering – an efficient algorithm*, J.M.A.C.T., 28 (1995), pp. 93–101.

[236] C. S. Regazzoni and A. Teschioni, *A new approach to vector median filtering based on space-filling curves*, IEEE Trans. Image Process., 6 (1997), pp. 1025–1037.

[237] C. H. Richardson and R. W. Schafer, *A lower bound for structuring element decompositions*, IEEE Trans. Patt. Anal. Mach. Intell., 13 (1991), pp. 365–369.

[238] F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Frederick Ungar Publishing Co., New York, 1955.

[239] C. Ronse and H. J. A. M. Heijmans, *The algebraic basis of mathematical morphology: II. openings and closings*, Computer Vision, Graphics, and Image Processing, 54 (1991), pp. 74–97.

[240] E. Rouy and A. Tourin, *A Viscosity Solutions Approach to Shape-From-Shading*, SIAM Journal on Numerical Analysis, 29 (1992), pp. 867–884.

[241] L. I. Rudin, *Images, Numerical Analysis of Singularities and Shock Filters*, PhD thesis, California Institute of Technology, Pasadena, CA, 1987.

[242] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.

[243] W. Rudin, *Real and Complex Analysis. Third edition*, McGraw-Hill Book Co., New York, 1987.

[244] A. H. Salden, B. M. ter Haar Romeny, and M. Viergever, *Affine and projective differential geometric invariants of space curves*, in Geometric Methods in Computer Vision II, B. Vemuri, ed., 1993, pp. 60–74.

[245] P. Salembier, *Structuring element adaptation for morphological filters*, Journal of Visual Communication and Image Representation, 3 (1992), pp. 115–136.

[246] P. Salembier and J. Serra, *Flat zones filtering, connected operators and filters by reconstruction*, IEEE Trans. Image Process., 8 (1995).

[247] G. Sapiro, A. Cohen, and A. M. Bruckstein, *A subdivision scheme for continuous scale B-splines and affine invariant progressive smoothing*, J. Math. Imaging Vision, 7 (1997), pp. 23–40.

[248] G. Sapiro, R. Kimmel, D. Shaked, B. Kimia, and A. M. Bruckstein, *Implementing continuous-scale morphology via curve evolution*, Pattern Recognition, 9 (1993), pp. 1363–1372.

[249] G. Sapiro and A. Tannenbaum, *Affine invariant scale space*, International Journal of Computer Vision, 11 (1993), pp. 25–44.

[250] G. Sapiro and A. Tannenbaum, *On invariant curve evolution and image analysis*, Indiana Univ. Math. J., 42 (1993), pp. 985–1009.

[251] G. Sapiro and A. Tannenbaum, *On affine plane curve evolution*, J. Funct. Anal., 119 (1994), pp. 79–120.

[252] D. Schonfeld, *Optimal structuring elements for the morphological pattern restoration of binary images*, IEEE Trans. Patt. Anal. Mach. Intell., 16 (1994), pp. 589–601.

[253] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, San Diego, CA, 1982.

[254] J. Serra, *Introduction to mathematical morphology*, Computer Vision, Graphics, and Image Processing, 35 (1986), pp. 283–305.

[255] J. Serra, *Image Analysis and Mathematical Morphology. Volume2: Theoretical Advances*, Academic Press, 1988.

[256] J. Sethian, *An analysis of flow propagation*, PhD thesis, UCLA, 1982.

[257] J. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Cambridge University Press, 1996.

[258] J. A. Sethian, *Curvature and the evolution of fronts*, Comm. Math. Phys., 101 (1985).

[259] F. Y. Shih and O. R. Mitchell, *A mathematical morphology approach to Euclidean distance transformation*, IEEE Trans. Image Process., 1 (1992), pp. 197–204.

[260] N. Sochen, R. Kimmel, and R. Malladi, *A general framework for low-level vision*, IEEE Trans. Image Process., 7 (1998), pp. 310–318.

[261] P. Soille, E. J. Breen, and R. Jones, *Recursive implementation of erosions and dilations along discrete lines at arbitrary angles*, IEEE Trans. Patt. Anal. Mach. Intell., 18 (1996), pp. 562–567.

[262] D. Struik, *Differential Geometry*, Addison-Wesley, 1950.

[263] T. Sun and Y. Neuvo, *Detail-preserving median based filters in image-processing*, Pattern Recognition Letters, 15.

[264] I. Svalbe and R. Jones, *The design of morphological filters using multiple structuring elements, part I: Openings and closings*, Pattern Recognition Letters, 13 (1992), pp. 123–129.

[265] H. D. Tagare and R. J. P. deFigueiredo, *Order filters*, Proceedings of IEEE, 73 (1985), pp. 163–165.

[266] M. E. Taylor, *Partial Differential Equations: Basic Theory*, Springer, New York, 1996.

[267] B. M. ter Haar Romeny, ed., vol. 1 of Computational Imaging and Vision, Kluwer Academic Publishers, 1994.

