Chapter 3

Multi-scale diffusion

3.1 Scale within diffusion

Despite the fact that solutions to (2.9) offer an improvement over Gaussian blurring, they still suffer from the problem of the unreliability of the local gradient measures that are used in the conductance term. In situations where the noise is substantial, poor local gradient measurements can drive the process to undesirable results. Consider how solutions behave in the case of the blurred disk in Fig. 2.7 and additive Gaussian white noise with a standard deviation at each pixel [1] that is one-quarter of the intensity of the original disk.

Fig. 3.1 shows the original image and samples of the solutions for three distinct values of k. The algorithm described above assumes that the scale of individual pixels (sometimes referred to as "inner scale" (Koenderink 1990)) is a meaningful scale at which to make measurements of the gradient. The additive noise introduced into the blurred disk makes the pixel scale a particularly poor scale for such measurements. The result is that the process is unable to capture the larger scale regularity in the image.

Under certain conditions the anisotropic diffusion process appears to produce reliable measures of image gradient for the purpose of making decisions about boundaries. However, in order to carry out this process one needs a best estimate of the image gradient at each point in the image and at each point in time. The problem appears to be circular. In the absence of any a priori information about the image, the best estimate of the gradient (for the conductance determination) for which one could hope is the gradient measured with a Gaussian kernel. This suggests that (2.9) should contain a scale parameter s,

$$\nabla \cdot g(|\nabla_s f|)\nabla f \stackrel{\text{def}}{=} \nabla \cdot g(|(\nabla G(s)) \otimes f|)\nabla f = \frac{\partial f}{\partial t}, \tag{3.1}$$

in which "$G(s)\otimes$" denotes a convolution (over $\vec{x}$) of the image at time t with a Gaussian kernel of scale s. It is important to distinguish t and s in (3.1). The time parameter, t, characterizes a particular level of nonuniform blurring. The scale parameter, s, describes a level of uniform blurring used to make a gradient measurement of the image at some value of t. The uniform blurring associated with s is not a transformation applied to the image, but a description of the size of the test function that is used to measure the gradient at each point in a single time slice.

The presence of this smoothing kernel is of theoretical importance even in the case of continuous images. Catte, Coll, Lions and Morel (1990) argue that well-posedness of (2.9) cannot be guaranteed

[1] These images have dimensions of 256 × 256 pixels. Images through this paper have the same dimensions unless otherwise noted.

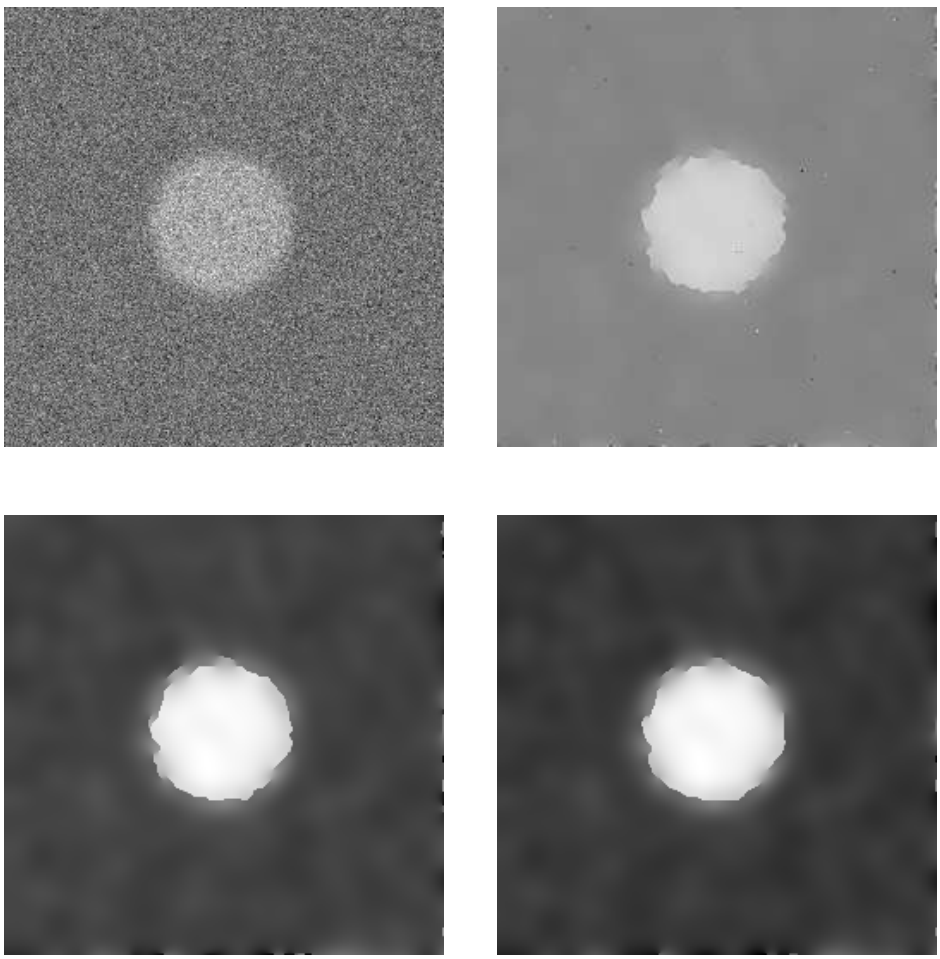

Figure 3.1: From the top left: A blurred disk with random noise (a) shows the susceptibility of anisotropic conductance diffusion. Results for $k =$ (b)0.5, (c)1.0, and (d)2.0 show that the large-scale structure is lost while the noise remains.

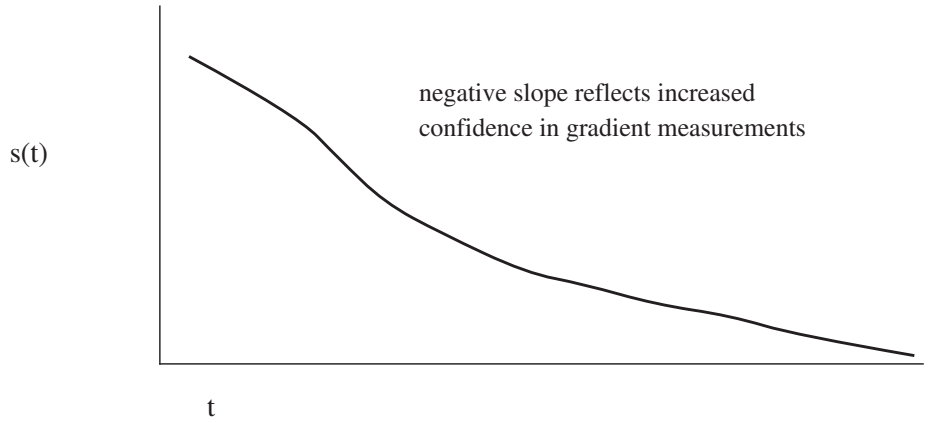

Figure 3.2: The scale parameter $s(t)$ should be some decreasing function of t, the *time* or *evolution* parameter.

in cases where $g(f_x)f_x$ is not monotonic. It is not monotonic for the conductance functions in (2.10). In such cases, however, unique solutions exist if a sufficiently smooth blurring kernel is included in the conductance term. The Gaussian kernel suffices for this purpose.

3.2 Scale as a function of evolution

Anisotropic diffusion is an iterative process that uses tentative estimates of the gradient to make incremental changes to the image. Because of unwanted, small-scale luminance fluctuations these tentative estimates are measured at some scale s. Recall that the scale s is the the size of the Gaussian kernel that is used to measure derivatives in images.

If the process works as we hope, the unwanted luminance fluctuations should diminish more rapidly than the signal that we wish to ultimately characterize. That is, gradient measurements should become more reliable as the process evolves. This suggests that one should not measure gradients at a single scale throughout the process but decrease the scale parameter s to reflect an increasing confidence in subsequent versions of the image. The argument above prescribes a diffusion equation of the form

$$\nabla \cdot g(|\nabla_{s(t)} f|)\nabla f = \frac{\partial f}{\partial t} \tag{3.2}$$

and suggests that $s(t)$, called the *scale recipe*, should be some decreasing function of the evolution parameter as in Fig. 3.2.

