

# Scale-Space Properties of Nonlinear Diffusion Filtering with a Diffusion Tensor

Joachim Weickert  
Laboratory of Technomathematics  
University of Kaiserslautern  
P.O. Box 3049  
67653 Kaiserslautern  
Germany

## Abstract

In spite of its lack of theoretical justification, nonlinear diffusion filtering has become a powerful image enhancement tool in recent years. The goal of the present paper is to provide a mathematical foundation for continuous nonlinear diffusion filtering as a scale-space transformation which is flexible enough to simplify images without losing the capability of enhancing edges. By studying the Lyapunov functionals, it is shown that nonlinear diffusion reduces  $L^p$  norms and central moments and increases the entropy of images. The proposed anisotropic class utilizes a diffusion tensor which may be adapted to the image structure. It permits existence, uniqueness and regularity results, the solution depends continuously on the initial image, and it satisfies an extremum principle. All considerations include linear and certain nonlinear isotropic models and apply to  $m$ -dimensional vector-valued images. The results are juxtaposed to linear and morphological scale-spaces.

**Keywords.** scale-space, image enhancement, nonlinear diffusion, maximum principle, Lyapunov functionals.

**AMS subject classifications.** 68U10, 35B30, 35B35, 35B50, 35K05, 35K55.

**Abbreviated title.** Scale-Space Properties of Diffusion Filtering.

## 1 Introduction

A frequent problem in low-level vision arises from the wish to eliminate noise and uninteresting small scale details from a degraded image, without blurring semantically important structures such as edges. Image enhancing preprocessing steps of this kind are often necessary for subsequent segmentation and feature extraction tasks.

Among the methods to achieve these goals, a large number of techniques based on partial differential equations (PDEs) has been proposed in the recent decade. Elliptic methods are

frequently related to variational problems via the Euler equations (see e.g. [32, 44]), while shock filters are important representatives from the hyperbolic area [34, 3]. The parabolic field is governed by mean curvature methods [2] and diffusion techniques, on which we will focus in the sequel.

The simplest diffusion filtering eliminates noise by a linear and isotropic process [30, 53, 26]. Let the original image be represented by a real-valued function  $f \in L^\infty(\mathbb{R}^m)$ . Then, one takes  $f$  as initial condition for the diffusion equation

$$\partial_t u = \Delta u \tag{1}$$

and interprets the solution  $u(x, t)$  as a filtered version with the time  $t$  as smoothing parameter. By means of the fundamental solution of the diffusion equation, this is equivalent to convolving the initial image with a Gaussian of width  $\sqrt{2t}$ :

$$K_\sigma(x) := \frac{1}{(2\pi\sigma^2)^{m/2}} \cdot \exp\left(-\frac{|x|^2}{2\sigma^2}\right), \tag{2}$$

$$u(x, t) = (K_{\sqrt{2t}} * f)(x) := \int_{\mathbb{R}^m} K_{\sqrt{2t}}(x-y) f(y) dy. \tag{3}$$

An obvious disadvantage of this proceeding is the fact that it does not only smooth noise, but also blurs important features such as edges.

To avoid this, Perona and Malik [35] suggest filtering by nonlinear diffusion. They apply a nonuniform process (which they name anisotropic) that reduces the diffusivity at those locations having a larger likelihood to be edges. This likelihood is measured by  $|\nabla u|$ . Hence, they use

$$\partial_t u = \operatorname{div}(g(|\nabla u|) \nabla u). \tag{4}$$

For the diffusivity, they propose for instance functions of type

$$g(s) = \exp\left(\frac{-s^2}{2\lambda^2}\right) \quad (\lambda > 0). \tag{5}$$

In our terminology, the Perona–Malik filter is regarded as an isotropic model, as it utilizes a scalar-valued diffusivity and not a diffusion tensor. Note that for such rapidly decreasing diffusivities like (5), the flux function  $\Phi(s) := sg(s)$  is increasing in  $[0, \lambda]$  and decreasing in  $(\lambda, \infty)$ . The latter property is explicitly intended for enhancing edges by acting locally like a backward diffusion equation [35]. Thus, in spite of its nonnegative diffusivity, the Perona–Malik model exhibits simultaneously forward and backward diffusion areas. As a consequence, the classical theory of existence and uniqueness does not apply anymore, and it seems to be unlikely that there exists a unique smooth solution [11]. Nevertheless, because of its visually impressing results, the Perona–Malik idea has triggered numerous modifications.

By means of spatial or temporal regularization, nonlinear diffusion filters were proposed which are more robust against noise [11, 31, 13, 14, 50, 47, 48, 49]. The isotropic nonlinear filter of Catté, Lions, Morel and Coll [11] is a well-investigated representative. These authors replace the diffusivity  $g(|\nabla u|)$  of the Perona–Malik model by  $g(|\nabla u_\sigma|)$  with  $u_\sigma := K_\sigma * u$  and establish existence, uniqueness and regularity of a solution for  $\sigma > 0$ . Whitaker and Pizer [50] suggest the regularization parameter  $\sigma$  to be a decreasing function in  $t$ .

Further improvements were achieved by anisotropic diffusion models which smooth preferentially *along* edges. This may be accomplished by using a diffusion tensor instead of a scalar

diffusivity [13, 14, 47, 48, 49] or by convolving the image with anisotropic Gaussians [31]. These methods offer either advantages at noisy edges with “jittery” contours or they are well-adapted for the processing of lower-dimensional features.

In order to avoid the difficulty of finding an appropriate stopping time for the diffusion process and to give a theoretical justification for nonlinear diffusion filtering, several authors introduced an additional reaction term [32, 16, 13, 43, 14]. This aims to allow nontrivial steady states. Diffusion filtering can then be regarded as a steepest descent method for some energy functional. (This idea may also be extended to study reaction-diffusion systems [45, 46, 36, 14].) An example for a diffusion filter with a reaction term is the Nordström model [32]

$$\partial_t u = \operatorname{div}(g(|\nabla u|) \nabla u) + \alpha(f - u) \quad (\alpha > 0).$$

However, this alteration just shifts the problem of finding a stopping time  $T$  to the problem of determining  $\alpha$ , a parameter which impacts the deviation of  $u$  from the initial image  $f$ . Concerning the second goal, the convergence to a nontrivial steady state, only conjectures are available [32]. Shah [44] points out that, for diffusivities like (5), the corresponding energy functionals for models of Nordström type are nonconvex and may possess numerous local minima. Thus, it is not clear whether a diffusion process with the original image as initial value will converge to the global minimum or not. Schnörr [43] renounces edge enhancing diffusivities in order to end up with (nonquadratic) convex functionals, for which the standard theory is applicable.

For other diffusion–reaction methods, the convergence problem has usually not been addressed. One exception is the paper of Cottet and Germain [13]. They show that their filter allows the possibility of having almost steady piecewise constant states and prove their attractivity. The stability of these solutions is indicated by showing that piecewise constant states are almost steady-state solutions of a semidiscretized version, where the diffusion tensor is constant with respect to time.

From all these considerations we recognize that, to provide a theoretical framework for edge enhancing nonlinear diffusion, an interpretation as a minimizer of some energy functional does leave several questions open. So in the following, we turn our attention to an alternative way for justifying nonlinear diffusion, namely an interpretation as a scale-space transformation.

In order to get an impression of the concept of scale-spaces, let us recall that images usually contain structures at a large variety of scales. In those cases where it is not clear in advance which is the right scale for the depicted information, it is desirable to have an image representation at multiple scales. Moreover, by comparing the structures at different scales, one obtains a hierarchy of image structures, which eases a subsequent image interpretation.

A scale-space is an image representation at a continuum of scales, embedding the image  $f$  into a family  $\{T_t f \mid t \geq 0\}$  of gradually simplified versions of it. The pioneering work of Alvarez, Guichard, Lions and Morel [1] shows that every scale-space fulfilling some fairly natural architectural, information reducing and invariance properties is governed by a partial differential evolution equation with the original image as initial condition.

The best investigated example of a scale-space is the Gaussian scale-space, which is obtained via the convolution with Gaussians of increasing variance [53]. By (3), we have already seen that this is equivalent to linear diffusion filtering. Numerous theoretical results indicate that this diffusion filter is the only “reasonable” way to define a *linear* scale-space [26, 54, 5, 28, 19, 1]. Nevertheless, linear diffusion filtering does not only blur edges, but also dislocates them when moving from finer to coarser scales. So, edges which are identified at a coarse

scale do not give the right location. Therefore, Witkin [53] proposes to identify edges at coarse scales and trace them back to the original image, but in practise this leads to several numerical difficulties which are generally denoted as the *correspondence problem*.

However, when renouncing the linearity requirement, other possibilities appear: Morphological equations represent an important class of nonlinear scale-spaces. They include the dilation/erosion equations [12, 9, 1, 4]

$$\partial_t u = \pm |\nabla u|,$$

the mean curvature motion (curve shortening flow, Euclidean geometric heat flow) [2, 1, 24]

$$\partial_t u = |\nabla u| \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right), \quad (6)$$

the affine invariant equation [1, 41]

$$\partial_t u = |\nabla u| \left( \operatorname{div} \left( \frac{\nabla u}{|\nabla u|} \right) \right)^{\frac{1}{3}} \quad (7)$$

and combinations of these ideas [1, 25].

An exhaustive axiomatic treatment of these equations may be found in [1]. It can be shown that for  $m = 2$ , the affine invariant equation (7) is the unique equation within this axiomatic framework fulfilling the additional requirement of projection invariance [1]. Therefore, it is sometimes named *fundamental equation of shape analysis* [29]. The morphological equations (6) and (7) can also be classified in a unique way by means of group invariances, see [33] and the references therein.

Morphological transformations possess the property that the filtering result depends only on the level lines of the image and, therefore, they are invariant under any nondecreasing grey level transformation (*grey scale invariance*) [1].

On the other hand, it is evident that a grey scale invariance requirement is not compatible with any contrast dependent image enhancement method. Hence, if one insists in having scale-spaces that allow contrast enhancement, one has to withdraw morphology. This excludes e.g. the nonlinear diffusion filter of Perona and Malik from being grey scale invariant. But this is not the main problem when regarding nonlinear diffusion as a scale-space: Perona and Malik justified the nonlinear diffusion scale-space by virtue of a causality reasoning which was based on a maximum–minimum principle [35]. This interpretation is not very satisfactory, as it is unlikely that the Perona–Malik filter possesses solutions which fulfil the smoothness requirements of the extremum principle.

As a promising alternative to construct nonlinear diffusion scale-spaces, Salden and Florack [40, 15] propose to carry over the linear scale-space theory to the nonlinear case by considering nonlinear diffusion processes which result from special rescalings of the linear one. Unfortunately, the Perona–Malik filter turns out not to belong to this class [15].

By now, we have observed that the theoretical justification of nonlinear diffusion filters both as an image enhancing tool and as a scale-space gives rise to some open questions. On the other hand, both fields appear to be contradictive: scale-spaces aim to smooth the image by reducing the amount of features, while image enhancement objects to “sharpen” the image, to discover structures which are blurred or obscured by noise.

The purpose of the present paper is to address this problem. We will see that the majority of scale-space ideas does not necessarily contradict certain demands on image enhancement,

and that a class of regularized anisotropic diffusion filters is well-suited to function as a bridge combining many advantages of these two areas. For this, we restrict ourselves to discussing the continuous scale-space theory of image enhancing anisotropic diffusion. Semidiscrete and discrete models as well as numerical algorithms are presented in forthcoming publications.

