# Mathematical Imaging with Applications to MRI and Diffusion Tensor MRI

PhD Thesis

## Johan Lie

Department of Mathematics

University of Bergen

January 2007

# Preface

This dissertation is submitted as a partial fulfilment of the requirements for the degree Doctor of Philosophy (PhD) at the University of Bergen. The dissertation consists of two parts. Part I provides a theoretical and methodological background for the articles in Part II. Part II is a collection of articles covering various aspects of mathematical imaging, with main focus on methods based on partial differential equations. The scientific contribution of the thesis is collected in part II.

# Acknowledgements

I will first of all thank my wife Siv Øvrebotten Lie and my daughter Ingrid Øvrebotten Lie for all the support and patience throughout the three years of my PhD project. Thanks also to my mother Brit Elin Lie, my father Inge Johan Lie and my sister Heidi Lie for your constant support. You all mean a lot to me. Thanks also to Heidi for drawing some of the illustrations in Chapter 2.

I acknowledge support from the Norwegian Research Council under grant *BeMatA*.* Thanks to Xue-Cheng Tai and Arvid Lundervold for letting me be the PhD candidate on the DTI project which is part of BeMatA. My supervisors on the PhD programme has been Arvid Lundervold, Hans Z. Munthe-Kaas and Xue-Cheng Tai. I am grateful for all your help. I am deeply grateful to Jarle Berntsen and Stein Arild Strømme for encouraging me to go on with my PhD programme instead of quitting in a state of desperacy during the summer of 2005.

Arvid Lundervold, thank you for all your help and support throughout my PhD programme. You provide an important bridge between the mathematical department and the medical department at the University of Bergen.

Hans Z. Munthe-Kaas, thank you for being willing to be my main supervisor during the last 18 months of my PhD programme. Together with Arvid Lundervold you have provided perfect conditions for the work on the thesis.

Jan Nordbotten, I have learned a lot from you when we worked together on one of the papers in this thesis. You have (at least tried to) make me think like a mathematician. Thank you very much.

Marius Lysaker, I have also learned a lot from you. Thanks for all the help you provided by phone during the first year of my PhD study. I also got my first introduction to mathematical image processing by reading one of your early papers somewhat five years ago.

Thanks to Ørjan Bergmann and Oddvar Christiansen for all the hours we have

---

* BeMatA is an acronym for Computational Mathematics for Applications (*Be*regningsorientert *Mat*ematikk for *A*nvendelser ).

worked together. We make a good "diffusion tensor team" together, and I have certainly learned a lot from you. Thanks also to Ketil Oppedal for all the hours at the computer/coffee lab at BBB, and for making the best coffee at "campus".

Thanks to Erlend Hodneland for all the fruitful discussions during these three years, and before. We have not formally worked together on any project, but you have always been willing to help me.

Thanks to Stanley Osher for hosting Jan Nordbotten and me at University of California, Los Angeles for one week during the winter of 2005, and to Joachim Weickert for hosting me at Universität des Saarlandes for two weeks during the summer of 2006. The visit to Saarlandes was funded by the "L. Meltzer Høyksolefond".

Thanks to Tony F. Chan and Tin-Man Lee for hosting Oddvar Christiansen and me at University of California, Los Angeles for two weeks in November 2006.

Thanks to Renate Grüner for providing access to the 3T MR scanner at Haukeland University Hospital, Bergen, for one of the projects in this thesis.

Bergen, 3. January 2007

Johan Lie

# Contents

# Part I

# Background

$\mathfrak{V}$anklok mann
veit alt, han trur,
der mellom bergi han bur.
Men når framandfolk
honom finn,
rådlaus røynest han då.

*(Håvamål)*

# Introduction and Outline

Digital images are used extensively in our society, and have become a natural part of our life. In medicine, images of different modalities are used both in diagnostics and research. The area of medical imaging is in fact an important motive power for the development of the area of mathematical imaging. In our project, we have studied mathematical models for enhancement and analysis of images. Medical imaging, in particular magnetic resonance imaging (MRI) has been the main motivation behind the work in the project.

We start with a basic introduction to mathematical imaging. A *digital image* can be interpreted as a discrete function sampled on a mesh. This mesh is usually, but not restricted to, a uniform quadratic (in 2D) or cubic (in 3D) mesh where all the sides of the elements have the same size. In two dimensions the elements are called *pixels*, whereas in three dimensions the elements are called *voxels*. Mathematically we describe the image by

$$f(x_i) : \Omega \rightarrow Q, \tag{1}$$

where $\Omega$ is the domain over which the image is sampled, and $Q$ is a subset of the nonnegative integers. The function $f$ can be a scalar-, vector- or tensor- valued function. We will treat all these possibilities later in this thesis. Digital images are described as discrete functions. But since the size of the pixels/voxels usually is very small compared to the domain $\Omega$, we may look at $f$ as a discretization of a function $f : \Omega \rightarrow Q$ where $\Omega$ is a continuous domain. Therefore it is customary to view the function as a continuous mapping from the imaging domain to (a subset of) the real numbers $\mathcal{R}$

$$f : \Omega \rightarrow \mathcal{R}. \tag{2}$$

In the analysis of the mathematical models used in this thesis, we implicitly treat the image as a continuous function with a continuous domain. But in all numerical implementations, the domain of $f$ is discrete, and $f$ takes floating point values.

The images considered herein are constructed from one or several measurements. We might model the image formation by the model

$$f = Ku + \eta, \tag{3}$$

where $u$ is the true image, $K$ is a deterministic degradation operator (e.g. a convolution operator) and $\eta$ is indeterministic degradation (noise) [3]. For a generic image processing application, we cannot say anything about neither the deterministic degradation operator $K$ or the indeterministic degradation $\eta$. However, for a specific application it is often possible to say something about the error. The deterministic degradation operator can for example be a blurring operator. The indeterministic measurement error is usually modeled by a random distribution. In the papers in this thesis we assume that the deterministic degradation operator is neglectable. Then the degradation model becomes

$$f = u + \eta. \tag{4}$$

We assume that $\eta$ is normally distributed noise with zero mean. This is a simple model which might not fully reflect the real world. In this thesis we mainly process Magnetic Resonance (MR) images. In this case the noise is not normally distributed. However, we model (approximate) the noise as normally distributed.