The result is a process that gradually "narrows in" on edges that belong to objects that do not blur away at some chosen scale. Consider the wide bands in the gradient image of Fig. 2.5. As the process continues and scale decreases, the bands become progressively narrower and the nonuniform blurring continues to smooth progressively closer to the boundary of the desired object. Thus, it is possible to retain very accurate descriptions of edges that belong to objects which are sufficiently large in the initial image. For noisy images it is important that $s(t)$ not decrease too quickly, or else structures that are not sufficiently large (noise) at some scale $s(t)$ might "resurface" at some scale $s(t + \delta t)$. Likewise, if $s(t)$ decreases slowly enough, edges that are sufficiently small at one scale $s(t)$ will be adequately smoothed so that they cannot reappear in measurements made at some later time. However, reducing scale too slowly could allow important information about large scale objects to be lost.

The above arguments do not prescribe a precise form for $s(t)$. Decisions about the rate at which to decrease scale $s(t)$ would likely depend on the noise level as well as the nature as the large-scale

structures in the image. It is conceivable that $s(t)$ might not be the same, or even have the same form, for every image but could vary according to some smoothness measurement made on the process itself. These notions can be quantified by considering a metric $\mathcal{M}(f(\vec{x},t),s)$ that measures the amount of noise (perhaps an average amount of noise) in the image at time t measured at scale s. If this metric is the relevant measure of noise, we could demand that the value of this metric should not increase as the process progresses. That is,

$$\frac{\partial \mathcal{M}}{\partial t} = \frac{\partial \mathcal{M}(f(\vec{x},t),s)}{\partial s}\frac{\partial s(t)}{\partial t} + \frac{\partial \mathcal{M}(f(\vec{x},t),s)}{t} \leq 0. \tag{3.3}$$

Equation (3.3) and the multi-scale diffusion equation (3.2) offer a pair of simultaneous equations for solving $s(t)$ and $f(\vec{x},t)$. However, prescribing such a metric $\mathcal{M}$ promises to be difficult. The issues of constructing suitable metrics and showing the existence of solutions are beyond the scope of this dissertation. Instead I will use some simple arguments to construct reasonable scale recipes, $s(t)$, and analyze the behavior of the multiscale diffusion process on that basis. Some examples scale recipes are discussed in Sect. 3.4, and the effects of various $s(t)$ reduction rates are examined experimentally in Chapt. 5.

Experiments show that the multiscale diffusion technique produces interesting results even with some very simple assumptions about the scale recipe. Figure 3.3 shows examples using the MRI head of Sect. 1.2, where the scale recipe $s(t)$ is a linearly decreasing function of t:

$$s(t) = s_0 - \alpha t. \tag{3.4}$$

Notice that with increasing s_0 the smaller structures such as the sulci become part of larger structures like the cerebral cortex, and yet the boundaries of those larger structures remain sharp and well defined. Figure 3.4 shows this processing on the noisy overlapping discs and the noisy blob that so confounds the "pixel-scale" version of this process.

3.2.1 Grouping and subjective boundaries

The choice of the initial value, s_0 and shape of $s(t)$ determine a lower limit on the size and intensity of objects or structures that remain in later stages of the diffusion process. This has a significant impact on the results of this type of processing. Its reasonable that s_0 should depend not only on the properties of the image but also on the task one wishes to perform. The reasoning underlying the multiscale diffusion equation makes no assumptions about what is considered noise or what form that noise should take; it assumes only that there is an improvement in the signal to noise ratio with applications of Gaussian kernels. This provides a great deal of flexibility in the kinds of image structure that can be treated as noise. In particular it is possible to choose very large Gaussian kernels in order to prevent correlated structures (structures that are large relative to the pixel size) from contributing to the gradient measurements. The basic requirement is that regions of interest must have measurable intensity differences at some scale that is large enough to suppress the noise. That is, the noise must be small relative to the objects in the image. The effects of this relationship are quantified in the empirical evaluation of Chapt. 5.

Consider the example in Fig. 3.5. It consists of 33 white dots on a black background. The dots are readily distinguishable from the background by any of a number of very simple boundary criteria. The picture also has another interpretation: It consists of two light forms, one small and the other large, on a dark background. In the context of this interpretation, the white dots can be considered a kind of highly structured noise which serves to obscure the two light forms. It is even possible to ascribe boundaries to these larger forms, but these boundaries would inevitably pass through black areas of the image that have no luminance variation; these are *subjective* boundaries.

Consider a choice of s_0 such that the initial gradient measurements employ kernels that are larger

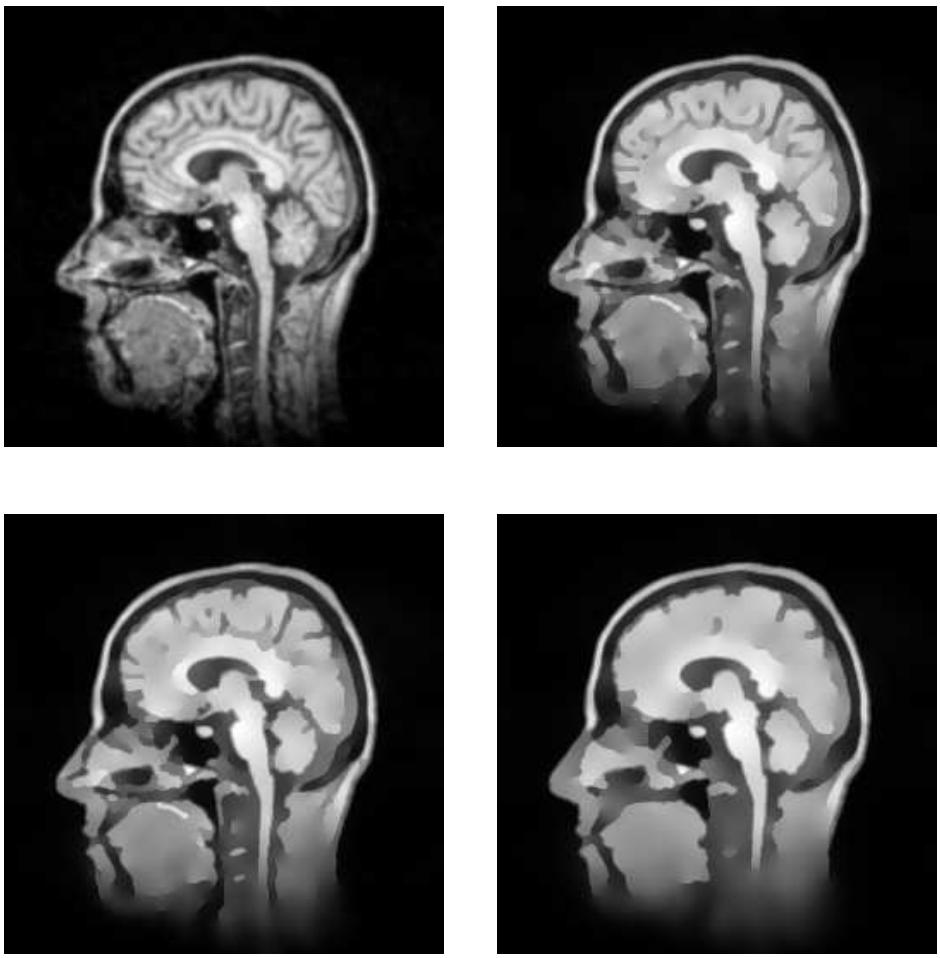

Figure 3.3: The multiscale diffusion equation acting on the MRI head image. From the top left: $\alpha = 3.0$ and the starting scales are (a) 1.0, (b) 2.0, (c) 4.0, and (d) 8.0.