The paper is organized as follows:

Section 2 investigates a general class of  $m$ -dimensional nonlinear filters with a diffusion tensor instead of a scalar diffusivity. It extends the existence, uniqueness and regularity results of the isotropic model of Catté, Lions, Morel and Coll [11] to the anisotropic case. We do not impose any additional reaction term, as we have seen that this might create more problems than it solves. In order to demonstrate that special representatives of our class allow the restoration of fairly noisy images, we investigate processes which smooth in an isotropic way within the interior of a region and act anisotropically near edges where they mainly diffuse along the edge structure.

In the third section, a scale-space interpretation is presented. After reviewing some basics of scale-space theory, the invariances and smoothing properties of anisotropic diffusion filtering are analysed and juxtaposed to linear and morphological scale-spaces. The numerous smoothing and information reducing properties are due to an extremum principle and the Lyapunov functionals associated with nonlinear diffusion filtering. As a consequence of the extremum principle, it is proved that the solution depends continuously on the initial image. Section 4 demonstrates how these ideas may be generalized to more sophisticated descriptors of local structure (allowing corner enhancement) and to vector-valued images.

## 2 Nonlinear anisotropic diffusion

### 2.1 The general model

Let  $\Omega := (0, a_1) \times \dots \times (0, a_m)$  be an  $m$ -dimensional image domain and  $\Gamma := \partial\Omega$  its boundary. Then, a filtered version  $u(x, t)$  of a real-valued image  $f(x) \in L^\infty(\Omega)$  may be obtained as the solution of a diffusion equation with  $f$  as initial condition and reflecting boundary conditions:

$$\partial_t u = \operatorname{div}(D(\nabla u_\sigma) \nabla u) \quad \text{on} \quad \Omega \times (0, \infty) \quad (8)$$

$$u(x, 0) = f(x) \quad \text{on} \quad \Omega \quad (9)$$

$$\langle D(\nabla u_\sigma) \nabla u, n \rangle = 0 \quad \text{on} \quad \Gamma \times (0, \infty) \quad (10)$$

Hereby,  $n$  denotes the outer normal and  $\langle \cdot, \cdot \rangle$  the usual inner product on  $\mathbb{R}^m$ . The diffusion tensor  $D \in \mathbb{R}^{m \times m}$  is chosen to be a function of the edge estimator  $\nabla u_\sigma$ , where

$$u_\sigma(x, t) := (K_\sigma * \tilde{u}(\cdot, t))(x) \quad (\sigma > 0) \quad (11)$$

and  $\tilde{u}$  denotes an extension of  $u$  from  $\Omega$  to  $\mathbb{R}^m$ , which may be obtained by mirroring at  $\Gamma$ . The regularization by convolving with a Gaussian makes the edge detection insensitive to noise at scales smaller than  $\sigma$  and helps to ensure existence and uniqueness results (cf. [11]). Now, the question arises of how to choose  $D(\nabla u_\sigma)$ . Let  $\{v_1, \dots, v_m\}$  be an orthonormal basis of  $\mathbb{R}^m$  with  $v_1 \parallel \nabla u_\sigma$ . Then we construct the matrix  $D(\nabla u_\sigma)$  in such a way that it is symmetric positive definite and  $\{v_1, \dots, v_m\}$  represent its eigenvectors. The corresponding eigenvalues  $\lambda_1, \dots, \lambda_m$  are chosen to be functions of  $|\nabla u_\sigma|$  only. Hence, we have

$$\begin{aligned} D(\nabla u_\sigma) &= \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix} \operatorname{diag}(\lambda_1(|\nabla u_\sigma|), \dots, \lambda_m(|\nabla u_\sigma|)) \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix}^T \\ &= S(\nabla u_\sigma) \Lambda(|\nabla u_\sigma|) S(\nabla u_\sigma)^T, \end{aligned} \quad (12)$$

where  $S(\nabla u_\sigma)$  is built of the row vectors  $v_1, \dots, v_m$  and  $\Lambda(|\nabla u_\sigma|)$  is a diagonal matrix with entries  $\lambda_1(|\nabla u_\sigma|), \dots, \lambda_m(|\nabla u_\sigma|)$ . In the present paper, we require the subsequent properties:

$$D : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m} \quad \text{is Lipschitz continuous,} \quad (13)$$

$$\lambda_i \in C^\infty([0, \infty); (0, 1]), \quad i = 1, \dots, m, \quad (14)$$

$$\lambda_2(z) = \dots = \lambda_m(z) \quad \forall z \geq 0. \quad (15)$$

So the problem we are concerned with is as follows:

$$\left. \begin{array}{l} \text{Solve the initial boundary value problem (8)–(10) with an initial} \\ \text{value } f \in L^\infty(\Omega) \text{ and a diffusion tensor according to (12)–(15).} \end{array} \right\} \quad (P)$$

In order to discuss existence and uniqueness results, we first have to introduce some notations. Let  $H^1(\Omega)$  be the Sobolev space of functions  $u(x) \in L^2(\Omega)$  with all distributional derivatives of first order being in  $L^2(\Omega)$ . We equip  $H^1(\Omega)$  with the norm

$$\|u\|_{H^1(\Omega)} := \left( \|u\|_{L^2(\Omega)}^2 + \sum_{i=1}^m \|\partial_{x_i} u\|_{L^2(\Omega)}^2 \right)^{1/2}$$

and identify it with its dual space. We denote by  $L^2(0, T; H^1(\Omega))$  the space of functions  $u$ , strongly measurable on  $[0, T]$  with range in  $H^1(\Omega)$  (for the Lebesgue measure  $dt$  on  $[0, T]$ ) such that

$$\|u\|_{L^2(0, T; H^1(\Omega))} := \left( \int_0^T \|u(t)\|_{H^1(\Omega)}^2 dt \right)^{1/2} < \infty.$$

In a similar way,  $C([0, T]; L^2(\Omega))$  is defined as the space of continuous functions  $u : [0, T] \rightarrow L^2(\Omega)$  provided with the norm

$$\|u\|_{C([0, T]; L^2(\Omega))} := \max_{[0, T]} \|u(t)\|_{L^2(\Omega)}.$$

With these notations, our filter class (P) satisfies the subsequent theorem:

**Theorem 1 (Existence, uniqueness and regularity).**

*For any  $T > 0$ , there exists a unique function  $u(x, t)$  that solves problem (P) in the distributional sense. This function verifies*

$$\begin{aligned} u &\in C([0, T]; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega)), \\ \partial_t u &\in L^2(0, T; H^1(\Omega)). \end{aligned}$$

Moreover,  $u \in C^\infty(\bar{\Omega} \times (0, T])$ .

**Proof:** This is a straightforward extension of the proof for the isotropic model [11] with a scalar diffusivity to the anisotropic case with a diffusion tensor.  $\square$

Before turning to a more detailed investigation of further theoretical properties of the filter (P), let us first investigate how it may be used for image enhancement.

## 2.2 Applications to image enhancement

The functional dependence of the eigenvalues on  $|\nabla u_\sigma|$  can be adapted to the desired purpose of the diffusion filter. Frequently, one chooses the eigenvalues in such a way that they are close to 1 in the interior of a region where one desires strong noise elimination by isotropic linear diffusion. At edges, one usually reduces diffusion in normal direction in order to preserve localization and structure of the edge. If noise at edges is to be eliminated as well, this may be accomplished by encouraging diffusion along the edge.

To give an example how such a diffusion tensor may be constructed from an isotropic model, let us consider a Lipschitz continuous scalar diffusivity function  $g \in C^\infty([0, \infty), (0, 1])$  which is represented in  $[0, \infty)$  by a convergent power series [48, 38]:

$$g(s) = \sum_{k=0}^{\infty} a_k s^k.$$

Since the tensor product

$$J_0(\nabla u_\sigma) := \nabla u_\sigma \otimes \nabla u_\sigma \quad (16)$$

gives a positive semidefinite matrix there exists a unique positive semidefinite matrix  $\sqrt{J_0}$  with

$$\sqrt{J_0} \cdot \sqrt{J_0} = J_0.$$

Hence,

$$D(\nabla u_\sigma) := g\left(\sqrt{J_0(\nabla u_\sigma)}\right) := \sum_{k=0}^{\infty} a_k \left(\sqrt{J_0(\nabla u_\sigma)}\right)^k \quad (17)$$

defines a diffusion tensor possessing an orthonormal system of eigenvectors  $v_1, \dots, v_m$  with  $v_1 \parallel \nabla u_\sigma$ . The corresponding eigenvalues are

$$\begin{aligned} \lambda_1(|\nabla u_\sigma|) &= g(|\nabla u_\sigma|). \\ \lambda_2(|\nabla u_\sigma|) &= \dots = \lambda_m(|\nabla u_\sigma|) = g(0). \end{aligned}$$

A class of admissible diffusivities (comprising (5)) is given by

$$g(s) = \exp\left(\frac{-s^\alpha}{\alpha \lambda^\alpha}\right) \quad (18)$$

with a contrast parameter  $\lambda > 0$  and a decay parameter  $\alpha \geq 1$ . The flux function  $\Phi(s) = sg(s)$  is increasing in  $[0, \lambda]$  and decreasing in  $(\lambda, \infty)$ . Larger values of  $\alpha$  correspond to a faster decay of  $g$  near  $\lambda$ . They lead to steeper edges with a longer “lifetime”, while smaller values diminish staircasing at edges. A detailed investigation of this family of diffusivities is given in [6]. All figures in the present paper were obtained with  $\alpha := 5$ , in order to demonstrate that the choice of  $\alpha$  is not critical and that this is not a crucial parameter.

In practise, anisotropic diffusion needs essentially two parameters, which are fairly natural for any image restoration method (cf. [11]): a contrast parameter  $\lambda$  and a resolution parameter  $\sigma$ . Gradients with modulus larger than  $\lambda$  are considered to be edges and, therefore, diffusion is reduced significantly. Details being smaller than  $\sigma$  are regarded as noise or unimportant features, which are to be removed.

The stopping time  $T$  can be related to  $\sigma$  [11]: in the interior of a region (where  $|\nabla u_\sigma|$  is close to 0), the diffusion process becomes almost linear and isotropic with diffusivity 1. Hence, it takes a time  $T = O(\sigma^2)$  to eliminate details of scale  $\sigma$ . Anyway, since diffusion is reduced at edges, one should regard  $O(\sigma^2)$  rather as a lower bound for the stopping time  $T$ . Note that

– especially in higher dimensions – the stopping time reasoning by virtue of linear diffusion fits the anisotropic case of nonlinear filtering better than the isotropic case: at (codimension 1) edges, the anisotropic model still possesses  $m - 1$  directions with diffusivity (eigenvalue) 1, while the isotropic model inhibits diffusion in all directions. For longer times the following is observed: edges are hardly affected, while the interior of each region becomes almost completely diffused. This leads to nearly piecewise constant images and qualifies anisotropic diffusion as a preprocessing step which makes subsequent segmentations very easy. However, as Theorem 5 will show, the diffusion process converges finally to a constant steady state. We will see that this behaviour is rather an advantage of the nonlinear diffusion scale-space than a shortcoming of a restoration method.