The field of mathematical imaging is vast. We may distinguish between at least three different methodologies, namely *transform*, *statistical* and *variational* methods. We stress that there are no exact borders between these methodologies; a specific method can sometimes be expressed in terms of several different mathematical languages. In the major part of the research of this thesis (Papers A-D) we focus on variational (PDE) methods. In these papers we process the images in the spatial domain. But Paper E in is written in the context of transform and statistically based methods. In that paper we work in both the spatial and the frequency domain.

In chapter 1 we introduce the reader to a few PDE models which are related to image processing applications. This chapter is central to four of the five papers in part II in this thesis. In chapter 2 we give a brief introduction to the field of diffusion tensor magnetic resonance imaging (DTMRI), since this imaging modality has been a main motive power in the project. Two of the papers in the second part of this thesis are directly connected to DTMRI applications.

# Chapter 1

# PDE Image Processing

Partial Differential Equations (PDEs) has proved to be an extremely flexible and powerful concept in mathematical modeling [10, 25, 74]. The mathematical modeling of a vast number of physical processes rely on PDEs. During the last 20 years, Partial Differential Equation (PDE)s have been used extensively also in image processing applications [63, 66, 61, 65, 20]. Modeling of natural phenomena has been elegantly adapted to modeling in image processing. Perhaps the most well-known image processing task that is inspired by physical modelling of natural phenomena is noise removal. As we will see, we can remove noise in images by equations very similar to the heat equation, which is the prototype of a parabolic (evolutionary) PDE [25, 74]. Similarly, curves (edges) can be moved by applying equations similar to the transport equation, which is the prototype of a hyperbolic PDE. We will touch upon this kind of equations towards the end of this chapter [61]. In this chapter we review a few important PDE image processing models that are directly or indirectly related to the papers in the second part of this thesis. For the sake of clarity, we first consider models for scalar valued images, i.e. grayscale images. As we will see later, most of the methods can be generalized to vector (color) and matrix valued images as well.

The *linear heat equation* (sometimes called the *linear diffusion equation*) is probably the simplest PDE that is practical for image modeling. Recall the heat equation with a given initial condition $f(x)$ and Neumann boundary conditions

$$
\begin{aligned}
u_t(x,t) &= \nabla^2 u(x,t) \text{ in } \Omega \times (0, \infty) \,, \\
u(x,0) &= f(x), \\
u_n(x,t) &= 0 \text{ on } \partial\Omega \times (0, \infty) \,.
\end{aligned}
\tag{1.1}
$$

Here $u_t(x,t)$ denotes time derivative of $u(x,t)$, $u_n(x,t)$ denotes the normal derivative of $u(x,t)$ and $\nabla^2 u(x,t)$ denotes the Laplacian of $u(x,t)$. In physical modeling, the heat equation describes how heat spreads in a spatial domain as time

evolves. From an initial temperature distribution $u(x, 0) = f(x)$, the equation (1.1) gives us the temperature distribution $u(x, t)$ at any time $t > 0$. From intuition we know that an initial temperature distribution gets averaged out with time *.

The fact that the heat equation has an averaging effect can be used for image denoising purposes. Noise is a random signal with substantially higher frequency than the signal itself. We might hope that it is possible to pick an optimal stopping time $t^*$ such that the noise is averaged out, but the important features of the underlying image are still kept. By solving the heat equation, we can construct multiscale representations which are gradually smoother as time evolves. This is a powerful idea, especially when the linear heat equation is interchanged with more adaptive nonlinear equations. By appropriately modifying the heat equation, we can construct equations which preserves, and even enhance the most important information in images: namely edges. In the next section we give an introduction to scale space methods, which we interpret as solving a generalization of the heat equation. This approach is sometimes called diffusion filtering, even though the two terms *scale space modeling* and *diffusion filtering* does not always mean the same thing.

## 1.1   Scale Space Formulation

As mentioned in the previous section, multiscale representations of images can be constructed by solving a PDE of parabolic character, i.e. an evolutionary PDE. The image $f(x)$ is used as initial data and $u$ evolves towards a steady-state solution with increasing time. We may formally write such a PDE as

$$
\begin{aligned}
u_t(x, t) &= F(D^2 u, \nabla u, u, x, t), \text{ in } \Omega \times (0, \infty), \\
u(x, 0) &= f(x), \\
u_n(x, t) &= 0 \text{ on } \partial\Omega \times (0, \infty).
\end{aligned}
\tag{1.2}
$$

as in [20, 60]. Here $\nabla u$ is the spatial gradient of $u$, and $D^2 u$ is a compact notation of the set of all second derivatives. The term *scale space* comes from the interpretation of the time variable $t$ as a scale parameter. Initially, especially in the presence of noise in the data, details at all scales exists in the image. But as time evolves, small scales disappear, and only the large scales persist. Eventually, as $t \to \infty$, $u$ converges to a steady state which is determined by the boundary condition on $u$. Thus, we construct a hierarchical or multilevel representation of

---

* Diffusion is modeled by the same equation. The temperature distribution in the heat equation is in diffusion modeling interchanged with a density distribution.

the image data. This can be written as the embedding of $f$ into a family of images

$$\{T_t f\}, \qquad t \geq 0, \tag{1.3}$$

ordered according to a continuously increasing $t$. The operator $T_0$ is the identity operator. We construct the hierarchical representation (1.3) by solving a PDE on the form (1.2). The PDE can be solved by a time-stepping scheme where the time step $\Delta t$ is a fixed small number. This means that we compute approximations to the embedding (1.3) at time instants $t = 0, \Delta t, 2\Delta t, \cdots, N\Delta t$. Note that this resembles the steepest descent method from nonlinear optimization [58]. We illustrate the concept of scale spaces in Figure 1.1, where two scale space flows of very different nature are depicted. The scale space methods has proved to be powerful methods for many image processing purposes [1, 2, 13, 87].