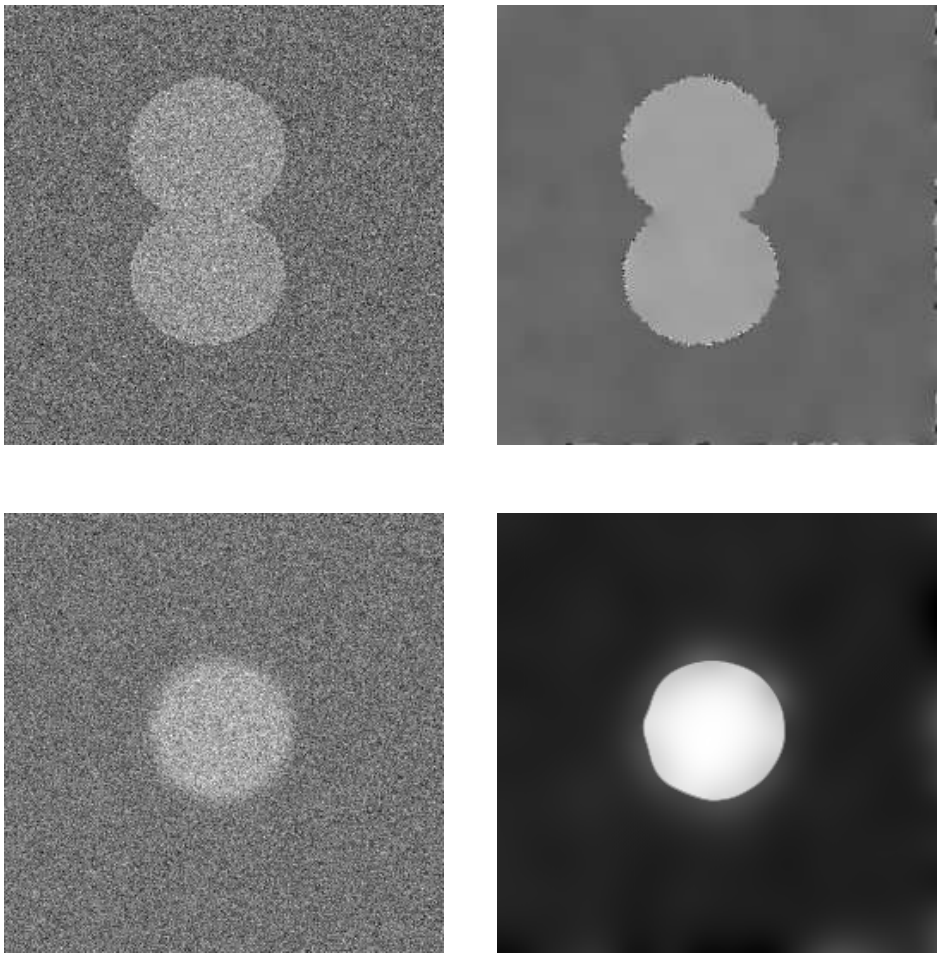

Figure 3.4: The multiscale diffusion equation acting on the test images from Figs. 2.4 and 3.1. From the top left: (a) original, (b) processed with starting scale 3.0, (c) original, (d) process with starting scale 2.0.

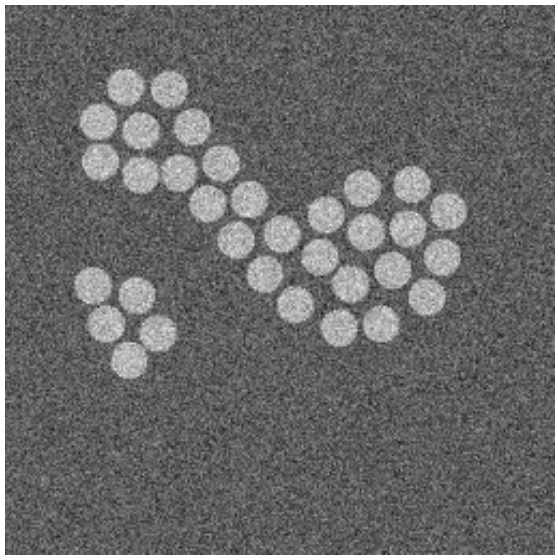

Figure 3.5: White dots on a black background form two larger groups or "objects."

than the white dots. Such a choice would enable the anisotropic diffusion to "fill in" luminance across the larger forms and complete the subjective boundaries. Figure 3.6 shows several time slices of the dotted image at different times in the process. These images show that the smoothing begins near the center of the object and flows outward toward the boundaries. The process groups dots based on their averaged intensities as measured as some large scale. The result in Fig. 3.6d is a pair of objects that have virtually flat luminance functions with well defined boundaries. The presence of the large-scale noise (dots) has influenced the shape of the boundary. The framework for diffusion described in this paper offers some flexibility in the amount of information retained at the boundaries. Allowing $s(t)$ to decrease more slowly, choosing larger values of k, or beginning the process with a Gaussian blurred version of the original image, can all result in smoother boundaries. Of course, these approaches do not discriminate between the effects of noise along the boundary and desirable features of objects.

3.3 Analysis

3.3.1 Space-stretching kernel

One way to describe Eq. (2.9) is the application of isotropic diffusion to the image as it undergoes local warping of space and time (Whitaker and Pizer 1993). In one dimension, let $\bar{x}$ and $\bar{t}$ be local coordinate systems about the point x', t' that vary as functions of the local gradient and let

$$\bar{x} = \left(\frac{1}{g^{\frac{1}{2}}(x,t)} \right) (x - x') \text{ and } \bar{t} = (t - t').$$

The derivatives in these local coordinates are

$$\begin{aligned} \frac{\partial}{\partial \bar{x}} &= \frac{\partial x}{\partial \bar{x}} \frac{\partial}{\partial x} = \left[\frac{1}{g^{\frac{1}{2}}(x,t)} - \frac{g_x(x,t)}{2g^{\frac{3}{2}}(x,t)} (x - x') \right]^{-1} \frac{\partial}{\partial x} \text{ and} \\ \frac{\partial}{\partial \bar{t}} &= \frac{\partial \bar{x}}{\partial t} \frac{\partial}{\partial \bar{x}} + \frac{\partial t}{\partial \bar{t}} \frac{\partial}{\partial t} = -\frac{g_t(x,t)}{2g^{\frac{3}{2}}(x,t)} (x - x') \frac{\partial}{\partial \bar{x}} + \frac{\partial}{\partial t} \end{aligned}$$

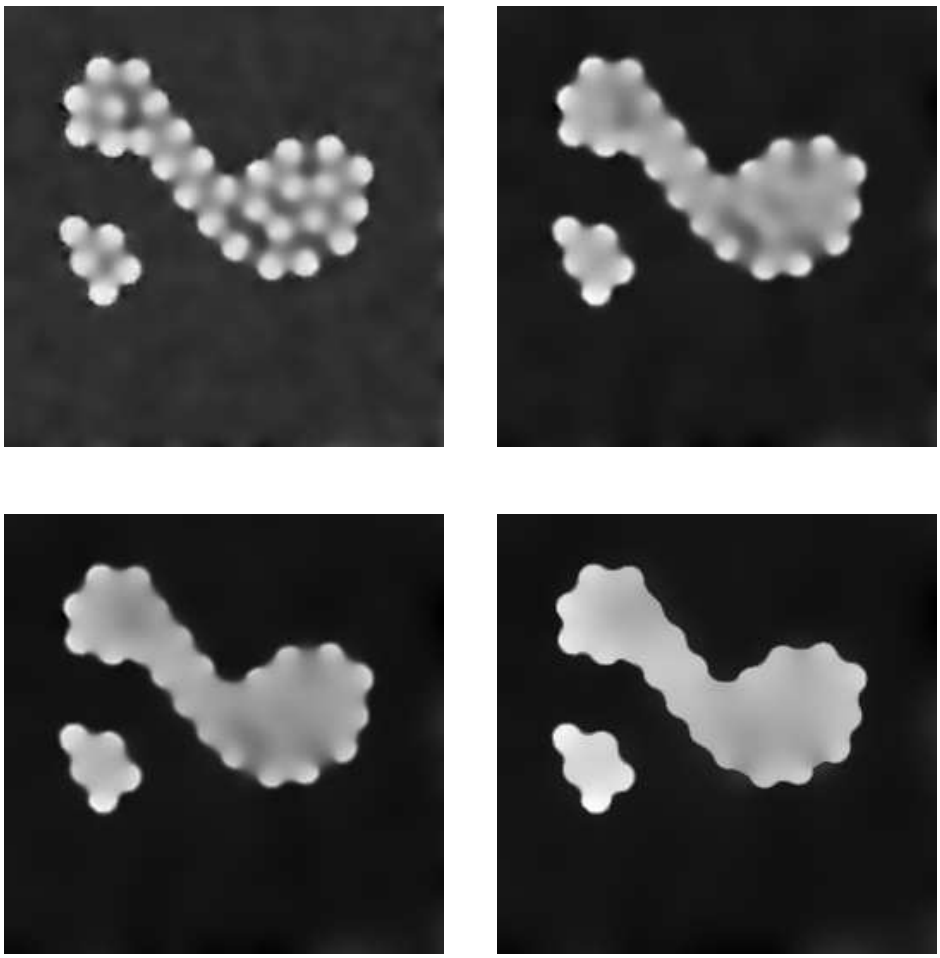

Figure 3.6: A multiscale diffusion process with starting scale 8.0 allows the blurring to proceed from the "inside out" forming two distinct objects. Images from top left show the results after processing time (a) 12.5, (b) 25 (c) 50 (d) 100, where the times are given in units of pixels squared.