Let us now investigate two examples which demonstrate the image enhancing potential of anisotropic diffusion filtering using tensors such as (17) with  $g$  according to (18). Figure 1 shows an image with a cross, degraded by additive Gaussian noise of standard deviation 100 and its processed version. Note that all images in the present paper are depicted such that the grey values 0 and 255 appear black and white, respectively. Except for small roundings at corners, anisotropic diffusion could restore the original image almost entirely. Figure 2 illustrates that regularized anisotropic diffusion filtering is still capable of possessing the contrast enhancing properties of the Perona–Malik filter (provided that  $\sigma$  is not too large). It depicts a 2D Gaussian-like function with its isolines before and after diffusion filtering. It can be observed that two regions with almost constant grey value evolve which are separated by a fairly steep edge. Edge enhancement is caused by the fact that due to the rapidly decreasing diffusivity, smoothing within each region is strongly preferred to diffusion between two adjacent regions.

The discussed model has several relations to other anisotropic ideas. An alternative anisotropic model also fulfilling the requirements (12)–(15) was proposed in earlier works of the author [47, 48]. In [48], it was also demonstrated that, at noisy edges, anisotropic diffusion performs much better than isotropic nonlinear diffusion. A different approach to anisotropic ideas is due to Nitzberg and Shiota [31]. They convolve the original image with elongated and displaced Gaussians which are adapted to the local structure. In a special scaling limit, this may also be regarded as an anisotropic diffusion filtering. Cottet and Germain [13] use a positive semidefinite diffusion tensor with  $\lambda_1 \equiv 0$  which offers certain advantages for the processing of one-dimensional features. Salden [40] proposes a diffusion tensor utilizing intrinsic properties of isophotes and flowlines.

### 3 Scale-space properties

#### 3.1 The concept of scale-space

As already mentioned, a scale-space has to fulfil certain architectural, smoothing and invariance properties. This will be illustrated briefly in the sequel. For more details, the reader is referred to [1] from which most of the following nomenclature is borrowed.

A typical representative for an architectural property is *recursivity*, i.e. for  $t = 0$ , the scale-space representation gives the original image  $f$ , and the filtering may be split into a sequence of filter banks:

$$\begin{aligned} T_0 f &= f, \\ T_{t+s} f &= T_t(T_s f) \quad \forall s, t \geq 0. \end{aligned}$$

This property is also sometimes referred to as the *semigroup property*. Other architectural



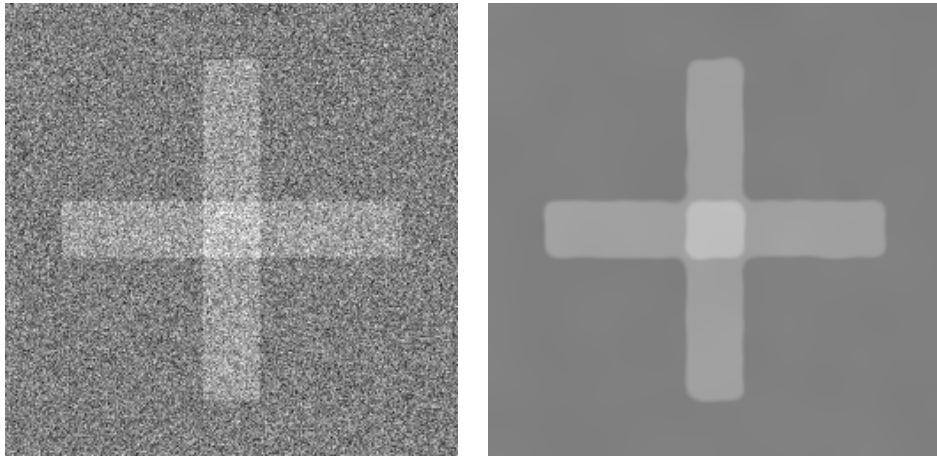


Figure 1: Restoration properties of anisotropic diffusion. LEFT: Cross degraded by noise,  $\Omega = (0, 255)^2$ . RIGHT: Restored,  $\lambda = 2$ ,  $\sigma = 2$ ,  $t = 80$ .

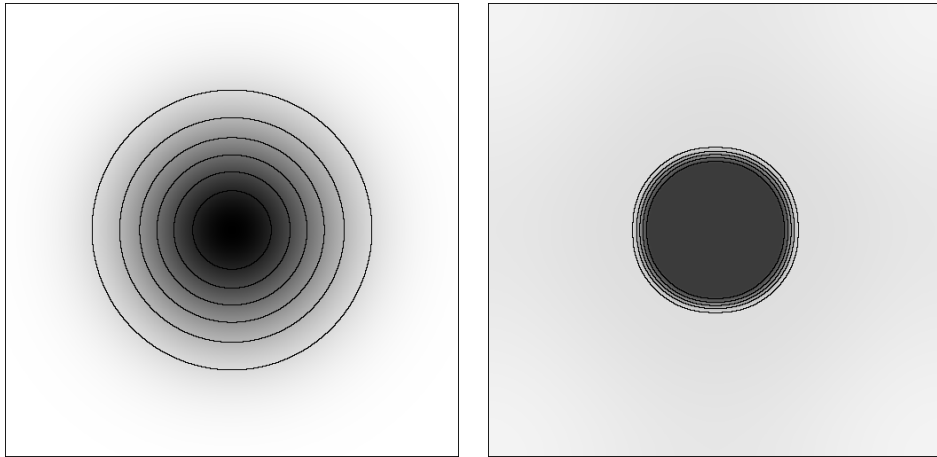


Figure 2: Edge enhancement of anisotropic diffusion. LEFT: Gaussian-type function,  $\Omega = (0, 255)^2$ . RIGHT: Filtered,  $\lambda = 3$ ,  $\sigma = 2$ ,  $t = 1800$ .

axioms are e.g. concerned with regularity properties of  $T_t$  and the local behaviour as  $t$  tends to 0.

Information reduction arises from the wish that the smoothing transformation should not create artefacts when passing from fine to coarse representation. Thus, at a coarse scale, we must not have additional structures which are caused by the filtering method itself and not by underlying structures at finer scales. This property is specified by numerous authors in different ways, using concepts like causality [26], nonenhancement of local extrema [28], no creation of new level-crossings [20, 28], comparison principle [1], maximum–minimum principle [35], decay of Euclidean absolute curvature [41], and decay of the number of extrema and zero-crossings of the curvature [41].

It is desirable that – except for information reduction by smoothing – the scale-space transform does not alter the image too much, i.e. it should be invariant with respect to many transformations. Examples of such transformations are grey-level shifts, translations and rotations, but also affine transformations of the space.

In the present paper, we shall not focus on further investigations of architectural requirements like recursivity, regularity and locality, as these qualities do not distinguish nonlinear diffusion scale-spaces from other ones. We start with briefly discussing invariances. Afterwards, we turn to a more crucial task, namely the question, in which sense our restoration method – which allows edge enhancement – can still be considered as a smoothing, information reducing image transformation.

## 3.2 Invariances

Let  $u(x,t)$  be the unique solution of (P) and define the scale-space operator  $T_t$  by

$$T_t f := u(t), \quad t \geq 0,$$

where  $u(t) := u(\cdot, t)$ .

The properties we discuss below illustrate that an invariance of  $T_t$  with respect to some image transformation  $P$  is characterized by the fact that  $T_t$  and  $P$  commute.

### 3.2.1 Grey level shift invariance

Since the diffusion tensor is only a function of  $\nabla u_\sigma$ , but not of  $u$ , we may shift the grey level range by an arbitrary constant  $C$ , and the filtered images will also be shifted by the same constant. Moreover, a constant function is not affected by diffusion filtering. Therefore, we have

$$\begin{aligned} T_t(0) &= 0, \\ T_t(f + C) &= T_t(f) + C \quad \forall t \geq 0. \end{aligned}$$

### 3.2.2 Reverse contrast invariance

By the construction of the diffusion tensor, we have  $D(-\nabla u_\sigma) = D(\nabla u_\sigma)$ , and it follows that

$$T_t(-f) = -T_t(f) \quad \forall t \geq 0.$$

This property is not fulfilled by classical morphological scale-space equations like dilation and erosion. When reversing the contrast, the role of dilation and erosion has to be exchanged as well.

### 3.2.3 Average grey level invariance

Average grey level invariance is a further property in which diffusion scale-spaces differ from morphological scale-spaces. In general, the evolution PDEs of the latter ones are not of divergence form and do not preserve the mean grey value.

**Proposition 1 (Conservation of average grey value).**

*The average grey level*

$$\mu := \frac{1}{|\Omega|} \int_{\Omega} f(x) dx \quad (19)$$

*is not affected by nonlinear diffusion filtering:*

$$\frac{1}{|\Omega|} \int_{\Omega} T_t f dx = \mu \quad \forall t > 0.$$

**Proof:** Define  $I(t) := \int_{\Omega} u(x, t) dx$  for all  $t \geq 0$ . Then the Cauchy–Schwarz inequality implies

$$|I(t) - I(0)| = \left| \int_{\Omega} (u(x, t) - f(x)) dx \right| \leq |\Omega|^{1/2} \|u(t) - f\|_{L^2(\Omega)}.$$

Since  $u \in C([0, T]; L^2(\Omega))$ , the preceding inequality gives the continuity of  $I(t)$  in 0. For  $t > 0$ , Theorem 1, the divergence theorem and the boundary conditions yield

$$\frac{dI}{dt} = \int_{\Omega} \partial_t u dx = \int_{\Gamma} \langle D(\nabla u_{\sigma}) \nabla u, n \rangle dS = 0.$$

Hence,  $I(t)$  must be constant for all  $t \geq 0$ . □

Average grey level invariance may be described by commuting operators, when introducing an averaging operator  $M : L^1(\Omega) \rightarrow L^1(\Omega)$  which maps  $f$  to a constant image with the same mean grey level:

$$(Mf)(y) := \frac{1}{|\Omega|} \int_{\Omega} f(x) dx \quad \forall y \in \Omega. \quad (20)$$

Then Proposition 1 and grey level shift invariance imply that the order of  $M$  and  $T_t$  is exchangeable:

$$M(T_t f) = T_t(Mf) \quad \forall t \geq 0.$$

When studying diffusion filtering as a pure initial value problem in the domain  $\mathbb{R}^m$ , it also makes sense to investigate Euclidean transformations of an image. This leads us to translation and isometry invariance.

### 3.2.4 Translation invariance

Define a translation  $\tau_h$  by  $(\tau_h f)(x) := f(x + h)$ . Then diffusion filtering fulfils

$$T_t(\tau_h f) = \tau_h(T_t f) \quad \forall t \geq 0.$$

This is a consequence of the fact that the diffusion tensor is a function of  $\nabla u_{\sigma}$  solely, but not of  $x$ .

### 3.2.5 Isometry invariance

Let  $R \in \mathbb{R}^{m \times m}$  be an orthogonal transformation. If we apply  $R$  to  $f$  by defining  $Rf(x) := f(Rx)$ , then the eigenvalues of the diffusion tensor are unaltered and any eigenvector  $v$  is transformed into  $Rv$ . Thus, it makes no difference whether the orthogonal transformation is applied before or after diffusion filtering:

$$T_t(Rf) = R(T_t f) \quad \forall t \geq 0.$$

## 3.3 Information reducing properties

### 3.3.1 Nonenhancement of local extrema

The requirement that a scale-space representation must not amplify local extrema was first pointed out by Lindeberg [28] for the linear diffusion scale-space. However, this condition is also verified by nonlinear anisotropic diffusion.