Scale spaces can be derived in various ways. In the work by Alvarez et. al, scale spaces are constructed by an axiomatic approach [1]. They divide the axioms into three different categories, *architectural, stability* and *morphological* axioms. Some of the axioms are similar to principles that are well-known in the theory of PDEs, e.g. the max/min principle and the causality principle. By requiring the different axioms to hold, they deduce which PDE that must be solved.

In this thesis we follow another approach and instead interpret the scale spaces via diffusion filter methods. This means that we construct the hierarchical family of images by solving PDEs related to the linear heat equation. We will in the rest of this section study a diffusion equation which is more general than the heat equation (1.1), but slightly less general than the equation (1.2), namely

$$\begin{aligned}
u_t(x,t) &= \nabla \cdot (D(\nabla u, u, x)\nabla u), \text{ in } \Omega \times (0, \infty), \\
u(x,0) &= f(x), \\
u_n(x,t) &= 0, \text{ on } \partial\Omega \times (0, \infty).
\end{aligned} \tag{1.4}$$

where $D$ is a diffusivity coefficient which prevents diffusion across edges, and allows for diffusion in smooth regions of $u$. This diffusion coefficient can be either scalar- or matrix-valued.

Perona and Malik were pioneers in scale space image processing [63]. The Perona-Malik flow can be used for both edge detection purposes and image denoising purposes. In the Perona-Malik (PM) model, the governing PDE is given by

$$u_t = \nabla \cdot \left(g(|\nabla u|^2)\nabla u\right), \tag{1.5}$$

where $g$ is a smooth function with the properties $g(0^+) = 1$ and $g(\infty) = 0$, and $g' \leq 0$. One specific choice of $g$ is

$$g(|\nabla u|^2) = e^{-|\nabla u|^2/2\sigma^2}, \tag{1.6}$$
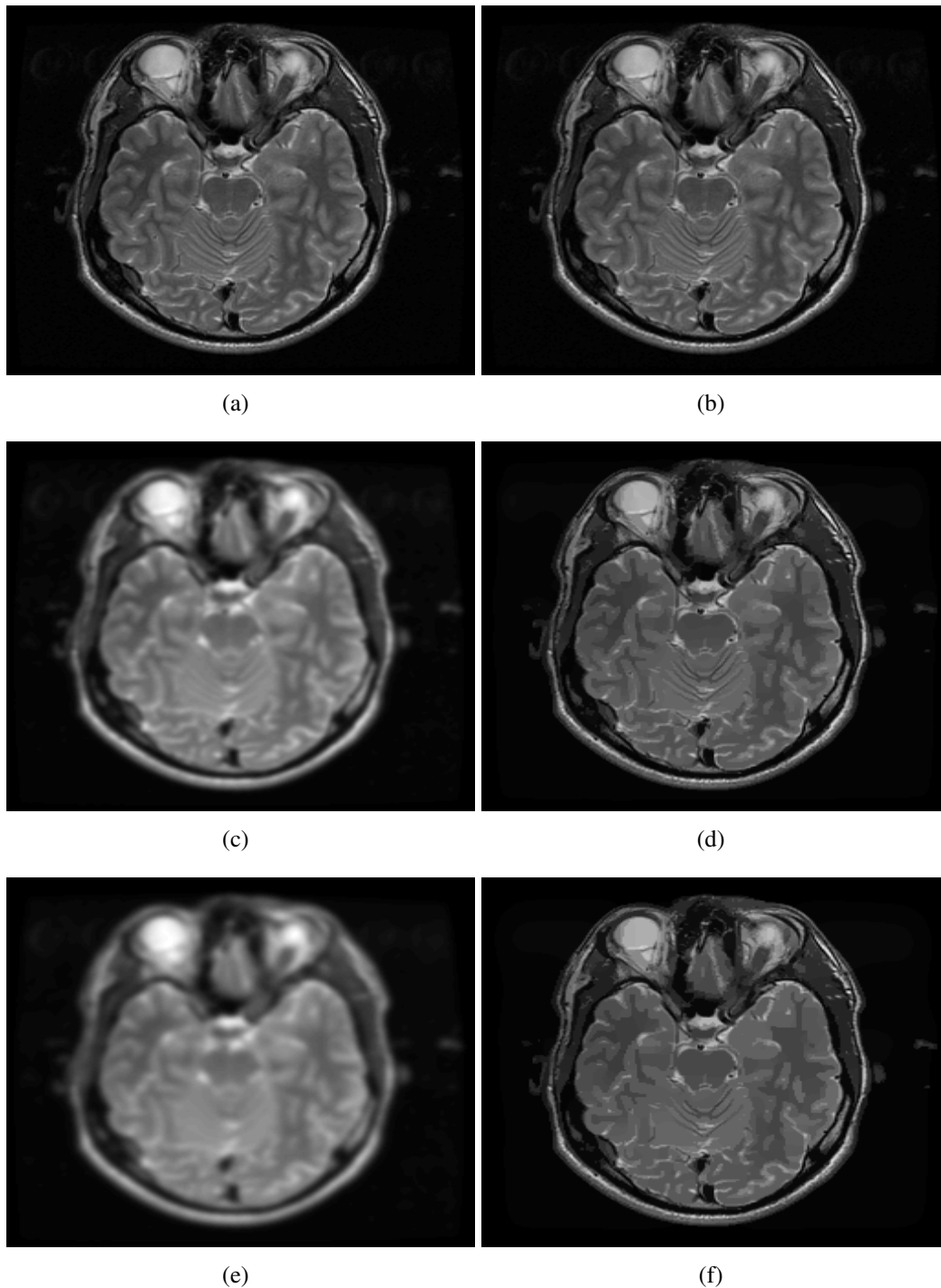
Figure 1.1: An illustration of two totally different scale spaces; Gaussian scale space in the left column and Perona-Malik scale space in the right column. Time is increasing from the upper to the lower rows. The small scale features are disappearing in both flows, but in totally different fashions. The image is one slice of a $T_2$ weighted MR image of the author.