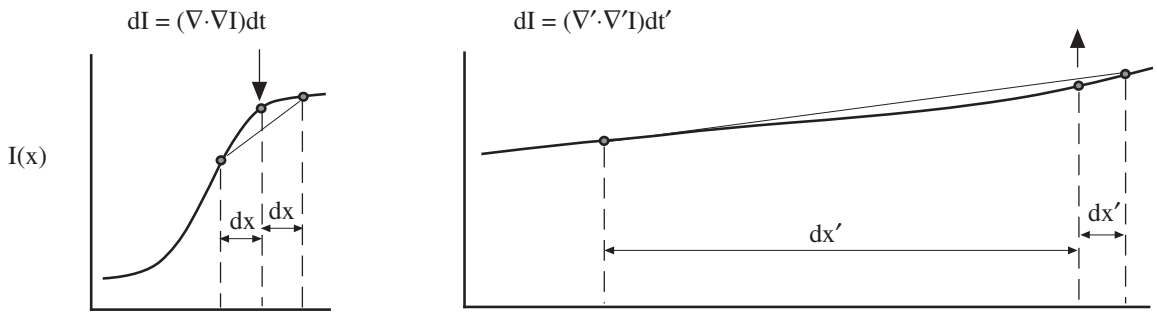

Figure 3.7: The local "stretching" of space can cause a point to move away from the average of its neighbors—contrary to Gaussian blurring.

The coordinate transformation is local, so when these derivatives are evaluated at $x = x'$, $t = t'$,

$$\left.\frac{\partial}{\partial \bar{x}}\right|_{x=x',\, t=t'} = g^{\frac{1}{2}}(x',t')\frac{\partial}{\partial x} \quad \text{and} \quad \left.\frac{\partial}{\partial \bar{t}}\right|_{x=x',\, t=t'} = \frac{\partial}{\partial t}$$

The second derivative with respect to the new spatial coordinates is

$$\begin{aligned}
\frac{\partial^2}{\partial \bar{x}^2} &= \frac{\partial x}{\partial \bar{x}}\frac{\partial}{\partial x}\left(\frac{\partial x}{\partial \bar{x}}\frac{\partial}{\partial x}\right) \\
&= \frac{\partial x}{\partial \bar{x}}\frac{\partial}{\partial x}\left(\left(\frac{1}{g^{\frac{1}{2}}(x,t)} - \frac{g_x(x,t)}{2g^{\frac{3}{2}}(x,t)}(x-x')\right)^{-1}\frac{\partial}{\partial x}\right) \\
&= \left(\frac{1}{g^{\frac{1}{2}}(x,t)} - \frac{g_x(x,t)}{2g^{\frac{3}{2}}(x,t)}(x-x')\right)^{-2}\frac{\partial^2}{\partial x^2} \\
&\quad + \left(\frac{1}{g^{\frac{1}{2}}(x,t)} - \frac{g_x(x,t)}{2g^{\frac{3}{2}}(x,t)}(x-x')\right)^{-3} \\
&\quad \left(\frac{g_x(x,t)}{g^{\frac{3}{2}}(x,t)} + \left(\frac{g_{xx}(x,t)}{2g^{\frac{3}{2}}(x,t)} - \frac{3g_x^2(x,t)}{4g^{\frac{5}{2}}(x,t)}\right)x\right)\frac{\partial}{\partial x}.
\end{aligned}$$

Evaluate the second derivative at $x = x'$, $t = t'$ and get

$$\left.\frac{\partial^2}{\partial \bar{x}^2}\right|_{x=x',\, t=t'} = g(x',t')\frac{\partial^2}{\partial x^2} + g_x(x',t')\frac{\partial}{\partial x} = \frac{\partial}{\partial x}\left.\left(g(x,t)\frac{\partial}{\partial x}\right)\right|_{x=x',\, t=t'}.$$

The evaluation of all these terms at $x = x'$, $t = t'$ is a notational convenience; the same result holds at each point at each point in the image space. Satisfying the diffusion equation in the new coordinates $\partial/\partial \bar{t} f = \partial^2/\partial \bar{x}^2 f$ gives the variable-conductance diffusion equation

$$\frac{\partial}{\partial t} f = \frac{\partial}{\partial x} g(x,t)\frac{\partial}{\partial x} f.$$

This local "stretching" of space and time near edges (Fig. 3.7) can reverse the sign of the Laplacian operator and allow the "sharpening" or "enhancement" of edges near places of high gradient.

3.3.2 Conductance and edge enhancement

When processing images with the anisotropic diffusion equation, properties of edge enhancement and blurring are sensitive to the choice of parameters, k and t. Not all edges in an image are enhanced—smaller edges are blurred away. Choices of k provide a lower limit on the gradients of edges that are enhanced.

In order to simplify the analysis, consider the one-dimensional version of (2.9). For this analysis I make the simplifying assumption that the gradients in the conductance term are computed at the same scale as the other derivatives in Eq. (3.2). Define the flow, $\phi(f_x) = g(|f_x|)f_x$ as in Perona and Malik (1990). The anisotropic diffusion equation is

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{\partial}{\partial x} g(|f_x|)f_x = g'(|f_x|)f_x f_{xx} + g(|f_x|)f_{xx} \\ &= \phi'(f_x)f_{xx}. \end{aligned} \tag{3.5}$$

For $\phi'(f_x) > 0$ this equation looks like Gaussian blurring, where the time parameter is modulated by the local gradient information. However, in regions where $\phi'(f_x) < 0$ this equation has just the opposite effect, and there is a kind of "reverse blurring," i.e., a sharpening. From Eq. (3.5) the change in image gradient over time is

$$\frac{\partial f_x}{\partial t} = \phi'' {f_{xx}}^2 + \phi' f_{xxx}. \tag{3.6}$$

The sign of the right-hand side of (3.6) indicates whether gradients are increasing or decreasing. For a local maximum of gradient $f_{xx} = 0$ and $f_{xxx} < 0$. Therefore, edges that are local maxima of gradient become progressively more steep over time ($\frac{\partial f_x}{\partial t} > 0$) provided that $\phi'(f_x) < 0$. Perona and Malik (1990) suggest choices of $g(\cdot)$ so that $\phi(\cdot)$ has the property that there is exactly one β where

$$\frac{\partial f_x}{\partial t} = \begin{cases} > 0 & \text{if } f_x > \beta \\ < 0 & \text{if } f_x < \beta \end{cases} . \tag{3.7}$$

The conductance functions of Eqs. (2.10) and (2.11) have this property, and $\beta = k$ in both cases. This suggests that one could choose k to reflect the lower bound on the steepness of edges that are enhanced; edges with less slope than β are blurred away while edges with slopes above β are enhanced.

This analysis of the role of k suggests that it should be set to reflect the steepness of edges in the image. Because it serves as a means of separating the signal and the noise, it will likely depend on both the contrast of the signal and the level of the noise. In this analysis k has been tied to the overall contrast in the image by using convention (Sect. 2.7) of specifying k as some multiple of the average gradient in the image. Chapter 5 will quantify the effects of varying k with different levels of noise.

3.3.3 Instability

Unfortunately, there are further constraints on the choice of k. Figure 3.8 shows anisotropic diffusion applied to the blurred disk from Fig. 2.7. For this analysis k was purposely chosen to be low; it is the root-mean-squared gradient magnitude of the image. The result is a "staircasing" effect. Instead of isolating a single steep portion of the disk boundary as in the examples above, the process has broken the boundary into many discrete steps.