#### Theorem 2 (Nonenhancement of local extrema).

Let  $u$  be the unique solution of (P) and consider some  $\theta > 0$ . Suppose that  $\xi \in \Omega$  is a local extremum of  $u(\cdot, \theta)$ . Then,

$$\begin{aligned} \partial_t u(\xi, \theta) &\leq 0, & \text{if } \xi \text{ is a local maximum,} \\ \partial_t u(\xi, \theta) &\geq 0, & \text{if } \xi \text{ is a local minimum.} \end{aligned}$$

**Proof:** Let  $D(\nabla u_\sigma) =: (d_{ij}(\nabla u_\sigma))$ . Then we have

$$\partial_t u = \sum_{i=1}^m \sum_{j=1}^m \left( \partial_{x_i} d_{ij}(\nabla u_\sigma) \right) \partial_{x_j} u + \sum_{i=1}^m \sum_{j=1}^m d_{ij}(\nabla u_\sigma) \partial_{x_i x_j} u. \quad (21)$$

Since  $\nabla u(\xi, \theta) = 0$  and  $\partial_{x_i} d_{ij}(\nabla u_\sigma(\xi, \theta))$  is bounded, the first term of the right hand side of (21) vanishes in  $(\xi, \theta)$ .

We know that the diffusion tensor  $D := D(\nabla u_\sigma(\xi, \theta))$  is positive definite. Hence, there exists an orthogonal transformation  $S \in \mathbb{R}^{m \times m}$  such that

$$S^T D S = \text{diag}(\lambda_1, \dots, \lambda_m) =: \Lambda$$

with  $\lambda_1, \dots, \lambda_m$  being the positive eigenvalues of  $D(\nabla u_\sigma(\xi, \theta))$  (cf. also (12)).

Now, let us assume that  $(\xi, \theta)$  is a local maximum. Then the Hessian  $H := \text{Hess}(u(\xi, \theta))$  and  $B := (b_{ij}) := S^T H S$  are negative semidefinite. Therefore, we have

$$b_{ii} \leq 0 \quad \forall i = 1, \dots, m$$

and by the invariance of the trace with respect to orthogonal transformations it follows that

$$\begin{aligned} \partial_t u(\xi, \theta) &= \text{trace}(DH) \\ &= \text{trace}(S^T D S S^T H S) \\ &= \text{trace}(\Lambda B) \\ &= \sum_{i=1}^m \lambda_i b_{ii} \\ &\leq 0. \end{aligned}$$

If  $\xi$  is a local minimum of  $u(x, \theta)$ , one proceeds in the same way utilizing the positive semidefiniteness of the Hessian.  $\square$

Nonenhancement of local extrema distinguishes anisotropic diffusion from classical contrast enhancing methods such as high-frequency emphasis ([18], pp. 182–183), which do violate this principle. Although possibly being in the backward diffusion region at edges, nonlinear diffusion is always in the forward region at extrema. This ensures its stability. However, it should be noted that the preceding theorem does not imply that the number of local extrema is nonincreasing. It is not very difficult to give illustrative counterexamples where this is yet violated for the linear diffusion case in two dimensions [27, 28].

### 3.3.2 The maximum–minimum principle

From the nonenhancement of local extrema, it can be shown by using a reasoning of Illner [21] that for  $t_2 \geq t_1 > 0$  we get

$$\min_{y \in \bar{\Omega}} u(y, t_1) \leq u(x, t_2) \leq \max_{y \in \bar{\Omega}} u(y, t_1) \quad \forall x \in \bar{\Omega}.$$

Instead of doing this in the sequel, we are interested a more general result which does also apply to extrema in the initial data.

A common way to prove extremum principles for nonlinear parabolic PDEs is the use of monotony results (see e.g. [37], pp. 186–188). The monotony condition

$$f \leq h \quad \Longrightarrow \quad T_t f \leq T_t h \quad \forall t > 0.$$

has also been proposed as a smoothing requirement for scale-spaces [1]. When combined with reverse contrast invariance and grey level shift invariance, this comparison principle implies an extremum principle. However, to establish the monotony condition for anisotropic diffusion filtering, we would need that the equation is always of forward parabolic type. Since this would forbid contrast enhancing processes, we will not pursue this idea any further.

In order to prove a maximum–minimum principle, we utilize Stampacchia’s truncation method instead (cf. [8], p. 211).

#### Theorem 3 (Extremum principle).

*Consider the problem (P) and define*

$$a := \operatorname{ess\,inf}_{x \in \Omega} f(x), \tag{22}$$

$$b := \operatorname{ess\,sup}_{x \in \Omega} f(x). \tag{23}$$

*Then the solution  $u$  of (P) verifies*

$$a \leq u(x, t) \leq b \quad \text{on} \quad \bar{\Omega} \times (0, \infty). \tag{24}$$

**Proof:** We restrict ourselves in proving only the maximum principle. The minimum principle follows from the reverse contrast invariance and the maximum principle when applied to the initial datum  $-f$ .

Let  $G \in C^1(\mathbb{R})$  be a function with  $G(s) = 0$  on  $(-\infty, 0]$  and  $0 < G'(s) \leq C$  on  $(0, \infty)$  for some constant  $C$ . Now, we define

$$H(s) := \int_0^s G(\sigma) d\sigma, \quad s \in \mathbb{R},$$

$$\varphi(t) := \int_{\Omega} H(u(x, t) - b) dx, \quad t \in [0, \infty).$$

By the Cauchy–Schwarz inequality, we have

$$\int_{\Omega} |G(u(x,t)-b) \partial_t u(x,t)| dx \leq C \cdot \|u(t)-b\|_{L^2(\Omega)} \cdot \|\partial_t u(t)\|_{L^2(\Omega)}$$

and from Theorem 1 we know that the right hand side of this estimate exists. Therefore,  $\varphi$  is differentiable for  $t > 0$ , and we get

$$\begin{aligned} \frac{d\varphi}{dt} &= \int_{\Omega} G(u-b) \partial_t u dx \\ &= \int_{\Omega} G(u-b) \operatorname{div} (D(\nabla u_{\sigma}) \nabla u) dx \\ &= \int_{\Gamma} G(u-b) \underbrace{\langle D(\nabla u_{\sigma}) \nabla u, n \rangle}_{=0} dS - \int_{\Omega} \underbrace{G'(u-b)}_{\geq 0} \underbrace{\langle \nabla u, D(\nabla u_{\sigma}) \nabla u \rangle}_{\geq 0} dx \\ &\leq 0. \end{aligned} \tag{25}$$

By means of  $H(s) \leq \frac{C}{2}s^2$ , we have

$$0 \leq \varphi(t) \leq \int_{\Omega} H(u(x,t)-f(x)) dx \leq \frac{C}{2} \|u(t)-f\|_{L^2(\Omega)}^2. \tag{26}$$

Since  $u \in C([0, T]; L^2(\Omega))$ , the right hand side of (26) tends to  $0 = \varphi(0)$  for  $t \rightarrow 0^+$  which proves the continuity of  $\varphi(t)$  in 0. Now from

$$\varphi \in C([0, \infty)), \quad \varphi(0) = 0, \quad \varphi \geq 0 \quad \text{on} \quad [0, \infty)$$

and (25), it follows that

$$\varphi \equiv 0 \quad \text{on} \quad [0, \infty).$$

Hence, for all  $t \geq 0$ , we obtain  $u(x, t) - b \leq 0$  almost everywhere (a.e.) on  $\Omega$ . Due to the smoothness of  $u$  for  $t > 0$ , we finally end up with the assertion

$$u(x, t) \leq b \quad \text{on} \quad \bar{\Omega} \times (0, \infty).$$

□

Hummel [20] shows that, under certain conditions, the extremum principle for parabolic operators is equivalent to the property that the corresponding scale-space never creates additional level-crossings for  $t > 0$ . This points out the importance of extremum principles for scale-spaces. Furthermore, Theorem 3 is a useful tool for establishing theoretical properties like continuous dependence of the solution on the initial data. To demonstrate this, we first have to prove two lemmas.

**Lemma 1 (Uniform gradient estimate for Gaussian convolution).**

Let  $\sigma > 0$  and  $A(c_1) := \{u \in L^\infty(\mathbb{R}^m) \mid \|u\|_{L^\infty(\mathbb{R}^m)} \leq c_1\}$ .

Then there exists a constant  $c_2 = c_2(\sigma, m, c_1)$  such that

$$|(\nabla K_{\sigma} * u)(x)| \leq c_2 \quad \forall u \in A(c_1), \quad \forall x \in \mathbb{R}^m.$$

**Proof:** Denoting the one-dimensional Gaussian by  $\kappa_{\sigma}(s) := \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{s^2}{2\sigma^2}\right)$  and using  $\int_{-\infty}^{\infty} \kappa_{\sigma}(s) ds = 1$  and  $K_{\sigma}(x) = \prod_{i=1}^m \kappa_{\sigma}(x_i)$ , one immediately obtains

$$\int_{\mathbb{R}^m} |(\partial_{x_i} K_{\sigma})(x)| dx = \int_{\mathbb{R}} |\partial_{x_i} \kappa_{\sigma}(x_i)| dx_i = \frac{2}{\sqrt{2\pi\sigma}}.$$

This gives the estimate

$$\begin{aligned}
|(\nabla K_\sigma * u)(x)| &\leq \sum_{i=1}^m |(\partial_{x_i} K_\sigma * u)(x)| \\
&\leq \sum_{i=1}^m \int_{\mathbb{R}^m} |\partial_{y_i} K_\sigma(y)| \cdot |u(x-y)| dy \\
&\leq \|u\|_{L^\infty(\mathbb{R}^m)} \cdot \sum_{i=1}^m \int_{\mathbb{R}^m} |\partial_{y_i} K_\sigma(y)| dy \\
&\leq \frac{2mc_1}{\sqrt{2\pi}\sigma}.
\end{aligned}$$

□

Now, by Theorem 3 and Lemma 1, we know that  $|\nabla u_\sigma(x, t)|$  is bounded on  $\bar{\Omega} \times (0, \infty)$ . This gives immediately

**Lemma 2 (Uniform positive definiteness of the diffusion tensor).**

Let  $u$  be the solution of (P). Then there exists a positive constant  $\nu = \nu(\sigma, m, \|f\|_{L^\infty(\Omega)})$  such that

$$\nu |y|^2 \leq \langle y, D(\nabla u_\sigma(x, t))y \rangle \quad \forall (x, t) \in \bar{\Omega} \times [0, \infty), \quad \forall y \in \mathbb{R}^m.$$

With this lemma, we may prove the subsequent theorem which is an important step for establishing the well-posedness of the method (P).

**Theorem 4 (Continuous dependence on the initial data).**

The unique solution of (P) depends continuously on the initial value with respect to  $\|\cdot\|_{L^2(\Omega)}$ .