where $\sigma$ is a scalar parameter. It is natural to interpret $\sigma$ as a noise-level parameter. The function $g(\cdot)$ acts like an edge detector. In regions where the gradient is steep, the diffusivity in equation (1.5) is small. Thus $g$ prevents (forward) diffusion across edges. In regions where the gradient is nonsteep the diffusivity becomes large. The Perona-Malik flow gives resulting images which are smoother than the input image, but the edges are well preserved and even *enhanced*. The flow can be better understood by a decomposition of the equation (1.5) [89]. We may write equation (1.5) on the form

$$u_t(x,t) = \left(g(|\nabla u|^2) + 2|\nabla u|^2 g'(|\nabla u|^2)\right) \partial_{\eta\eta} u \tag{1.7}$$

$$+ g(|\nabla u|^2) \left(\nabla^2 u - \partial_{\eta\eta} u\right), \tag{1.8}$$

where $\eta = (\nabla u/|\nabla u|)$ is the direction which is normal to the level curves of $u$. We now observe that in the case where $|\nabla u| \leq \lambda$, the equation is a traditional diffusion equation. But when $|\nabla u| > \lambda$, the factor in front of $u_{\eta\eta}$ becomes negative. This in turn implies that the Perona-Malik filter in fact works as a backward in time diffusion filter across large discontinuities in the data. This backward diffusion is however surprisingly stable [89].

Weickert and his coworkers are also pioneers in scale space imaging [88, 90]. In a series of papers, they have studied isotropic and anisotropic diffusion in image processing. Weickert also introduced the so-called structure tensor for image processing applications, where the diffusion coefficient $D$ is matrix valued [90, 88]. In their work, the scale space is constructed by solving anisotropic equations of the form

$$u_t = \nabla \cdot \left(D(\nabla u \nabla u^T)\nabla u\right) \tag{1.9}$$

In physics, a scalar valued diffusion is called isotropic, since it is not direction specific. Thus the PM diffusion is *isotropic*. A diffusion that depends on the direction is called *anisotropic*.

Scale space methods can be used directly for constructing multi scale representations of images, as shown for example in Figure 1.1. Then we are interested in the whole family $\{T_t f\}$, or at least a subset of it. But when scale space methods are used for image denoising purposes, then one member of the family is used as the solution to the denoising problem. Selecting a good stopping time is crucial, in order to pick out the best member $u(x, t^*)$ in the family of scales which is closest to a noise free image in a given norm. If we stop the flow at a time $t << t^*$, we get a result which still contains noise. On the contrary, if we stop the flow at a time $t >> t^*$, the resulting image is over-simplified. The optimal stopping time in scale space methods is strongly related to the noise level in the image. Thus, if a good estimate for the noise level is known, then a good estimate of the optimal stopping time can be found. Estimates of good stopping times for scale space models have been constructed by Mrázek and coworkers [49, 50].

To summarize, by scale space modeling we can construct a multi scale representation of images by solving a diffusion-like PDE. In the case where scale space methods are used for denoising purposes, a single member of the family of images must be picked out by a stopping criteria. Estimation of the optimal stopping time is in general difficult, and the stopping criteria often includes a free parameter.

The scale space methods have counterparts which are formulated as minimization problems. The minimization functionals are constructed in such a way that both *large deviations* from the initial data and *rapid oscillations* in the solution are penalized. We will in this thesis refer to these methods as variational or regularization methods. Some scale spaces corresponds to minimization problems, like the linear heat equation, but a scale space solution is not required to correspond to a minimum of a functional. This is the case for the Perona-Malik equation.

## 1.2 Regularization Formulation

A *functional* is a rule that assigns a real number to a function. Sometimes the solution of a PDE corresponds to the minimum of a energy-like functional. This minimum can be found by calculus of variations [28]. In four of the papers in the second part of this thesis, we solve minimization problems on the abstract form

$$\min_u \left\{ F(u, f) + \lambda R(u) \right\}, \tag{1.10}$$

where $R$ is a *regularization* functional which measures the smoothness of the solution $u$, and $F$ is a *fidelity* functional measuring the closeness of the solution $u$ to the input data $f$ in a given norm. The positive scalar $\lambda$ controls whether the solution should be close to the input data, or a smooth solution. We see that if $\lambda = 0$, the minimum of the functional in Equation (1.10) is the input data $f$. If $\lambda$ is very large, the solution depends almost exclusively on the regularization functional. In this case the solution will be (piecewise) smooth[†].

In a slightly different context, Tikhonov (1906-1993) developed a method to find reasonable solutions to ill-posed problems on the form

$$\min_u \| Ku - f \|^2. \tag{1.11}$$

Here $K$ is a ill-posed matrix, typically a discretization of a Fredholm integral equation of the first kind. The term ill-posed comes from the property that small

---

[†] Sometimes the regularization parameter $\lambda$ is written in front of $F(u, f)$. Then a small $\lambda$ gives a smooth solution, and a large $\lambda$ gives a solution which is close to the data. We use both formulations in the papers in the second part of this thesis.

perturbations in $f$ could lead to large perturbations in the solution. Thus a straightforward inversion of the operator $K$ gives results which by no means are meaningful. Tikhonov realized in the 1960's that by solving the modified problem (1.12)

$$\min_u \left\{ ||Ku - f||_2^2 + \lambda^2 ||Lu||_2^2 \right\},$$ (1.12)

with $L$ being a linear operator (e.g. $Lu = u$ or $Lu = \nabla u$), you get a well-posed solution as long as $\lambda$ is large enough. By requiring $||Lu||_2^2$ to be bounded, we "force" the solution to be smooth [80, 29].