I will study this effect by examining the local behavior (time and space) of one-dimensional functions. The sharpness of an edge (with $f_x > 0$) is indicated by the third derivative f_{xxx}, which is negative. A more negative the third derivative indicates a sharper edge. The following analysis

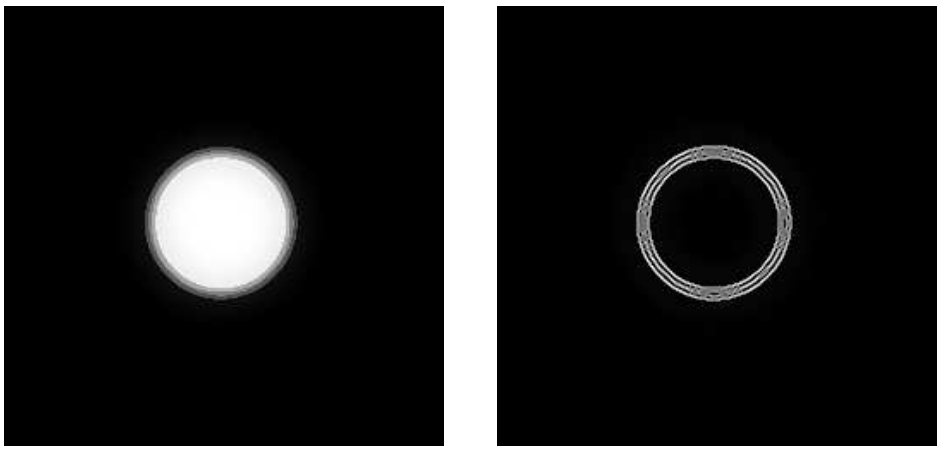

Figure 3.8: Results from the blurred disk show (a) the image after processing and (b) the gradient magnitude of the image. This image, which is processed $t = 20$ with $k = 1.0$, shows that the anisotropic diffusion can produce an unstable "staircasing" effect.

will show that for sufficiently steep one-dimensional edges, the third derivative f_{xxx} can increase (become less negative) through time ($f_{xxxt} > 0$). In this manner steep edges can form "ramp-like" functions (locally) which are unstable and can "break" or "fracture" into even more step edges. As those edges become progressively more steep, they break to form more edges, and so on. Thus, as edges are enhanced beyond a certain steepness they also become more flat, and flat, steep edges are unstable. The resolution to this problem will be to limit the magnitude of conductance measurements by using finite scale, an essential aspect of mutiscale diffusion, and by employing higher-order geometric measures as described in Chapt. 4.

The strategy for understanding this instability is to first examine the particular case of an analytically modeled edge function (a Gaussian-blurred step edge), examine the Taylor series expansion of the change in first derivative over time, f_{xt}, and then characterize the steepening in terms of the flow ϕ. The step-edge analysis will examine the behavior of the first derivative in the neighborhood of a locally steep point. The behavior of that first derivative near maximally steep points will indicate that for very steep edges *relative steepness* (relative to a local neighborhood) at the point is *decreasing* over time, i.e., edges can become more flat. I will use these results to construct a more general formulation of this analysis which applies to other types of edges.

Consider a one-dimensional image $f(x)$ that has a single "edge."

$$f(x) = \frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}} \int_{-\infty}^{x} \exp\left(\frac{-l^2}{2\sigma^2}\right) \mathrm{d}l \tag{3.8}$$

The free parameter σ determines the "smoothness" of this luminance function.

This function is a step function of height A that is blurred with a Gaussian kernel of size σ. The advantage of using this function is that it has closed form analytical expressions for all of its derivatives. The derivative of $f(x)$,

$$f_x = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp\left(\frac{-x^2}{2\sigma^2}\right),$$

is the familiar bell shaped Gaussian; $f(x)$ has a maximum gradient at $x = x_0 = 0$. The gradient at

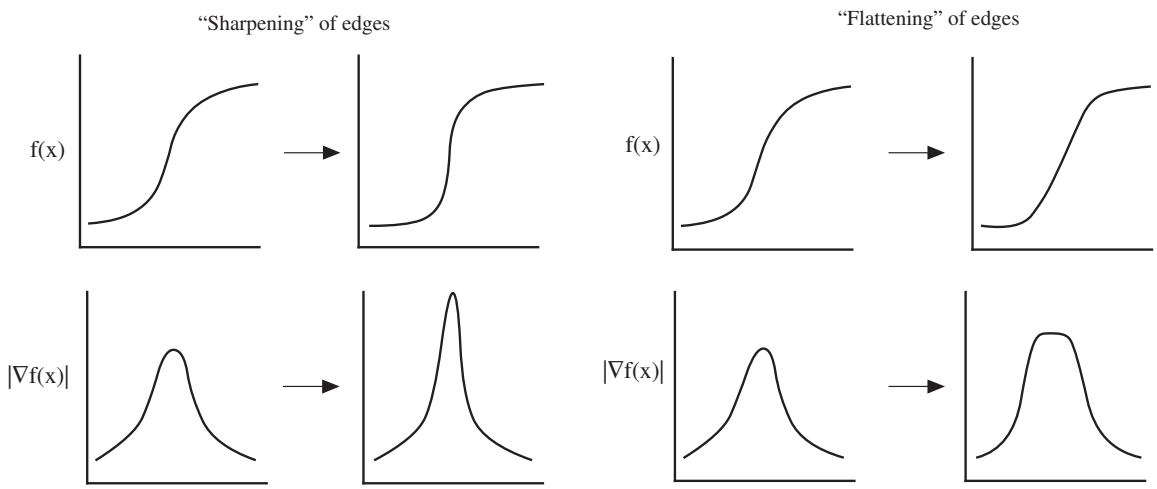

Figure 3.9: Sharpening is distinguished from flattening by examining the manner in which the first derivative changes in a local neighborhood.

the steepest point, x_0, is increasing for appropriate choices of the conductance parameter (Eq. (3.7)). However, this says nothing about the way in which f_x changes in the neighborhood of x_0. In order for this edge to become "sharper," the gradient at x_0 should increase more quickly than the gradients in a local neighborhood of x_0. If the local neighborhood of x_0 increases in steepness more quickly than $f_x(x_0)$, then the edge f becomes more steep at x_0 but also more "flat" or "straight": sharp edges can become more like ramp functions. This distinction is depicted in Fig. 3.9.

The term $\frac{\partial f_x}{\partial t}$ indicates whether or not any point in the image is becoming more steep over time. Sharpening, however, is slightly more subtle. For the sharpening of an edge, places with higher f_x should have higher $\frac{\partial f_x}{\partial t}$. Otherwise, less steep places could "catch up" to places that are more steep and undermine the possibility that there is one point that is locally most steep. Characterizing this behavior requires analyzing the manner in which higher order terms change over time. From (3.6) and (3.7) and expanding the exponential in a Taylor series, we have

$$\begin{aligned}\frac{\partial f_x}{\partial t} &= \left[\phi'' {f_{xx}}^2 \frac{x^2}{\sigma^4}\frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left(\frac{-l^2}{2\sigma^2}\right) + \phi'(\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})\right]\frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left(\frac{-x^2}{2\sigma^2}\right) \\ &= \frac{1}{\sigma^2}\frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}}\left[-\phi' + (\phi''\frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}} + \frac{3}{2}\phi')\frac{x^2}{\sigma^2} + \mathcal{O}(x^4)\right] \qquad (3.9)\end{aligned}$$

To second order, this function is a quadratic in x—it is a parabola. The issue is whether that parabola faces up or down. If it faces down, then it describes the sharpening behavior. The local behavior of $\frac{\partial f_x}{\partial t}$ is dependent on the sign of the expression $(\phi''\frac{A}{(2\pi\sigma^2)^{\frac{1}{2}}} + \frac{3}{2}\phi')$. For $\phi'' < -\frac{3(2\pi\sigma^2)^{\frac{1}{2}}}{2A\phi'}$ the edge becomes sharper in the neighborhood of x_0, while $\phi'' > -\frac{3(2\pi\sigma^2)^{\frac{1}{2}}}{2A\phi'}$ creates the "flattening" effect shown in Fig.3.9.