**Proof:** Let  $f, h \in L^\infty(\Omega)$  be two initial values for problem (P) and  $u, w$  the corresponding solutions. In the same way as in the uniqueness proof in [11], one shows that there exists some constant  $c > 0$  such that

$$\frac{d}{dt} \left( \|u(t) - w(t)\|_{L^2(\Omega)}^2 \right) \leq c \cdot \|\nabla u(t)\|_{L^2(\Omega)}^2 \cdot \|u(t) - w(t)\|_{L^2(\Omega)}^2.$$

Applying the Gronwall–Bellman lemma ([7], pp. 156–157) yields

$$\|u(t) - w(t)\|_{L^2(\Omega)}^2 \leq \|f - h\|_{L^2(\Omega)}^2 \cdot \exp \left( c \cdot \int_0^t \|\nabla u(s)\|_{L^2(\Omega)}^2 ds \right).$$

Then for  $t \in [0, T]$ , Lemma 2 implies that there exists some constant  $\nu > 0$ , such that

$$\begin{aligned}
\int_0^t \|\nabla u(s)\|_{L^2(\Omega)}^2 ds &\leq \int_0^T \|\nabla u(s)\|_{L^2(\Omega)}^2 ds \\
&\leq \frac{1}{\nu} \int_0^T \left| \int_{\Omega} \langle \nabla u(x, s), D(\nabla u_\sigma(x, s)) \nabla u(x, s) \rangle dx \right| ds \\
&= \frac{1}{\nu} \int_0^T \left| \int_{\Omega} u(x, s) \cdot \operatorname{div} \left( D(\nabla u_\sigma(x, s)) \nabla u(x, s) \right) dx \right| ds \\
&\leq \frac{1}{\nu} \int_0^T \|u(s)\|_{L^2(\Omega)} \|\partial_t u(s)\|_{L^2(\Omega)} ds \\
&\leq \frac{1}{\nu} \|u\|_{L^2(0, T; H^1(\Omega))} \|\partial_t u\|_{L^2(0, T; H^1(\Omega))}.
\end{aligned}$$

By virtue of Theorem 1, we know that the right hand side of this estimate exists. Now, let  $\epsilon > 0$  and choose

$$\delta := \epsilon \cdot \exp\left(\frac{-c}{2\nu} \|u\|_{L^2(0,T;H^1(\Omega))} \|\partial_t u\|_{L^2(0,T;H^1(\Omega))}\right).$$

Then for  $\|f - h\|_{L^2(\Omega)} < \delta$ , the preceding results imply

$$\|u(t) - w(t)\|_{L^2(\Omega)} < \epsilon \quad \forall t \in [0, T].$$

This proves the assertion.  $\square$

Theorem 4 is especially important when applying anisotropic diffusion in areas like stereo vision or analysis of image sequences, as it guarantees that similar images remain similar after being processed.

### 3.3.3 Lyapunov functionals and behaviour for $t \rightarrow \infty$

Since scale-spaces are intended to simplify an image, it is of importance, that for  $t \rightarrow \infty$ , we obtain the simplest possible image representation, namely a constant image with the same average grey value as the original one. Unlike morphological equations of type (6) and (7), the following theorem states, that anisotropic diffusion filtering always leads to a constant steady-state. This is due to the class of Lyapunov functionals associated with the diffusion process.

#### Theorem 5 (Lyapunov functionals and behaviour for $t \rightarrow \infty$ ).

Suppose that  $u$  is the solution of (P) and let  $a, b, \mu$  and  $M$  be defined as in (22), (23), (19) and (20), respectively. Then the following properties are valid:

(a) (Lyapunov functionals)

For all  $r \in C^2([a, b])$  with  $r'' \geq 0$  on  $[a, b]$ , the function

$$V(t) := \Phi(u(t)) := \int_{\Omega} r(u(x, t)) dx$$

is a weak Lyapunov functional:

- (i)  $\Phi(u(t)) \geq \Phi(Mf)$  for all  $t \geq 0$ .
- (ii)  $V \in C([0, \infty)) \cap C^1((0, \infty))$  and  $V'(t) \leq 0$  for all  $t > 0$ .

Moreover, if  $r'' > 0$  on  $[a, b]$ , then  $V(t) = \Phi(u(t))$  is a strict Lyapunov functional:

- (iii)  $\Phi(u(t)) = \Phi(Mf) \iff \begin{cases} u(t) = Mf & \text{on } \bar{\Omega} & (\text{if } t > 0) \\ u(t) = Mf & \text{a.e. on } \Omega & (\text{if } t = 0) \end{cases}$
- (iv) If  $t > 0$ , then  $V'(t) = 0$  if and only if  $u(t) = Mf$  on  $\bar{\Omega}$ .
- (v)  $V(0) = V(T)$  for  $T > 0 \iff \begin{cases} f = Mf & \text{a.e. on } \Omega & \text{and} \\ u(t) = Mf & \text{on } \bar{\Omega} \times (0, T] \end{cases}$

(b) (Convergence)

- (i)  $\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^p(\Omega)} = 0$  for  $1 \leq p < \infty$ .
- (ii) For  $m = 1$ , the convergence  $\lim_{t \rightarrow \infty} u(x, t) = \mu$  is uniform on  $\bar{\Omega}$ .



**Proof:**

- (a) (i) Average grey level invariance and the convexity of  $r$  give for  $t \geq 0$ ,

$$\begin{aligned}
\Phi(Mf) &= \int_{\Omega} r \left( \frac{1}{|\Omega|} \int_{\Omega} u(x, t) dx \right) dy \\
&\leq \int_{\Omega} \left( \frac{1}{|\Omega|} \int_{\Omega} r(u(x, t)) dx \right) dy \\
&= \int_{\Omega} r(u(x, t)) dx \\
&= \Phi(u(t)).
\end{aligned} \tag{27}$$

- (ii) Let us start by proving the continuity of  $V(t)$  in 0. Thanks to the maximum–minimum principle, we may choose a constant

$$L := \max_{s \in [a, b]} |r'(s)|$$

such that for all  $t > 0$ , the Lipschitz condition

$$|r(u(x, t)) - r(f(x))| \leq L |u(x, t) - f(x)|$$

is verified a.e. on  $\Omega$ . From this and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
|V(t) - V(0)| &\leq |\Omega|^{1/2} \|r(u(t)) - r(f)\|_{L^2(\Omega)} \\
&\leq |\Omega|^{1/2} L \|u(t) - f\|_{L^2(\Omega)},
\end{aligned}$$

and by virtue of  $u \in C([0, T]; L^2(\Omega))$ , the limit  $t \rightarrow 0^+$  gives the announced continuity in 0.

By Theorem 1 and the boundedness of  $r'$  on  $[a, b]$ , we know that  $V$  is differentiable for  $t > 0$  and  $V'(t) = \int_{\Omega} r'(u) u_t dx$ . Thus, the divergence theorem yields

$$\begin{aligned}
V'(t) &= \int_{\Omega} r'(u) \operatorname{div}(D(\nabla u_{\sigma}) \nabla u) dx \\
&= \int_{\Gamma} r'(u) \underbrace{\langle D(\nabla u_{\sigma}) \nabla u, n \rangle}_{=0} dS - \int_{\Omega} \underbrace{r''(u)}_{\geq 0} \underbrace{\langle \nabla u, D(\nabla u_{\sigma}) \nabla u \rangle}_{\geq 0} dx \\
&\leq 0.
\end{aligned}$$

- (iii) Let  $\Phi(u(t)) = \Phi(Mf)$ .

If  $t > 0$ , then  $u(t)$  is continuous in  $\bar{\Omega}$ . By virtue of the strict convexity of  $r$ , we know that equality in the estimate (27) holds if and only if  $u(t) = Mf$  on  $\bar{\Omega}$ .

So let us turn to the case  $t = 0$ . From (i) and (ii), we conclude that  $\Phi(u(\theta)) = \Phi(Mf)$  for all  $\theta > 0$ . Thus, we have  $u(\theta) = Mf$  for all  $\theta > 0$ .

For  $\theta > 0$ , the Cauchy-Schwarz inequality gives

$$\int_{\Omega} |u(x, \theta) - \mu| dx \leq |\Omega|^{1/2} \|u(\theta) - Mf\|_{L^2(\Omega)} = 0.$$

Since  $u \in C([0, T]; L^2(\Omega))$ , the limit  $\theta \rightarrow 0^+$  finally yields  $u(0) = Mf$  a.e. on  $\Omega$ . Conversely, it is obvious that  $u(t) = Mf$  (a.e.) on  $\Omega$  implies  $\Phi(u(t)) = \Phi(Mf)$ .

(iv) Let  $t > 0$  and  $V'(t) = 0$ . Then from

$$0 = V'(t) = - \int_{\Omega} \underbrace{r''(u(x, t))}_{>0} \langle \nabla u(x, t), D(\nabla u_{\sigma}(x, t)) \nabla u(x, t) \rangle dx$$

we obtain

$$\langle \nabla u(x, t), D(\nabla u_{\sigma}(x, t)) \nabla u(x, t) \rangle = 0 \quad \text{a.e. on } \Omega.$$

By Lemma 2, there exists some constant  $\nu > 0$ , such that

$$\nu |\nabla u(x, t)|^2 \leq \langle \nabla u(x, t), D(\nabla u_{\sigma}(x, t)) \nabla u(x, t) \rangle \quad \text{on } \bar{\Omega} \times (0, \infty).$$

Thus, we have  $\nabla u(x, t) = 0$  a.e. on  $\Omega$ . Due to the continuity of  $\nabla u$ , this yields  $u(x, t) = \text{const.}$  for all  $x \in \Omega$ , and the average grey level invariance finally gives  $u(x, t) = \mu$  on  $\Omega$ .

Conversely, let  $u(x, t) = \mu$  on  $\Omega$ . Then,

$$V'(t) = - \int_{\Omega} r''(u(x, t)) \langle \nabla u(x, t), D(\nabla u_{\sigma}(x, t)) \nabla u(x, t) \rangle dx = 0.$$

(v) Suppose that  $V(T) = V(0)$ . Since  $V$  is decreasing, we have

$$V(t) = \text{const.} \quad \text{on } [0, T].$$

Let  $\epsilon > 0$ . Then for any  $t \in [\epsilon, T]$ , we have  $V'(t) = 0$ , and part (iv) implies that  $u(t) = Mf$  on  $\Omega$ . Now, the Cauchy-Schwarz inequality gives

$$\int_{\Omega} |f - Mf| dx \leq |\Omega|^{1/2} \|f - u(t)\|_{L^2(\Omega)}.$$

As  $u \in C([0, T]; L^2(\Omega))$ , the limit  $t \rightarrow 0^+$  yields  $f = Mf$  a.e. on  $\Omega$ .

Conversely, if  $u(t) = Mf$  (a.e.) on  $\Omega$  holds for all  $t \in [0, T]$ , it is evident that  $V(0) = V(T)$ .

(b) To proof the behaviour of  $u(t)$  for  $t \rightarrow \infty$ , we can adapt the methods of Illner and Neunzert [22] for directed linear diffusion to our (undirected) nonlinear case.