Image deblurring is an example of a ill-posed problem where we can directly employ the framework of Tikhonov. But even in the case where the operator $K$ is not at all ill-posed (i.e. when $K = I$) we can use the methodology of Tikhonov to construct solutions which in some sense are regular. The notion of regularity however depends on the regularization operator. Typically the solution of a PDE image processing problem is the minimizer of a functional of the form

$$E(u) = \int_\Omega \left( (u - f)^2 + \lambda \Psi(|\nabla u|) \right) \mathrm{d}x.$$ (1.13)

Here the first term in the integral in Equation (1.13) is the fidelity integral, while the second term is the regularization integral. The properties of the minimizer of $E$ heavily depends on the choice of the regularization function $\Psi$. In section 1.2.2 we study a few specific choices of $\Psi$.

We will in this section introduce the framework that we need to find the minimum of a functional. We mainly follow the exposition in the book [28] and lecture notes from lectures held by Joachim Weickert. The functionals that we want to minimize are of the form

$$E(u) = \int_\Omega F(x, u, \nabla u) \mathrm{d}x.$$ (1.14)

However, to introduce the concepts of calculus of variations, we first look at the functional

$$E(u) = \int_a^b F(x, u, u') \mathrm{d}x.$$ (1.15)

We want to find an expression for the function $u$ that minimize the functional of Equation (1.15). In standard calculus, a necessary criterium for $x^*$ to be a minimum point of the function $f$ is that $f'(x^*) = 0$. In the calculus of variations the corresponding criterium is a differential equation. This equation is called the Euler-Lagrange equation which is attributed to Euler (1707-1783) and Lagrange

(1736-1813). We will now see that integration by parts and other standard methods from calculus are sufficient to derive the Euler-Lagrange equation corresponding to the minimum of the functional in Equation (1.15).

Let us assume that $v(x)$ is a minimizer of the functional (1.15), that is

$$E(v) < E(w) \quad \forall\, w. \tag{1.16}$$

We might perturb the function $v(x)$ with another function $h(x)$

$$u(x, \epsilon) = v(x) + \epsilon h(x). \tag{1.17}$$

Since the function $v(x)$ is a minimum of $E$, we know that

$$g(\epsilon) = E(u(x, \epsilon)) \tag{1.18}$$

has a minimum at $\epsilon = 0$. This means that the derivative $g'(\epsilon)$ of this function must vanish at $\epsilon = 0$. Thus we must have

$$0 = g'(0) = \frac{d}{d\epsilon} E(v + \epsilon h)|_{\epsilon=0} = \frac{d}{d\epsilon} \int_a^b F(x, v + \epsilon h, v' + \epsilon h') dx|_{\epsilon=0} \tag{1.19}$$

Using the chain rule, we see that the derivative (w.r.t $\epsilon$) of the last term of Equation (1.19) becomes

$$0 = \int_a^b [F_u h + F_{u'} h'] dx. \tag{1.20}$$

Using integration by parts, we get

$$0 = \int_a^b [F_u h - \frac{d}{dx} F_{u'} h] dx + F_{u'} h|_a^b \tag{1.21}$$

Since $v$ is the minimizer of the functional from Equation (1.15), it must be the minimizer for any perturbation $h$. We can in particular choose $h$ in such a way that $h(a) = h(b) = 0$. A more general and natural boundary condition is however

$$F_{u'}|_a^b = 0. \tag{1.22}$$

This boundary condition is called the *natural* boundary condition. Thus the Equation (1.21) gives us what is called the Euler-Lagrange equation corresponding to the functional of Equation (1.15).

$$F_u - \frac{d}{dx} F_{u'} = 0. \tag{1.23}$$

The Euler-Lagrange equation corresponding to Equation (1.14) for $d$ spatial dimensions is given by

$$F_u - \sum_{i=1}^{d} \partial_{x_i} F_{u_{x_i}} = 0, \tag{1.24}$$

with the natural boundary condition

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial \Omega. \tag{1.25}$$

In the remaining part of this section, we use the shorthand notation

$$E'(u) := F_u - \sum_{i=1}^{d} \partial_{x_i} F_{u_{x_i}} \tag{1.26}$$

for the Euler-Lagrange equation. At a minimum $v$, we have

$$E'(v) = 0, \tag{1.27}$$

thus we mimic the notation from standard calculus.

In Papers A-D in this thesis the solution strategy is to solve a parabolic PDE, which at steady-state corresponds to $E'(v) = 0$. Note that this strategy resembles the scale space formulation with an additional fidelity term. Thus it is clear that the regularization formulation and the scale space formulation are closely related. This connection has been thoroughly studied by Weickert and Scherzer [70].

## 1.2.1  Solution Strategy

We use the *method of steepest descent* to find a solution $u$ that satisfies the Euler-Lagrange equation corresponding to an energy functional. This methodology fits well with the diffusion formulation from the previous chapter.

The method of steepest descent is probably the simplest method in nonlinear minimization [58]. From a starting point $u_0$, we construct a sequence of solutions $u_k$ from the equation

$$u_{k+1} = u_k - \alpha g_k, \tag{1.28}$$

where $\alpha$ is a small stepsize parameter, and $g$ is the derivative of the minimization functional. If we interpret the parameter $\alpha$ as a time step and reorder terms in Equation (1.28), we see that we have the equation

$$\frac{u_{k+1} - u_k}{\alpha} = -g_k, \tag{1.29}$$

which is nothing but a discretization of the equation

$$u_t = -g. \tag{1.30}$$

If we interpret the derivative $g$ as

$$g = E'(u), \tag{1.31}$$

we see that if we are able to reach a steady-state to the equation (1.30) then $u$ is a solution to the Euler-Lagrange equation corresponding to the minimization functional $E$. Thus we construct the solution to the minimization problem by solving the equation (1.30) to steady-state.

We stress that the steepest descent methods is inefficient w.r.t computation time. The convergence of the steepest descent is only of first order. On the other hand, the steepest descent method is robust in the sense that it leads to simple numerical schemes that are easy to implement.