Whether or not an edge sharpens near a maximally steep place depends on the relationship between $\phi'(f_x)$ and $\phi''(f_x)$. Recall that $\phi(f_x) = g(f_x)f_x$ is called the flow. It has several properties. It is positive (for positive gradients) and bounded but is not monotonic. For small gradients $\phi' > 0$ while for large gradients $\phi' < 0$, resulting in the steepening at edges. The flattening depends on the second derivative of flow, and this flattening happens at the steepest point of the blurred step edge when

$$\phi'' > -\frac{3(2\pi\sigma^2)^{\frac{1}{2}}}{2A\phi'(f_x(x_0))} = -\frac{3}{2f_x(x_0)\phi'(f_x(x_0))}. \qquad (3.10)$$

For any conductance function with a conductance parameter k, Eq. 3.10 establishes a limit $f_x(x_0) = \lambda$. Edges that are more steep than this (and adhere to the blurred step-edge model) will result in the instabilities shown in Fig. 3.8.

I generalize this analysis to more general types of edges by observing that the possibility for this flattening occurs when $\phi'' > 0$. Define a new function e,

$$e(f_x, f_{xx}, f_{xxx}) = \frac{\partial f_x}{\partial t} = \phi'' {f_{xx}}^2 + \phi' f_{xxx}, \tag{3.11}$$

which is Eq. (3.6) with a change in formal parameters. This function describes the rate of change of the first derivative at any location in the image, given the local properties up to third order. In order for the process to accentuate the most steep portions of a smooth edge and reduce the steepness of other portions of the same edge, the function $e(f_x, f_{xx}, f_{xxx})$ should penalize locations that are not local maxima of gradient. Consider the behavior of (3.11) in a neighborhood of x_0 where $f_x(x_0)$ is a local maximum. If f_x and f_{xxx} are fixed in this neighborhood and this function is a quadratic in $f_{xx}(x)$,

$$e(f_{xx}) = \phi'' f_{xx}(x)^2 + \phi' f_{xxx}.$$

This quadratic has a local maximum at $f_{xx} = 0$ only if $\phi'' < 0$. Assume that first and third order properties are relatively fixed in this neighborhood, i.e., $(f_x)_x = 0$ and $(f_{xxx})_x \approx 0$. Then local maxima of intensity gradient ($f_{xx} = 0$ and $f_{xxx} < 0$) increase maximally only if $\phi'' < 0$. For $\phi'' > 0$ places that are nearby local maxima of f_x can actually increase their gradients more quickly than those places that are the most steep.

Every reasonable choice of $g(f_x)$ creates the possibility for a positive ϕ'' and therefore the possibility of staircasing. The flow, ϕ, is strictly positive and has a negative slope for all $f_x > k$. Any acceptable choice of $g(|f_x|)$ will result in a positive ϕ'' for f_x sufficiently large. Otherwise, the function $g(f_x)$ would eventually cross zero at some point.

A given conductance function and an associated parameter k establish a bound on gradient value above which instabilities should be expected. For the conductance term $g(|\nabla f|) = \exp\left(\frac{-|\nabla f|^2}{2k^2}\right)$ that was used in these examples, $\phi'' > 0$ only for $f_x < \sqrt{3}k$. When smooth edges in an image have gradients larger than $\sqrt{3}k$, the staircasing described above can result.

The blurred step edge above demonstrates that multiscale diffusion can be used to control staircasing effects. The gradient of the step edge is proportional to $\frac{A}{\sigma}$ when measured at a scale σ. If A is the maximum contrast expected in the initial image, then by the maximum principle it cannot exceed that value as the process evolves. Thus, the measurement scale s and the conductance parameter k can be tuned to prevent such staircasing.

This analysis suggests that for a given conductance function there is a limited range of gradient values that will produce edge enhancement without the staircasing of smooth edges as shown in Fig. 3.10. Choosing k can be very difficult for images that have a great variation in gradient values. For such images it might be impossible to characterize both dim and bright objects. Experiments have shown that this staircasing tends to occur only on wide flat edges as in the blurred disk. For a wide range of images that have only narrow or sharp boundary regions, this phenomenon was not evident.

3.3.4 Two-dimensional analysis

The one-dimensional analysis offers some interesting insights and predicts many of the experimental results on two-dimensional images. A similar two-dimensional analysis, although somewhat more complex, provides more information. This two-dimensional analysis provides information about blurring along the level curves (isophotes) of an image and shows that this level set blurring does

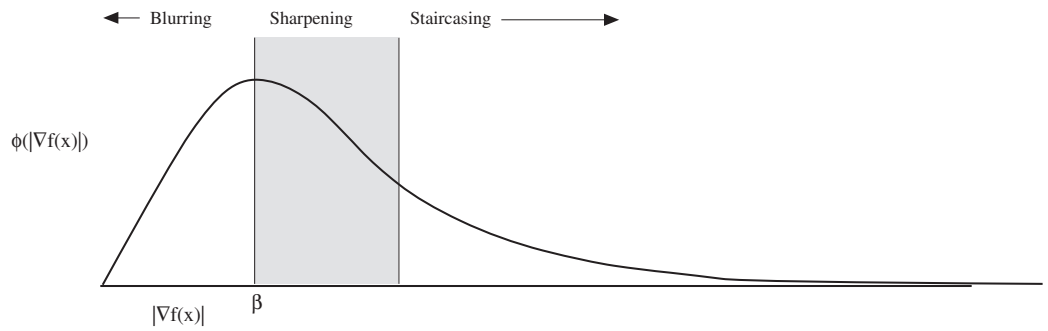

Figure 3.10: Because of the staircasing phenomenon there is a limited range of gradient values that produce desirable results. The darkened area on the graph indicates the range of edge values that will become steeper without becoming less sharp. The left-hand side of that area is the value β, which is the threshold between those edges that are enhanced and those that are blurred.

not undergo the reverse blurring that is associated with the enhancement of edges. As a result, anisotropic diffusion tends to blur away corners of objects, even those corners that are associated with very steep edges.

For this analysis it is useful to express the derivatives in local gauge coordinates. For this I use the convention of ter Haar Romeny, Florack, Koenderink and Viergever (1991a) of denoting the gradient direction as w and the level set (or isophote) direction as v. The Laplacian, expressed in such local coordinates, is $\nabla \cdot \nabla f = f_{ww} + f_{vv}$, the gradient magnitude is f_w, and $f_v = 0$ by definition. The edge-affected diffusion equation is

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial w} g(f_w) f_w + \frac{\partial}{\partial v} g(f_w) f_v \tag{3.12}$$

$$= \phi'(f_w) f_{ww} + g(f_w) f_{vv}, \tag{3.13}$$

where $\phi = g(f_w) f_w$ as in (3.6). This analysis suggests that there are two actions associated with this process. One action, in the gradient direction, is the same as the one-dimensional case where gradients of sufficient magnitude undergo the "reverse" blurring ($\phi' < 0$) referred to in Sect. 3.3.2. The other action, in the isophote direction, undergoes a blurring that is *modulated* but always positive because $g(f_w) > 0$.

Consider the way in which the gradient changes at Canny edges ($f_{ww} = 0$ and $f_{www} < 0$).

$$\frac{\partial f_w}{\partial t} = \phi''(f_w) {f_{ww}}^2 + \phi'(f_w) f_{www} + g'(f_w) f_{ww} f_{vv} + g(f_w) f_{vvw} \tag{3.14}$$

$$= \phi'(f_w) f_{www} + g(f_w) f_{vvw} \tag{3.15}$$

The first term is analogous to the one-dimensional case—gradients tend to become more steep if they are above the threshold $f_w > k$. The second term indicates that the effect along the level set is different. Note that $f_{vvw} = (f_w)_{vv}$, so the gradient tends to become more like gradients in the neighborhood that lies tangent to the level curve. This effect is the same in Gaussian blurring—there is averaging within the tangent space to the level curve. In the case of anisotropic diffusion the level-set blurring is reduced at places of high gradient, but it does not change sign. For relatively flat objects with sharp corners (high level set curvature), $(f_w)_{vv}$ is likely to be negative and would reduce or offset the enhancing effects of the first term. In the level-curve direction there is always some kind of averaging or blurring, *so the process has a tendency to round off corners of flat objects by blurring them away.* If the edges of the initial image are sufficiently steep, the effects of this "rounding" are slow, but the action of blurring along level curves does not change sign.