(i) Choosing  $r(s) = s^2$  and differentiating the corresponding Lyapunov function  $V(t)$  gives

$$V'(t) = -2 \int_{\Omega} \langle \nabla u, D(\nabla u_{\sigma}) \nabla u \rangle dx \leq 0. \quad (28)$$

From part (a) we know that  $\lim_{t \rightarrow \infty} V(t)$  exists. Therefore, the expression

$$\lim_{T \rightarrow \infty} V(T) - V(1) = -2 \int_1^{\infty} \int_{\Omega} \langle \nabla u, D(\nabla u_{\sigma}) \nabla u \rangle dx dt$$

is bounded. Hence, there exists a sequence  $(t_i) \rightarrow \infty$  with

$$\int_{\Omega} \langle \nabla u(x, t_i), D(\nabla u_{\sigma}(x, t_i)) \nabla u(x, t_i) \rangle dx \rightarrow 0 \quad \text{for } i \rightarrow \infty. \quad (29)$$

By Lemma 2, we can find some constant  $\nu > 0$  such that

$$\nu \|\nabla u(t_i)\|_{L^2(\Omega)}^2 \leq \int_{\Omega} \langle \nabla u(x, t_i), D(\nabla u_{\sigma}(x, t_i)) \nabla u(x, t_i) \rangle dx$$

and (29) gives

$$\|\nabla u(t_i)\|_{L^2(\Omega)} \rightarrow 0 \quad \text{for } i \rightarrow \infty. \quad (30)$$

Since  $V(t) = \int_{\Omega} u^2(x, t) dx$  is decreasing and bounded from below, we know that the set  $\{u(t_i) \mid i \in \mathbb{N}\}$  is bounded in the Hilbert space  $L^2(\Omega)$ . By the theorem of Banach–Alaoglu ([8], p.42) this bounded set is weakly compact, i.e. there exists a subsequence (which we denote for simplicity by  $(t_i)$ ) and some function  $g \in L^2(\Omega)$  such that

$$u(t_i) \rightharpoonup g \quad \text{weakly.}$$

By (30), we know that  $\{u(t_i) \mid i \in \mathbb{N}\} \in H^1(\Omega)$ . Since there exists a compact imbedding from  $H^1(\Omega)$  into  $L^2(\Omega)$  by the Rellich–Kondrachov theorem ([8], p. 169), we find a subsequence (again denoted by  $(t_i)$ ) such that

$$u(t_i) \rightarrow g \quad \text{strongly in } L^2(\Omega).$$

Hence, we get the convergence

$$\nabla u(t_i) \rightarrow \nabla g \quad \text{in the distributional sense.}$$

From (30) we obtain  $\nabla g = 0$  in the sense of distributions. Therefore,  $g$  is constant almost everywhere and

$$u(t_i) \rightarrow \text{const.} \quad \text{strongly in } L^2(\Omega).$$

The average grey level invariance implies

$$u(t_i) \rightarrow Mf \quad \text{strongly in } L^2(\Omega).$$

By the maximum principle, we know that

$$\|u(t) - Mf\|_{L^\infty(\Omega)} \leq \|f - Mf\|_{L^\infty(\Omega)} \quad \forall t > 0.$$

Now, let  $p \in \mathbb{N}$  with  $p \geq 2$ . For the sequence  $(t_i) \rightarrow \infty$ , we get

$$\|u(t_i) - Mf\|_{L^p(\Omega)}^p \leq \|f - Mf\|_{L^\infty(\Omega)}^{p-2} \cdot \|u(t_i) - Mf\|_{L^2(\Omega)}^2 \rightarrow 0. \quad (31)$$

For  $r(s) = |s - \mu|^p$ , we have  $r''(s) \geq 0$  and in (a) it was shown that the corresponding weak Lyapunov function  $V_p(t) = \|u(t) - Mf\|_{L^p(\Omega)}^p$  is decreasing and bounded from below. Hence,  $\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^p(\Omega)}$  exists and (31) implies

$$\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^p(\Omega)} = 0 \quad \text{for } 2 \leq p < \infty.$$

By the Cauchy-Schwarz inequality, we have

$$\|u(t) - Mf\|_{L^1(\Omega)} \leq |\Omega|^{1/2} \cdot \|u(t) - Mf\|_{L^2(\Omega)}$$

and therefore we end up with the assertion

$$\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^p(\Omega)} = 0 \quad \text{for } 1 \leq p < \infty.$$

- (ii) In the one-dimensional case with  $\Omega = (0, a)$ , we can find in analogy to (30) a sequence  $(t_i) \rightarrow \infty$  with

$$\|u_x(t_i)\|_{L^2(\Omega)} \rightarrow 0.$$

For  $x \in [0, a]$  the Cauchy-Schwarz inequality gives

$$|u(x, t_i) - u(0, t_i)| = \left| \int_0^x u_\xi(t_i) d\xi \right| \leq \sqrt{a} \cdot \|u_x(t_i)\|_{L^2(\Omega)} \rightarrow 0$$

So if  $\lim_{i \rightarrow \infty} u(x, t_i)$  or  $\lim_{i \rightarrow \infty} u(0, t_i)$  exists, the other limit must exist as well and both are equal. Furthermore, for every  $\epsilon > 0$ , there exists an  $i_0(\epsilon)$  such that

$$|u(x, t_i) - u(0, t_i)| < \epsilon \quad \forall i > i_0(\epsilon). \quad (32)$$

This means that  $u(x, t_i)$  is constant of order  $\epsilon$  with respect to  $x$ .

Therefore,  $\lim_{i \rightarrow \infty} u(x, t_i)$  must actually exist, because otherwise  $\lim_{t \rightarrow \infty} \int_0^a u^2(x, t) dx$  could not exist, which contradicts the Lyapunov properties of  $V(t) = \int_0^a u^2(x, t) dx$ . From this and the average grey level conservation we conclude that

$$\lim_{i \rightarrow \infty} u(x, t_i) = \mu \quad (33)$$

and due to (32), this convergence is uniform on  $\bar{\Omega}$ .

From part (a), we know that  $\|u(t) - Mf\|_{L^p(\Omega)}^p$  is a weak Lyapunov function for  $p \geq 2$ . Thus,

$$\|u(t) - Mf\|_{L^\infty(\Omega)} = \lim_{p \rightarrow \infty} \|u(t) - Mf\|_{L^p(\Omega)}$$

is also decreasing in  $t$ . Hence,  $\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^\infty(\Omega)}$  exists and (33) implies

$$\lim_{t \rightarrow \infty} \|u(t) - Mf\|_{L^\infty(\Omega)} = 0.$$

The smoothness of  $u$  establishes finally that

$$\lim_{t \rightarrow \infty} u(x, t) = \mu$$

uniformly on  $\bar{\Omega}$ . □

We have seen that for edge enhancing anisotropic diffusion we are faced with a competition between forward and backward diffusion. Theorem 5 is of importance, as it states that the regularization tames the backward diffusion in such a way that the forward diffusion wins in the long run. Moreover, the fight evolves the whole time in a certain direction: although backward diffusion may be locally superior, the global result – denoted by the Lyapunov functional – becomes permanently better for forward diffusion. So let us have a closer look at what might be the meaning of this global result in the frame of image processing.

Considering the Lyapunov functions associated with  $r(s) := |s|^p$ ,  $r(s) := (s - \mu)^{2n}$  and  $r(s) := s \ln s$ , respectively, the preceding theorem gives the following corollary.

**Corollary 1 (Special Lyapunov functionals).**

*Let  $u$  be the solution of (P) and  $a$  and  $\mu$  be defined as in (22) and (19). Then the following functions are decreasing for  $t \in [0, \infty)$ :*

- (a)  $\|u(t)\|_{L^p(\Omega)}$  for all  $p \geq 2$ .
- (b)  $M_{2n}[u(t)] := \frac{1}{|\Omega|} \int_{\Omega} (u(x, t) - \mu)^{2n} dx$  for all  $n \in \mathbb{N}$ .
- (c)  $H[u(t)] := \int_{\Omega} u(x, t) \ln(u(x, t)) dx$ , if  $a > 0$ .

Corollary 1 offers multiple possibilities of how to interpret nonlinear anisotropic diffusion filtering as a smoothing transformation.

As a special case of (a) it follows that the energy  $\|u(t)\|_{L^2(\Omega)}^2$  is reduced by diffusion.

Part (b) gives a probabilistic interpretation of anisotropic diffusion filtering. Consider the intensity in an image  $f$  as a random variable  $Z_f$  with distribution  $F_f(z)$ , i.e.  $F_f(z)$  is the probability, that an arbitrary grey value  $Z_f$  of  $f$  does not exceed  $z$ . By the average grey level invariance,  $\mu$  is equal to the expected value

$$EZ_{u(t)} := \int_{\mathbb{R}} z dF_{u(t)}(z),$$

and it follows that  $M_{2n}[u(t)]$  is just the even central moment

$$\int_{\mathbb{R}} (z - EZ_{u(t)})^{2n} dF_{u(t)}(z).$$

The second central moment (the variance) characterizes the spread of the intensity about its mean. It is a common tool for constructing measures for the relative smoothness of the intensity distribution. The fourth moment is frequently used to describe the relative flatness of the grey value distribution. Higher moments are more difficult to interpret, although they do provide important information for tasks like texture discrimination ([18], pp. 414–415). All decreasing even moments demonstrate that the image becomes smoother during diffusion filtering. Hence, local effects such as edge enhancement, which object to increase central moments, are overcompensated by a strong smoothing in other areas.

If we choose another probabilistic model of images, then part (c) characterizes the information-theoretical side of our scale-space. Suppose that the intensity of the image  $f$  is caused by photons being absorbed by some image recording device, and consider the location where the photon hits the image domain  $\bar{\Omega}$  as an  $m$ -dimensional random variable  $Z_f$ . If  $\int_{\Omega} f dx = 1$ , then  $f$  may be regarded as the density of this random variable. Then  $S[f] := -H[f]$  is nothing else but the entropy of  $f$ . An exhaustive treatment of entropy ideas may be found in [10]. Relevant for us is the fact that entropy is a measure of uncertainty and of missing information. This means that the scale-space created by anisotropic diffusion filtering embeds the genuine image  $f$  into a family of subsequently likelier versions of it which contain less information. Moreover, for  $t \rightarrow \infty$ , the process reaches the state with the lowest possible information, namely a constant image.

From these considerations, we observe that, in spite of its edge enhancing properties, anisotropic diffusion does really simplify the original image.

### 3.4 Examples

Figure 3 compares scale-spaces generated by isotropic linear diffusion, anisotropic nonlinear diffusion and the fundamental equation in image processing, respectively. The original image is taken from a magnetic resonance imaging (MRI) sequence of the human head.

It is observed that linear diffusion does not only blur all structures in an equal amount, but also dislocates them with increasing scale: for longer times, the head shape tends to a blurred disk. However, the average grey level is unaltered.

This is not true for the affine invariant scale-space: we observe that the image becomes more and more black as the time evolves. On the other hand, edges are not blurred as strong as in the linear diffusion case. This is due to the fact, that (7) smoothes anisotropically *along* level lines: indeed, it can be rewritten as

$$\partial_t u = |\nabla u|^{\frac{2}{3}} u^{\frac{1}{3}}_{\xi\xi}, \quad (34)$$

with  $\xi$  representing the tangent direction. Moreover, (7) implies that the fundamental equation propagates level lines in order to reduce their curvature: low-curved object boundaries are less affected by this process, while high-curved structures (e.g. the nose) exhibit roundings at an earlier stage. In Figure 3 we recognize that level lines shrink to convex curves and tend asymptotically to ellipses before they finally disappear. This is also confirmed by the theory, see [1] and the references therein. Thus, although edges are easier detectable than in the linear diffusion case, the correspondence problem remains: structures which are identified at coarse scales do not give the correct location and must be traced back. A modification of the affine invariant flow avoiding the shrinkage problem for the evolution of a *curve* is proposed by Sapiro and Tannenbaum [42]. Nevertheless, for *grey-scale images*, the practical use of this modification is yet under research. Since the fundamental equation of shape analysis involves no additional parameters and offers numerous invariances, it induces an ideal scale-space for uncommitted image analysis and shape recognition. However, its grey scale invariance prevents it from being contrast enhancing.