There are many strategies for speeding up the convergence, and perhaps the most intuitive (traditional) way is by employing Newton methods or quasi-Newton methods [58, 7].These methods are known to have quadratic convergence behaviour, at least in a neighborhood of the solution. Alternatively, multigrid methods might be employed [84, 18]. Chambolle has developed quick algorithms for solving regularization problems based on a duality formulation, where the solution can be seen as a projection of the data to a convex set [15]. Recently, Darbon and Sigelle have developed even quicker (and even more exotic) methods for the solution of a class of regularization problems based on graph theory [24]. So there are many ways to accelerate the solution process. However, in this thesis we do not focus on quick algorithms, we rather study the underlying properties of the mathematical models. Therefore we employ simple finite difference steepest descent schemes.

## 1.2.2   Common Regularization Functionals

In scale space modeling, the resulting image depends heavily on the properties of the underlying diffusion equation, as we have seen for example in Figure (1.1). In the following we will introduce the reader to a few regularization functionals that are commonly used for image denoising purposes.

**Laplacian with $L_2$ fidelity**

In regularization modeling, an analogue to the heat equation is the Laplacian regularization functional with the $L_2$ fidelity functional

$$G(u, f) = \int_\Omega |\nabla u|^2 \, dx + \frac{\lambda}{2} \int_\Omega (u - f)^2 \, dx. \tag{1.32}$$

By differentiation, we see that the Euler-Lagrange equation corresponding to the functional (1.32) is

$$G'(u) = -\nabla^2 u + \lambda(u - f). \tag{1.33}$$

At a minimum $v$ of the functional we have

$$G'(v) = 0. \tag{1.34}$$

By the gradient descent method, the minimum $v$ corresponds to the steady-state solution to the problem

$$
\begin{aligned}
u_t(x, t) &= \nabla^2 u(x, t) - \lambda(u(x, t) - f(x)), \\
u(x, 0) &= f(x), \\
u_n(x, t) &= 0.
\end{aligned}
\tag{1.35}
$$

The steady state solution to Equation (1.35) with a varying regularization parameter $\lambda$ gives solutions of similar character as the Gaussian scale space with varying stopping time $t$. But this means that using the regularization functional (1.32), we cannot expect to keep the edge information in the resulting image. This follows naturally from the properties of the Laplacian operator $\nabla^2 u$. It is a well-known fact that the solution of the linear heat equation is smooth for all $t > 0$ even when the initial condition $u(x, 0)$ is a nonsmooth function [25, 74]. More specifically, the solution to the equation (1.35) belongs to the Sobolev space $W^{1,2}(\Omega)$ which is defined by

$$W^{1,2}(\Omega) = \left\{ u \,:\, u \in L_2(\Omega),\, \nabla u \in L_2(\Omega)^d \right\}, \tag{1.36}$$

where $d$ is the dimension of the domain $\Omega$ [3]. The additional fidelity term does not change this fundamental property of the solution. To allow for edges in the solution, the regularization functional must be modified, i.e. we need to search for the solution in another function space which allows for discontinuities. To do this we replace the Diriclet functional with the total variation functional, and instead of searching for a continuous solution we search for a solution which is in the space of bounded variations.

## Total variation with $L_2$ fidelity

Among the most successful image denoising models is the total variation model that was first used for image processing by Rudin, Osher and Fatemi (ROF) in [65]. The idea of this model is to use the Total Variation (TV) functional

$$R(u) = \int_\Omega |\nabla u| \, \mathrm{d}x \tag{1.37}$$

as a regularization functional instead of the Diriclet integral. Adding the $L_2$ fidelity term, the ROF functional becomes

$$G(u, f) = \int |\nabla u| \mathrm{d}x + \frac{\lambda}{2} \int (u - f)^2 \mathrm{d}x. \tag{1.38}$$

Despite the minor difference in notation compared to the functional (1.32), the modification of the regularization term in the functional has a major impact on the solution to the denoising problem. The reason is that the TV functional penalizes jumps in the solution in a completely different way than the Diriclet functional.

We have already formally introduced the notion of the TV functional. For a one dimensional signal $u$, the total variation functional is rigorously defined as

$$TV(u) = \sup_{h > 0} \int \left| \frac{u(x + h) - u(x)}{h} \right| \mathrm{d}x. \tag{1.39}$$

However, in practice the TV functional of a one dimensional function is treated as

$$TV(u) = \int |u_x| \mathrm{d}x. \tag{1.40}$$

This not so rigorous definition can be used for higher dimensions, such that we get the definition of the multi dimensional TV functional as in (1.37). More rigorously we can define the total variation of a function $u$ by

$$\int_\Omega |\nabla u| = \sup \left\{ \int_\Omega u \nabla \cdot g \mathrm{d}x \; : \; g \in C_c^1(\Omega, R^n), |g(x)| \leq 1 \forall x \in \Omega \right\}, \tag{1.41}$$

where $C_c^1$ denotes the set of at least one time differentiable functions with compact support.

A nice thing about TV$(u)$ is that it is finite for any function that is bounded, even for discontinuous functions [60]. As long as a function $u$ does not have infinitely large jumps, then TV$(u)$ is finite. In image processing applications this property of the TV functional allows a regularized image to be a discontinuous function. This is in strong contrast to the Diriclet functional (i.e. $p = 2$), which is unbounded for all discontinuous functions. This important property of the $TV$ functional in fact only holds for $p = 1$ for functionals on the form

$$R_p(u) = \int |\nabla u|^p \mathrm{d}x. \tag{1.42}$$

As noted by Chan and others, we may interpret the TV norm of a function as integration along all level sets of the function. This implies that the TV norm depends on both the size of jumps and geometry of the level sets [17]. Since

edges as well as geometry are among the most prominent features for human perception, functions with finite total variation are attractive building blocks for imaging purposes, rather than continuous functions.