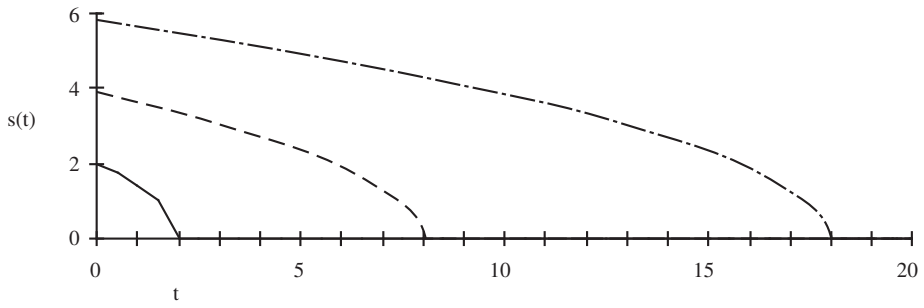

Figure 3.11: $s(t)$ for values of $s_0 = 2.0, 4.0, 8.0$ and $\alpha = 1.0$.

This analysis suggests that the process could include higher order terms that could control the blurring at corners and preserve shapes with sharp corners. There are a number of options for including higher order information. Using the regularization approach described in Sect. 2.7.1, one might construct energy functions that favor the higher curvature in the isophote direction that is associated with corners. The use of second and third order information is discussed in greater detail in Chapt. 4.

3.4 Scale tradeoff

The process of Sect. 3.2 describes a continuous "tradeoff" between the isotropic and anisotropic scale, where the scale of isotropic measurements used in the conductance term decreases as the image undergoes progressively more isotropic blurring. The scale recipe, $s(t)$, indicates how the scale of gradient measurements changes as the process progresses.

One possibility for choosing $s(t)$ is to make an exact tradeoff between isotropic and anisotropic scale and allow one unit of isotropic scale to equal a comparable number of units of anisotropic scale. Given a starting scale $s_0 = s(0)$, where s is the standard deviation of a Gaussian kernel used to make derivative measurements, set the sum of anisotropic and isotropic scale to a constant. Then

$$s(t)^2 + 2\alpha t = s_0^2,$$

so that

$$s(t) = \sqrt{s_0^2 - 2\alpha t}, \tag{3.16}$$

where α is a dimensionless constant which describes the relationship between a unit of isotropic and a unit of anisotropic scale. For a given starting scale this relationship prescribes an amount of anisotropic blurring that could effectively replace the effects of the initial kernel $G(s_0)$. For a continuous range of starting scales this process describes an alternate *anisotropic* scale space.

The *scale* of an anisotropic process is therefore the starting scale s_0, and this scale indicates the appropriate amount of time t that the process runs in order to achieve an amount s_0 of anisotropic blurring. Figure 3.11 shows examples of $s(t)$ for several values of s_0. Notice two things about these graphs. First, the process spends relatively more time at larger values of s. Second, these graphs indicate that there is a specific, finite amount of processing for any given starting scale s_0, and the amount of processing increases with increasing starting scale.

Equation (3.16) derives from a formalism. It follows from equating the scale and evolution parameters in the two processes on the basis of the dimensions of these parameters. This relationship between variable-conductance diffusion and Gaussian scale relies on the simplifying assumptions that the two types of scale are commensurate and there is a single constant of proportionality α between those two processes. Of course, this is not the case. The variable-conductance diffusion

process is interesting *because* it is nonstationary, so the relationship in (3.16) embodies a substantial simplification.

This simplification has tremendous practical implications, because at a single point in time t, it allows one to measure the gradient magnitude at a single scale throughout the entire image. Alternatively, one could compute the entire Gaussian scale-space for the entire image at each point in time. The use of this scale-space (the weighting of the different scales) might depend on local image structure. Besides the enormous computational burden, this strategy introduces (or revives) the very difficult problem of how to combine information from different scales. Instead, I choose a single relationship between s and t, and accept the consequences of this simplification.

The following discussion develops a relationship between s and t on the basis of the principle: *Noise that has been reduced by the application of the Gaussian kernel (size s) should not resurface as t increases and s decreases.* The challenge is determine the metric for deciding the level of noise. I develop this relationship so that $s(t)$ *guarantees* the reduction of specific levels of noise *in the absence of interesting image structure*. I proceed as follows. First, establish the incremental changes in dissimilarity that result from both the reduction of $s(t)$ and the increase in t. Second, set these incremental changes to offset each other. Third, choose a level of conductance (squared gradient magnitude) that represents the average behavior of this process in a field of uncorrelated white noise that is removed from any "interesting" image structure. Finally, calculate the relationship that guarantees that a *specified percentage* of the structure in this field undergoes an amount of blurring of *at least* s_0, the starting scale. That is, establish a relationship $s(t)$ so that "very little" of the noise in a field of white noise could resurface as $s(t)$ is reduced.

The strategy is to reduce $s(t)$ at such a rate that specified levels of noise remain unaffected by the tradeoff between isotropic and anisotropic scale. This would prevent such noise from "resurfacing" as $s(t)$ decreases. Such levels of noise would undergo a total amount of blurring of at least s_0 when the process is complete. The diffusion equation that describes the application of an isotropic kernel of size s is

$$\frac{\partial f}{\partial u} = f_{xx}, \tag{3.17}$$

where $2u = s^2$. The diffusion equation that describes the application of a "unit" of anisotropic scale is

$$\frac{\partial f}{\partial t} = \phi'(f_x) f_{xx}. \tag{3.18}$$

At a single point in time the two processes described by (3.17) and (3.18) both affect successive gradient measurements. Set the net result of these two blurring processes to zero (∂f's sum to 0). This gives the relationship $s(t)$ that must be met so that a point is unaffected by the tradeoff between s and t:

$$f_{xx}\partial u + \phi'(f_x) f_{xx}\partial t = 0.$$

The relationship between u and t is

$$\frac{\partial u}{\partial t} = \phi'(f_x), \text{ and } \frac{\partial u}{\partial t} = \frac{\partial s^2}{2\partial t} = s\frac{\partial s}{\partial t},$$

so that

$$\frac{\partial s}{\partial t} = -\frac{1}{s}\phi'(f_x). \tag{3.19}$$

Solutions depend on $f_x(x,t)$. For this analysis a single f_x is chosen to represent the entire image and is fixed at some level that insures that unwanted structures (noise) undergo the desired amount of blurring.

Blom (1992) shows that in the presence of Gaussian white noise, first derivative measurements

made at scale s' have a variance of

$$\sigma_{s'}^2 = \frac{\sigma_N^2}{8\pi s'^4} \tag{3.20}$$

Consider a field of Gaussian white noise g_N that is isolated from any coherent image structure. Assume that this noise has undergone a level of blurring s'. The expected value of first derivatives $\langle g_{Nx} \rangle$ in that field is zero, and the expected value of the square of first derivatives $\langle (g_{Nx})^2 \rangle$ is the variance $\sigma_{s'}^2$ in the expression above, so that

$$\langle (g_{Nx})^2 \rangle = \frac{\sigma_N^2}{8\pi s'^4} \tag{3.21}$$

Now choose a scale recipe that guarantees that a certain percentage of the random field undergoes the appropriate amount of blurring. Define this percentage to be those points that have values $(g_{Nx})^2$ less than $\gamma \langle (g_{Nx})^2 \rangle$. For this example choose the conductance function in (2.10) so that $\phi'(f_x) = (1 - {f_x}^2/k^2) \exp\left(-\frac{{f_x}^2}{2k^2}\right)$, and introduce the squared gradient value that serves as a cutoff. The result is

$$\frac{\partial s}{\partial t} = -\frac{1}{s}\left(1 - \frac{\gamma\sigma_N^2}{8\pi k^2 s'^4}\right) \exp\left(-\frac{\gamma\sigma_N^2}{16\pi k^2 s'^4}\right). \tag{3.22}$$

All that remains is to calculate s', the amount of blurring that this Gaussian noise has undergone before the gradient measurement. This blurring has three components. The first component is the inner scale s_I of the noise as it exists in the initial image. This accounts for the fact that the noise might be somewhat correlated from the start. The second component is the scale introduced by the anisotropic blurring and the third component is the blurring introduced in calculating conductance. The second and third components have been constructed so that their sum is *at least* the starting scale s_0. The full equation is

$$\frac{\partial s}{\partial t} = -\frac{1}{s}(1 - \kappa) \exp\left(-\frac{\kappa}{2}\right). \tag{3.23}$$

where

$$\kappa = \frac{\gamma\sigma_N^2}{8\pi k^2 ({s_0}^2 + {s_I}^2)^2} \tag{3.24}$$

Solutions have the same form as (3.16) with $\alpha = (1 - \kappa) \exp\left(-\frac{\kappa}{2}\right)$.