Nonlinear anisotropic diffusion filtering does offer this option, as it can be observed at the chin. Due to its drastically reduced diffusion in normal direction, edges are preserved and even enhanced for a fairly long time. Their localization is hardly affected until they finally disappear, and roundings at corners are less pronounced than in the affine invariant scale-space. Thus, the correspondence problem is no more of significance. The tendency to produce piecewise constant regions demonstrates that this scale-space is a well-suited preprocessing tool for segmentation. Unlike diffusion-reaction models aiming to produce *one* segmentation-like result for  $t \rightarrow \infty$ , the temporal evolution of our model generates a complete *hierarchical family* of segmentation-like images. From Figure 3 we recognize that most of the depicted “segments” coincide with the semantically correct objects. As for linear diffusion filtering, the average grey level remains constant. The contrast and scale parameters allow an adaptation of the anisotropic diffusion scale-space to the desired purpose, in order to preserve characteristic features for a longer time. In this sense, the time  $t$  is rather a parameter of importance, with respect to the specified task, than just a scale descriptor. Let us recall that this is exactly the basic idea of scale-spaces: to provide a family of subsequently simplified versions of the original image, in order to pick up the relevant information from a certain level and to get a hierarchy of structures. The fact that we use similar parameters for the very different images in Figure 1–3 demonstrates that, in practise, the parameter adaptation is not very critical.

Let us conclude this chapter with briefly discussing the difference between anisotropic diffusion and related morphological image-enhancement methods. Equation (34) indicates that anisotropic smoothing can also be obtained using morphological operators. A simpler example is given by the mean curvature flow (6), which can be written as

$$\partial_t u = u_{\xi\xi}$$

with  $\xi$  as above. Thus, one may construct a filter which diffuses linearly inside regions and smoothes by mean curvature motion along edges [2]. This approach also provides a contrast and a scale parameter, but it is not clear whether it permits edge enhancement. Anyway,

due to the linear diffusion, morphological grey scale invariance gets lost. A combination between mean curvature motion and hyperbolic shock filters is edge enhancing and gives experimentally nontrivial steady states [3]. It is intended as a pure restoration method without scale-space ambitions.

## 4 Generalizations

The results discussed above can be extended to a more general setting. The following three examples point out the directions, how this may be achieved.

### 4.1 Temporally varying kernel sizes

All previous considerations remain valid when allowing  $\sigma$  to be a smooth function in time, such that there exists some constant  $\sigma_0 > 0$  with  $\sigma(t) \geq \sigma_0$ . This includes the idea of Whitaker and Pizer [50] into our framework.

### 4.2 More sophisticated structure analysing methods

The edge detector  $\nabla u_\sigma$  enables us to adapt the diffusion to magnitude and direction of edges, but it cannot distinguish between edges and corners. Hence, anisotropic processes which smooth preferentially along the edge contour (like those in section 2.2), may give some undesired roundings at corners.

To identify corners, one can use second derivative descriptors like curvature or one may investigate fluctuations of the first derivatives within a certain neighbourhood. In order to illustrate the latter strategy, we may for instance average each component of the tensor product (16) by convolving it with a Gaussian. The resulting matrix

$$J_\rho(\nabla u_\sigma) := K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma) \quad (\rho > 0)$$

is called moment tensor, scatter matrix or structure tensor. It is a useful tool for texture description [39] and local structure analysis of spatio-temporal image sequences [23]. As an application to image restoration, Nitzberg and Shiota [31] used the quadratic form induced by the structure tensor to determine the shape of their anisotropic Gaussian kernel.

What makes the structure tensor so attractive for these applications? To understand this, recall that the symmetric matrix  $J_0(\nabla u_\sigma)$  has one eigenvector  $\nabla u_\sigma$  with corresponding eigenvalue  $|\nabla u_\sigma|^2$  and all eigenvalues belonging to other eigenvectors are 0. Hence, the eigenvalues give just the contrast (i.e. the squared gradient) in the eigendirections. By convolving  $J_0(\nabla u_\sigma)$  with a Gaussian  $K_\rho$ , the eigenvalues measure the variation of the grey values within a window size of order  $\rho$ . In the interior of a region, all contrasts nearly vanish and, thus, each eigenvalue is close to 0. At a hyperplane separating adjacent regions, we have  $m - 1$  orthogonal directions with no contrast variation and one direction (normal to the hyperplane) with large variation. Therefore, at (codimension 1) edges, one eigenvalue of the structure tensor is large and all the others are close to 0. Strongly curved structures possess more than one direction with large grey value variation within a certain neighbourhood. Thus, corner-like features result in a structure tensor with multiple nonvanishing eigenvalues.

In order to use the structure tensor for constructing a diffusion tensor, we can take (17) and replace  $g\left(\sqrt{J_0(\nabla u_\sigma)}\right)$  by  $g\left(\sqrt{J_\rho(\nabla u_\sigma)}\right)$ . Then the diffusion tensor possesses the same eigenvectors as  $\sqrt{J_\rho(\nabla u_\sigma)}$  and its eigenvalues are given by

$$\lambda_1 = g(\mu_1), \quad \dots, \quad \lambda_m = g(\mu_m),$$

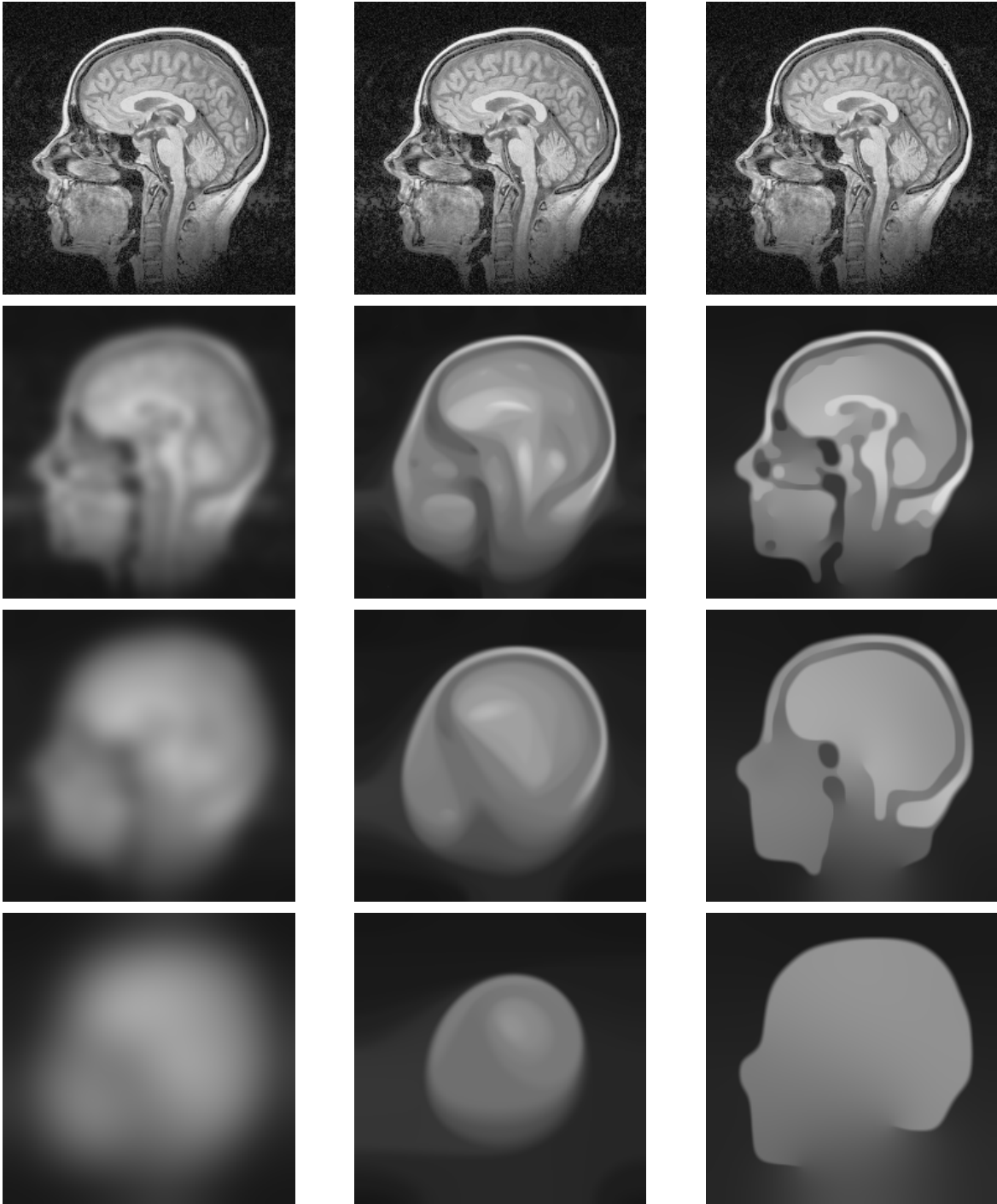


Figure 3: Comparison of scale-spaces. TOP: Original image,  $\Omega = (0, 235)^2$ . LEFT COLUMN: Linear diffusion, top to bottom:  $t = 0, 12.5, 50, 200$ . MIDDLE COLUMN: Fundamental equation,  $t = 0, 20, 50, 140$ . RIGHT COLUMN: Anisotropic diffusion,  $\lambda = 3, \sigma = 1, t = 0, 150, 800, 2400$ .



where  $\mu_1, \dots, \mu_m$  denote the eigenvalues of  $\sqrt{J_\rho(\nabla u_\sigma)}$ . Unlike at edges, diffusion at corners is now inhibited in more than one direction.

All our established results carry over when using the structure tensor instead of the edge detector  $\nabla u_\sigma$ .

### 4.3 Vector-valued images

Vector-valued images can arise either from devices measuring multiple physical properties or from a feature analysis of one single image. Examples for the first category are colour images, multi-spectral Landsat exposures and multi-spin echo MRI, whereas representatives of the second class are given by statistical moments or the jet space induced by the image itself and its partial derivatives up to a given order. Feature vectors play an important role for tasks like texture segmentation.

Isotropic vector-valued diffusion models were proposed by Gerig, Kübler, Kikinis, Jolesz [17] and Whitaker [51, 52] in the context of medical imagery. Our results extend their work to the anisotropic case by taking into account the direction of the structures.

The simplest idea how to apply diffusion filtering to multichannel images would be to diffuse all channels separately and independently from each other. This is frequently not recommendable since edges may be formed at different locations for the different channels. In order to avoid this, one should use a common diffusion tensor which combines information from all channels. One approach is to sum up the structure tensors of each channel to a common tensor. This may be regarded as collecting the contrast information of all channels. (Of course, the channels should be normalized in such a way that all values lie in a comparable range.)

Thus, for a multi-channel image  $(u_1, \dots, u_q)^T$ , the anisotropic diffusion equation is as follows:

$$\partial_t u_i = \operatorname{div} \left( g \left( \sqrt{\sum_{j=1}^q J_\rho(\nabla K_\sigma * u_j)} \right) \cdot \nabla u_i \right) \quad (i = 1, \dots, q).$$

It is evident that each channel of the diffusion process fulfils all properties of scalar-valued diffusion which we have investigated so far.

## 5 Conclusions

We have studied a class of regularized anisotropic nonlinear diffusion processes. Special representatives of them allow image enhancement by smoothing isotropically inside a region, while diffusing in an anisotropic way along edges. They can be applied in arbitrary dimensions, and the higher the dimension is, the stronger will the noise-eliminating capabilities at edges be, compared to the isotropic case.