By differentiation, we find the Euler-Lagrange equation corresponding to Equation (1.38)

$$G'(u, f) = -\nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) + (u - f). \tag{1.43}$$

Again, following the standard gradient descent approach, we find the zero point of the Euler-Lagrange equation by solving the equation

$$u_t = \nabla \cdot \left( \frac{\nabla u}{|\nabla u|} \right) - (u - f) \tag{1.44}$$

$$u(x, 0) = f(x) \tag{1.45}$$

$$u_n(x, t) = 0 \tag{1.46}$$

to steady state. The minimizer of the ROF problem is in the space of bounded variations

$$\mathrm{BV}(\Omega) = \left\{ u \in L^1 \,|\, \int_\Omega |\nabla u| < \infty \right\} \tag{1.47}$$

Even though the ROF model for denoising purposes is superior to the heat equation with the $L_2$ fidelity term, the method has a few drawbacks, namely *the staircase effect* and loss of contrast and geometry. Additionally, textures are not well handled with the original ROF model [3, 93].

In paper (A)-(D) in this thesis, we have used the total variation regularizer for various applications. We consider this regularizer as a "standard" or "generic" regularizer. It is however clear that the total variation regularizer can be interchanged with other regularizers in some applications. Regarding the fidelity functional, we have only used the $L_2$ norm as a measure of closeness. This is a very natural norm to use for many applications. However, for some applications the $L_1$ norm can give better results [19].

The result from regularization by functional minimization depends on the properties of the minimization functional (1.10). Since the equations are solved to steady state, we do not have a stopping criteria in the regularization formulation. However, the regularization parameter $\lambda$ has a similar role as the stopping time $t$ in the scale space formulation. So the two approaches are closely related. By selection of $\lambda$, the minimum of the functional (1.10) is a compromise between a completely smooth solution and a solution which is close to the initial data $f$. Correspondingly, if we choose a small stopping time $t$ in the scale space formulation, we get a solution which is close to the initial data. A large stopping time

corresponds to a very smooth solution. Thus we see that a parameter selection is important in both scale space methods and the variational methods. In the scale space setting, we have to pick out a good stopping time $t$, while in the variational setting we have to choose a proper fidelity parameter $\lambda$.

## PDE Regularization of Matrix Valued Images

So far in this thesis, we have considered methods for scalar valued images, i.e. grayscale images. Blomgren and Chan among others have developed total variation based methods for vector valued images[68, 9] . When generalizing methods for grayscale images to color images, the simplest approach is to employ scalar methods to each channel independently. Blomgren and Chan solve a *coupled* set of three PDEs for regularization of color images via TV flow [9]. They show that this is a better approach than solving a set of three uncoupled PDEs. The reason is that a coupling of the channels is needed if the PDE flow is supposed to preserve edge information in the regularized color image. They observe that if the ROF model is used on each channel individually, i.e.

$$\frac{\partial u_i}{\partial t} = \nabla \cdot \left( \frac{\nabla u_i}{|\nabla u_i|} \right) - \lambda (u_i - f_i) \qquad i = 1, 2, 3, \tag{1.48}$$

we get nonappealing results. This is because the three image channels are regularized in the same fashion, using the same $\lambda$ in each channel, even if the intensity of the different channels could differ substantially. To overcome this problem they incorporate information from all channels into the flow of each individual channel by introducing the weighting factor

$$w_j = \frac{\text{TV}[u_j]}{\sqrt{\sum_{i=1}^{3} \left( \text{TV}[u_i] \right)^2}}, \qquad j = 1, 2, 3, \tag{1.49}$$

and solve the set of modified equations

$$\frac{\partial u_j}{\partial t} = w_j \nabla \cdot \left( \frac{\nabla u_j}{|\nabla u_j|} \right) - \lambda (u_j - f_j), \qquad \text{in } \Omega \times (0, \infty)$$

$$u_j = f_j \qquad \text{in } \Omega \times \{0\}$$

$$\frac{\partial u_j}{\partial n} = 0 \qquad \text{on } \partial \Omega \times (0, \infty),$$

for $j = 1, 2, 3$. The steadystate solution to this system of equations corresponds to the minimization problem

$$\min_u \left\{ \sqrt{\sum_{j=1}^{3} \text{TV}[u_j]^2} + \frac{\lambda}{2} \sum_{j=1}^{3} ||u_j - f_j||_2^2 \right\}. \tag{1.50}$$

As noted by Blomgren and Chan, the weighting term $w_j$ in each channel $u_j$ is equivalent to the mapping

$$\lambda_i \leftarrow \lambda \frac{\sqrt{\sum_{i=1}^{3} (\text{TV}[u_i])^2}}{\text{TV}[u_i]}. \tag{1.51}$$

Using the model Blomgren-Chan model, a channel with a high total variation norm is stronger regularized than a channel with low total variation norm, which is quite intuitively correct. In paper D we make a natural extension of the Blomgren Chan model to yield a regularization method for matrix valued images. Our extension has two essential ingredients; the information flow between the different imaging channels is ensured by the Blomgren-Chan coupling and the positive definiteness of the diffusion matrix $D$ is ensured via the implicit representation $D = LL^T$, where $L$ is a lower triangular matrix.

Recall that the Frobenius norm of a matrix $A \in R^{M \times N}$ is defined by

$$\|A\|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} |a_{ij}|^2}. \tag{1.52}$$

Inspired by the Blomgren Chan model, we use this norm as a model for a corresponding norm in the total variation setting by defining the total variation norm of a $M \times M$ matrix valued function $D$ as

$$\text{TV}(D) = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{M} (\text{TV}[d_{ij}])^2}. \tag{1.53}$$

Using this definition, we formulate a minimization problem for matrix valued functions that is equivalent to the vector valued minimization problem of Blomgren and Chan

$$\min_{d_{ij}} \left\{ \text{TV}(D) + \frac{\lambda}{2} \sum_{ij} \|d_{ij} - f_{ij}\|_2^2 \right\} \tag{1.54}$$