As a review consider the following. The purpose of the process is to blur away the noise. The measurement scale $s(t)$ is reduced so that unwanted structures do not "resurface" as the process progresses. Unwanted structures are modeled as a field of blurred Gaussian white noise, for which the statistics of Gaussian derivative measurements are known. A level γ indicates, in terms of the expected value of the field, the very largest gradients that will be blurred the full amount s_0. A larger γ makes it less likely that noise will "slip through" the blurring process.

These equations establish a relationship between the starting scale s_0, the magnitude of the noise σ_N, the degree of correlation s_I in the noise, the conductance parameter k, and γ, which indicates a degree of confidence that noise is being reduced. A requirement for the process to eventually finish is that $\frac{\partial s}{\partial t} < 0$. This requires $k > \left(\frac{\gamma\sigma_N^2}{8\pi}\right)^{\frac{1}{2}} \frac{1}{{s_0}^2 + {s_I}^2}$. More specifically, (3.23) specifies a relation between the starting scale and the conductance parameter as shown in Fig. 3.12. In practice, after characterizing the noise, one would choose a starting scale that provides the best distinction between the signal and the noise and then choose values of k and γ that offer an acceptable compromise among the following:

- k low enough to preserve image structure,
- γ large enough to ensure that a large percentage of the noise is blurred away, and
- κ small enough so that the process finishes in a reasonable amount of time.

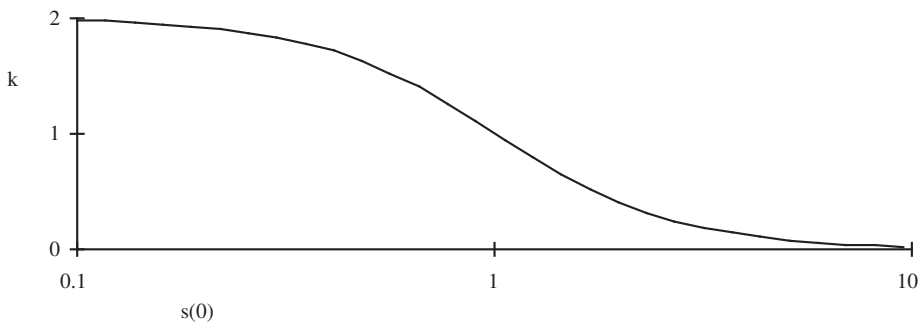

Figure 3.12: The reduction of random noise in finite amount of time ($\alpha > 0$) establishes a relationship between the conductance parameter and scale. This relationship is shown for $\sigma_N = 4.0$, $\gamma = 4.0$, $s_I = 2.0$.

This model of noise is more general than it first appears. It applies not only to white noise, *but any noise that, at some level of scale, resembles Gaussian blurred white noise.* It can be shown (ter Haar Romeny, Florack, Koenderink and Viergever 1991b) experimentally that even in images that have coherent structure, the annihilation of extrema (at some appropriately large scale) is similar to what is expected in white noise. When an image is blurred with kernels that are larger than the correlated structure in that image, the result tends to resemble blurred white noise.

Figure 3.13a,b shows two images. The first is a pattern of dots that incorporates a great deal of structure, and the second is an image of white noise. Figure 3.13c,d shows the same images, blurred so that the similarity in large-scale structure is apparent. The generality of this analysis comes from that fact that many different kinds of noise resemble Gaussian white noise when compared at some large scale.

This chapter began by demonstrating one of the basic problems of anisotropic diffusion (as proposed by Perona and Malik (1990) as well as Nordstrom (1990), Gerig, Kubler, Kikinis and Jolesz (1991), and Shah (1991)); it does not account for the relevant scale of the structures that are being measured. The gradient magnitude in the conductance term $g(|\nabla f|)$ must include a scale which controls the size of the structures that affect the "flow" of image intensity. I have proposed a method, called multiscale diffusion, wherein the measurement scale $s(t)$ decreases over time in order to capture the small-scale structure associated with the boundaries of large-scale objects. The examples in this chapter indicate that the multiscale diffusion process performs quite well in the presence of uncorrelated as well correlated noise. An objective analysis of this behavior is presented in Chapt. 5.

Recall from Sect. 2.3 that there are two problems associated with conventional edges as measured by the gradient magnitude. One is the sensitivity to noise, which is addressed in this chapter. The other is the need for segmentations based on properties other than luminance. The next chapter addresses this second issue by generalizing the variable-conductance diffusion process to apply to multi-valued images. It presents a framework for constructing image segmentations that rely on higher-order geometric structure.

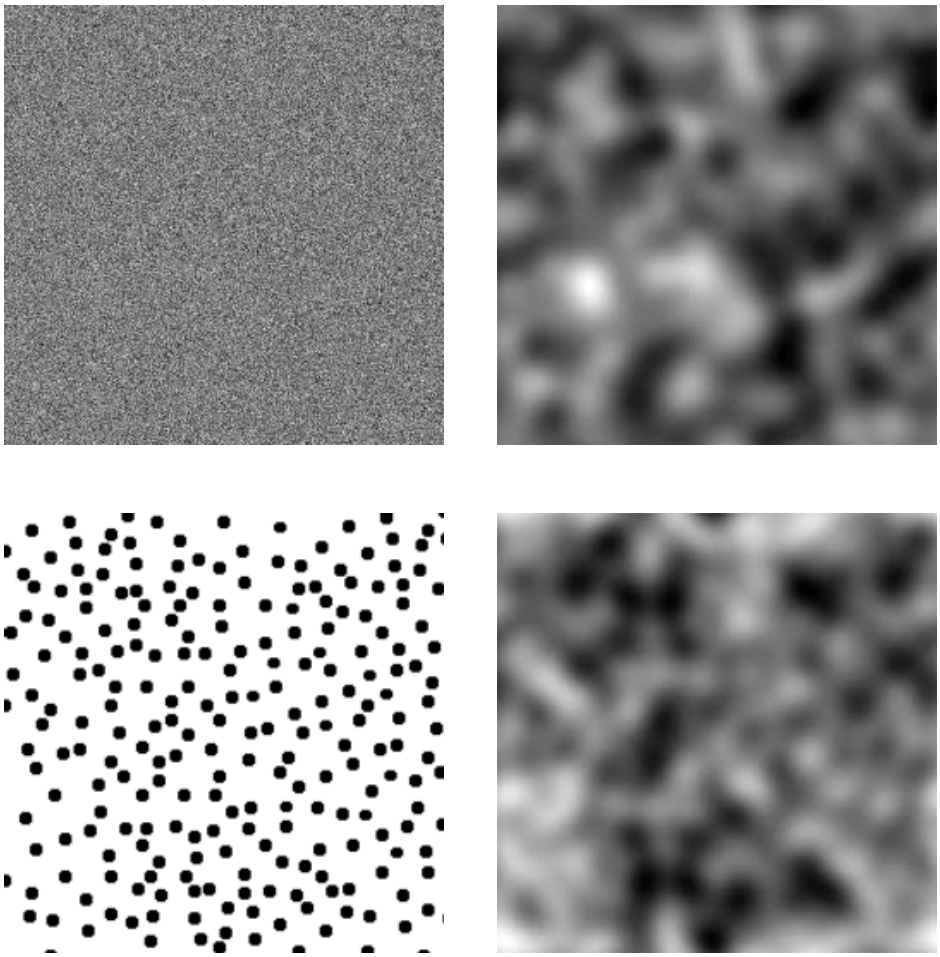

Figure 3.13: A field of white dots is shown to resemble blurred Gaussian white noise. Top: original images. Bottom: images blurred at scale 10.0.