[268] B. M. ter Haar Romeny, *Front-End Vision and Multiscale Image Analysis: Introduction to Scale-Space Theory*, Kluwer Academic Publishers, Dordrecht, NL, In preparation.

[269] 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, Aalborg, DK, 1991, pp. 346–353.

[270] A. Toet, *A morphological pyramidal image decomposition*, Pattern Recognition Letters, 9 (1989), pp. 255–261.

[271] A. Toet, *A hierarchical morphological image decomposition*, Pattern Recognition Letters, 11 (1990), pp. 267–274.

[272] R. van den Boomgaard, *Numerical Solution Schemes for Continuous-Scale Morphology*, Proceedings of the Second International Conference on Scale-Space Theories in Computer Vision, (1999), pp. 199–210.

[273] R. van den Boomgaard and A. Smeulders, *The morphological structure of images: The differential equations of morphological scale-space*, IEEE Trans. Patt. Anal. Mach. Intell., 16 (1994), pp. 1101–1113.

[274] L. J. VanGool, T. Moons, E. Pauwels, and A. Oosterlinck, *Semi-differential invariants*, in Geometric Invariance in Computer Vision, J. L. Mundy and A. Zisserman, eds., The MIT Press, 1992, pp. 157–192.

[275] B. Vasselle, G. Giraudon, and M. Berthod, *Following corners on curves and surfaces in the scale space*, in Proc. Third European Conference on Computer Vision, 1994, pp. 109–114.

[276] L. Vincent, *Grayscale area openings and closings, their efficient implementation and applications*, in Proceedings of the 1^{st} Workshop on Mathematical Morphology and its Applications to Signal Processing, J. Serra and P. Salembier, eds., Barcelona, Spain, 1993, pp. 22–27.

[277] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton U. Press, Princeton, NJ, 1947.

[278] J. Weickert, *Nonlinear diffusion scale-spaces: from the continuous to the discrete setting*, Proc. ICAOS96, pp. 111–118.

[279] J. Weickert, *Theoretical foundations of anisotropic diffusion in image processing*, Computing Suppl., 11 (1996), pp. 221–236.

[280] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI Series, Teubner-Verlag, Stuttgart, 1998.

[281] J. Weickert and B. Benhamouda, *A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox*, in Advances in Computer Vision, F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, eds., Springer, 1997, pp. 1–10.

[282] J. Weickert, S. Ishikawa, and A. Imiya, *On the history of Gaussian scale-space axiomatics*, in Gaussian Scale-Space Theory, Kluwer, 1997, pp. 45–59.

[283] J. Weickert, S. Ishikawa, and A. Imiya, *Linear scale-space has first been proposed in japan*, J. Math. Imaging Vision, 10 (1999), pp. 237–252.

[284] J. Weickert, B. Romeny, and M. Viergever, *Efficient and reliable schemes for nonlinear diffusion filtering*, Image Processing, IEEE Transactions on, 7 (1998), pp. 398–410.

[285] I. Weiss, *Local projective and affine invariants*, AMAI, 13 (1995), pp. 203–225.

[286] P. D. Wendt, E. J. Coyle, and N. C. Gallagher, Jr., *Stack filters*, IEEE Trans. Acoust. Speech Signal Process., 34 (1986), pp. 898–911.

[287] M. Wertheimer, *Untersuchungen zur Lehre der Gestalt, II*, Psychologische Forschung, 4 (1923), pp. 301–350.

[288] J. Wilburn, *Developments in generalized ranked-order filters*, Journal of the Optical Society of America, 15 (1998), pp. 1084–1099.

[289] A. P. Witkin, *Scale-space filtering*, in Proceedings of the International Joint Conference on Artificial Intelligence, Karlsruhe, Germany, 1983, Morgan Kaufmann Publishers, San Francisco, pp. 1019–1021.

[290] 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*, IEEE Trans. Signal Process., 39 (1991), pp. 395–410.

[291] A. L. Yuille, *The creation of structure in dynamic shape*, in Proceedings Second International Conference on Computer Vision, 1988, pp. 685–689.

[292] A. L. Yuille and T. A. Poggio, *Fingerprint theorems for zero-crossings*, J. Opt. Soc. Amer., 2 (1985), pp. 683–692.

[293] A. L. Yuille and T. A. Poggio, *Scaling theorems for zero-crossings*, IEEE Trans. Patt. Anal. Mach. Intell., 8 (1986), pp. 15–25.

[294] R. Zhu, *Existence uniqueness and regularity for viscosity solutions of fully nonlinear parabolic equations under natural structure conditions. (Chinese. English summary)*, J. Beijing Norm. Univ., Nat. Sci., 30 (1994), pp. 290–300.

[295] X. Zhuang, *Decomposition of morphological structuring elements*, J. Math. Imaging Vision, 4 (1994), pp. 5–18.

[296] X. Zhuang and R. M. Haralick, *Morphological structural element decomposition*, Computer Vision, Graphics, and Image Processing, 35 (1986), pp. 370–382.