The proposed class includes linear diffusion filtering and the nonlinear isotropic model of Catté, Lions, Morel, Coll [11] and Whitaker and Pizer [50]. It can be generalized to vector-valued images and may utilize more advanced structure descriptors than the gradient. For bounded initial images, there exists a unique smooth solution of the diffusion process which obeys an extremum principle and depends continuously on the initial data.

Due to their invariances and numerous information reducing properties, anisotropic diffusion filtering creates a scale-space representation of the original image. Unlike scale-spaces which can be uniquely classified (cf. (1), (6), (7)), it gives the user the liberty to incorporate a-priori demands concerning size and contrast of especially interesting features. Within the

scale-space hierarchy, these features are then rewarded by a stable location and an increased lifetime.

The theoretical results demonstrate that anisotropic diffusion is much more than an ad-hoc strategy for transforming a degraded image into a more pleasant looking one. It is a mathematically sound method which ties the advantages of two worlds: scale-space analysis and image restoration.

**Acknowledgement.** This work was supported by a grant from “Stiftung Rheinland-Pfalz für Innovation”. Thanks go to Andrea Bechtold and Jan Mohring for helpful hints.

## References

- [1] L. ALVAREZ, F. GUICHARD, P.-L. LIONS, AND J.-M. MOREL, *Axioms and fundamental equations in image processing*, Arch. Rational Mech. Anal., 123 (1993), pp. 199–257.
- [2] 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.
- [3] L. ALVAREZ AND L. MAZORRA, *Signal and image restoration using shock filters and anisotropic diffusion*, SIAM J. Numer. Anal., 31 (1994), pp. 590–605.
- [4] A. B. AREHART, L. VINCENT, AND B. B. KIMIA, *Mathematical morphology: The Hamilton–Jacobi connection*, in Proc. Fourth Int. Conf. on Computer Vision, IEEE Computer Society Press, Los Alamitos, pp. 215–219, 1993.
- [5] J. BABAUD, A. P. WITKIN, M. BAUDIN, AND R. O. DUDA, *Uniqueness of the Gaussian kernel for scale space filtering*, IEEE Trans. Pattern Anal. Mach. Intell., 8 (1986), pp. 26–33.
- [6] B. BENHAMOUDA, *Parameter adaptation for nonlinear diffusion in image processing*, master thesis, Dept. of Mathematics, University of Kaiserslautern, P.O. Box 3049, 67653 Kaiserslautern, Germany, 1994.
- [7] H. BREZIS, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North Holland, Amsterdam, 1973.
- [8] H. BREZIS, *Analyse fonctionnelle*, Masson, Paris, 1983.
- [9] R. W. BROCKETT AND P. MARAGOS, *Evolution equations for continuous-scale morphology*, in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 3 (1992), pp. 125–128.
- [10] B. BUCK AND V. MACAULAY, eds., *Maximum entropy in action*, Clarendon, Oxford, 1991.
- [11] F. CATTÉ, P.-L. LIONS, J.-M. MOREL, AND T. COLL, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., 29 (1992), pp. 182–193.
- [12] M.-H. CHEN AND P.-F. YAN, *A multiscaling approach based on morphological filtering*, IEEE Trans. Pattern Anal. Mach. Intell., 11 (1989), pp. 694–700.
- [13] G.-H. COTTET AND L. GERMAIN, *Image processing through reaction combined with nonlinear diffusion*, Math. Comp., 61 (1993), pp. 659–673.

- [14] G.-H. COTTET, *Neural networks: Continuous approach and applications to image processing*, technical report no. 113, LMC – IMAG, Université Joseph Fourier, B.P. 53, 38041 Grenoble Cédex 9, France, 1994 (submitted to Proc. 2nd Europ. Conf. on Mathematics applied to Biology and Medicine, Lyon, Dec. 15–18, 1993).
- [15] L. M. J. FLORACK, *The syntactical structure of scalar images*, Ph.D. thesis, Universiteit Utrecht, Faculteit Geneeskunde, Utrecht, 1993.
- [16] D. GEIGER AND A. YUILLE, *A common framework for image segmentation*, Int. J. of Computer Vision, 6 (1991), pp. 227–243.
- [17] G. GERIG, O. KÜBLER, R. KIKINIS, AND F. A. JOLESZ, *Nonlinear anisotropic filtering of MRI data*, IEEE Trans. Medical Imaging, 11 (1992), pp. 221–232.
- [18] R. C. GONZALEZ AND P. WINTZ, *Digital image processing*, Addison–Wesley, Reading, 1987.
- [19] B. M. TER HAAR ROMENY, L. M. J. FLORACK, J. J. KOENDERINK, AND M. A. VIERGEVER, *Scale space: Its natural operators and differential invariants*, in Information processing in medical imaging, A. C. F. Colchester and D. J. Hawkes, eds., Lecture Notes in Comp. Science 511, Springer, Berlin, pp. 239–255, 1991.
- [20] R. A. HUMMEL, *Representations based on zero-crossings in scale space*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 204–209, 1986.
- [21] R. ILLNER, *Global existence for two-velocity models of the Boltzmann equation*, Math. Meth. Appl. Sci., 1 (1979), pp. 187–193.
- [22] R. ILLNER AND H. NEUNZERT, *Relative entropy maximization and directed diffusion equations*, Math. Meth. Appl. Sci., 16 (1993), pp. 545–554.
- [23] B. JÄHNE, *Spatio-temporal image processing*, Lecture Notes in Comp. Science 751, Springer, Berlin, 1993.
- [24] B. B. KIMIA AND K. SIDDIQI, *Geometric heat equation and non-linear diffusion of shapes and images*, Technical Report LEMS-124, Division of Engineering, Brown University, Providence, R.I. 02912, U.S.A., 1993.
- [25] B. B. KIMIA, A. R. TANNENBAUM, AND S. W. ZUCKER, *Non-linear shape approximation via the entropy scale space*, in Geometric methods in computer vision II, B. C. Vemuri, ed., SPIE 2031, pp. 218–233, 1993.
- [26] J. J. KOENDERINK, *The structure of images*, Biol. Cybern., 50 (1984), pp. 363–370.
- [27] L. M. LIFSHITZ AND S. M. PIZER, *A multiresolution hierarchical approach to image segmentation based on intensity extrema*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 529–540.
- [28] T. LINDBERG, *Scale-space for discrete signals*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 234–254.
- [29] C. LOPEZ AND J.-M. MOREL, *Axiomatization of shape analysis and applications to texture hyperdiscrimination*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 646–647, 1993.

- [30] D. MARR AND E. HILDRETH, *Theory of edge detection*, Proc. R. Soc. Lond. B, 207 (1980), pp. 187–217.
- [31] M. NITZBERG AND T. SHIOTA, *Nonlinear image filtering with edge and corner enhancement*, IEEE Trans. Pattern Anal. Mach. Intell., 14 (1992), pp. 826–833.
- [32] N. NORDSTRÖM, *Biased anisotropic diffusion – a unified regularization and diffusion approach to edge detection*, Image and Vision Computing, 8 (1990), pp. 318–327.
- [33] P. J. OLVER, G. SAPIRO, AND A. TANNENBAUM, *Classification and uniqueness of invariant geometric flows*, C. R. Acad. Sci. Paris, Série I, 319 (1994), pp. 339–344.
- [34] S. OSHER AND L. RUDIN, *Feature-oriented image enhancement using shock filters*, SIAM J. Numer. Anal., 27 (1990), pp. 919–949.
- [35] P. PERONA AND J. MALIK, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639.
- [36] M. PROESMANS, E. J. PAUWELS, L. J. VAN GOOL, T. MOONS, AND A. OOSTERLINCK, *Image enhancement using non-linear diffusion*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 680–681, 1993.
- [37] M. H. PROTTER AND H. F. WEINBERGER, *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, 1978.
- [38] I. RAMBAUX AND P. GARÇON, *Nonlinear anisotropic diffusion filtering of 3D images*, project work, Département Génie Mathématique, INSA de Rouen and Laboratory of Technomathematics, University of Kaiserslautern, 1994.
- [39] A. R. RAO AND B. G. SCHUNCK, *Computing oriented texture fields*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 61–68, 1989.
- [40] A. H. SALDEN, *The nonlinear rescaling process*, report 3DCV 92-28, 3D Computer Vision, Academisch Ziekenhuis Utrecht, kamer E.02.222, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands, 1992.
- [41] G. SAPIRO AND A. TANNENBAUM, *Affine invariant scale-space*, Intern. J. of Computer Vision, 11 (1993), pp. 25–44.
- [42] G. SAPIRO AND A. TANNENBAUM, *Area and length preserving geometric invariant scale-spaces*, in Computer Vision – ECCV '94, J.-O. Eklundh, ed., Lecture Notes in Comp. Science 801, Springer, Berlin, pp. 449–458, 1994.
- [43] C. SCHNÖRR, *Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals*, J. Math. Imag. Vision, 4 (1994), pp. 189–198.
- [44] J. SHAH, *Parameter estimation, multiscale representation and algorithms for energy-minimizing segmentations*, report, Mathematics Dept., Northeastern Univ., 360 Huntington Av., Boston, MA 02115, U.S.A., 1990.
- [45] J. SHAH, *Segmentation by nonlinear diffusion*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 202–207, 1991.
- [46] J. SHAH, *Segmentation by nonlinear diffusion, II*, in Proc. IEEE Computer Society Conf. on Comp. Vision and Pattern Recognition, pp. 644–647, 1992.

- [47] J. WEICKERT, *Zwischenbericht zum Projekt "Nichtlineare Diffusionsfilter"*, in Bericht über die wissenschaftliche Tätigkeit Januar 1991 – Dezember 1991, Center for Applied Mathematics, Darmstadt – Kaiserslautern, pp. 133–142, 1992.
- [48] J. WEICKERT, *Abschlußbericht zum Projekt "Nichtlineare Diffusionsfilter"*, in Abschlußbericht und Bericht über die wissenschaftliche Tätigkeit Januar 1992 – Dezember 1993, Center for Applied Mathematics, Darmstadt – Kaiserslautern, pp. 191–209, 1994.
- [49] J. WEICKERT, *Anisotropic diffusion filters for image processing based quality control*, in Proc. Seventh European Conf. on Mathematics in Industry, A. Fasano and M. Primicerio, eds., Teubner, Stuttgart, pp. 355–362, 1994.
- [50] R. T. WHITAKER AND S. M. PIZER, *A multi-scale approach to nonuniform diffusion*, CVGIP: Image Understanding, 57 (1993), pp. 99–110.
- [51] R. T. WHITAKER, *Geometry limited diffusion in the characterization of geometric patches in images*, CVGIP: Image Understanding, 57 (1993), pp. 111–120.
- [52] R. T. WHITAKER, *Characterizing first and second-order patches using geometry-limited diffusion*, in Information processing in medical imaging, H. H. Barrett and A. F. Gmitro, eds., Lecture Notes in Comp. Science 687, Springer, Berlin, pp. 149–167, 1993.
- [53] A. P. WITKIN, *Scale-space filtering*, in Proc. Eighth Int. Joint Conf. on Artificial Intell., Karlsruhe, pp. 1019–1022, 1983.
- [54] A. L. YUILLE AND T. A. POGGIO, *Scaling theorems for zero crossings*, IEEE Trans. Pattern Anal. Mach. Intell., 8 (1986), pp. 15–25.