We see that this formulation reduces to the Blomgren Chan model if $D$ is interchanged with a $3 \times 1$ vector function and to the ROF model if $D$ is interchanged with a scalar function. From a physical modelling viewpoint, the matrix $D$ is required to be symmetric and positive definite, at least if the model is supposed to be used for diffusion tensor regularization. We go into details on diffusion tensor images later in this text. To ensure positive definiteness and also symmetry, we have chosen to represent the tensor implicitly on the form of a Cholesky factorization

$$D = LL^T. \tag{1.55}$$

Instead of working directly with the tensor elements $D_{ij}$, we work with the factors $L_{ij}$. The Cholesky factorization ensures that the matrix $D$ is both symmetric and positive definite. Even though we work implicitly on the factors $L_{ij}$, we minimize the functional of equation (3.15), which depends on the full matrix $D$, not only the factor $L$. This fully justifies the term *regularization of matrix valued data*. This is in contrast to the work of Wang and coauthors [86]. Instead of regularizing the matrix $D$ they regularize the Cholesky factor $L$. It is more natural to regularize the whole matrix $D$. However, when we introduce the $LL^T$ factorization, we cannot prove that the functional is convex, or even quasi-convex. But from numerical experiments, we observe that our formulation for all practical purposes converges to the same minimum with different (in fact completely random) initial conditions. Thus it seems like our formulation is at least quasi-convex. Another important difference between our proposed model and the model by Wang and coauthors is that we solve a coupled set of PDEs, while they solve for each channel individually. In Paper D we show by numerical experiments that this in fact makes a difference.

## 1.3   Inverse Scale Space Formulation

During the last couple of years, attention has been drawn to inverse scale space methods, which might be interpreted as a hybrid between scale space methods and regularization methods. Consider the scale space flow depicted in the diagram in Figure 1.1. Initially the data $f$ contains information at all scales, and $u$ becomes increasingly regular with time. If no stopping criteria is imposed on the solution, $u$ will converge to mean($f$) at every point in the imaging domain. Thus in scale space modelling, information is lost as time is increased. This was illustrated for two different scale spaces in Figure 1.1. In inverse scale space modeling the flow is in the opposite direction, as shown in Figure 1.2.

From the initial condition $u = $ mean($f$), information of decreasing scale is added to $u$ as time evolves. Thus in inverse scale space modeling, information is gained as time evolves. The inverse scale space methods have not been as thoroughly studied as forward scale space methods. The first attempts to construct inverse scale space methods for image processing purposes was done by Scherzer and Groetch in 2001 [69]. In their work, they show similarities between scale space modeling and regularization of ill-posed inverse problems. The methodology of inverse scale space methods is related to regularization methods for ill-posed inverse problems, in particular to Tikhonov regularization [80, 30, 85]. For a detailed description of the field of regularization of inverse problems, we refer the reader to the work of Per Christian Hansen[30], and the book of Curtis R. Vogel [85].

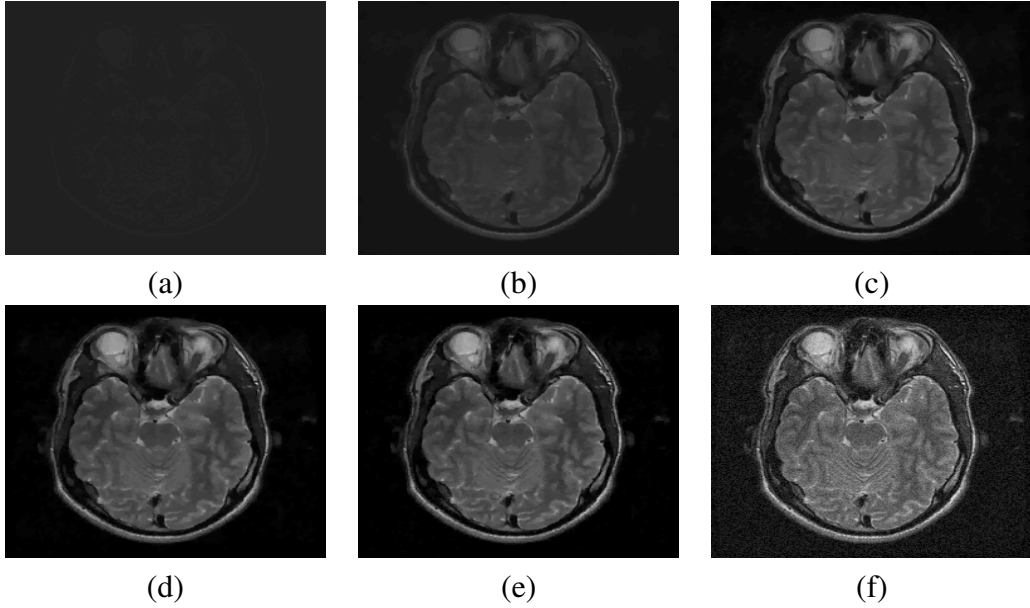In 2003, Tasdizen and coauthors introduced a novel techique for image

<div align="center">(a)            (b)            (c)</div>

<div align="center">(d)            (e)            (f)</div>

Figure 1.2: We here show how the relaxed inverse scale space flow forms a series of coarse to fine approximations (a)-(e) to the noisy input image $f$ (lower right corner).

processing, where they smooth the field of the unit normals of the image. Afterwards they construct new surface based on the smoothed normals [78]. This idea was further developed by Lysaker, Osher and Tai [44]. Osher and coauthors maturized the ideas from these two papers. They developed the so-called *iterative refinement technique* for image regularization [59]. This is closely related to inverse scale space methods. Burger, Osher and others studied inverse scale space methods in the papers [12]. They study the flow given by

$$\partial_t p = -\partial_u F(f, u) \quad p \in \partial_u R(u), \tag{1.56}$$

where $F$ is a fidelity functional, $R$ is a regularization functional and $\partial_u R(u)$ denotes an element of the subgradient of $R$,

$$\partial_u R(u) = \{p \in BV(\Omega)^* | R(v) \geq R(u) + \langle p, v - u \rangle, \forall v \in BV(\Omega) , \tag{1.57}$$

and $\partial_u F(f, u)$ denotes an element of the subgradient of $F$

$$\partial_u F(u) = \left\{ q \in L^2(\Omega) | F(v, f) \geq F(u, f) + \langle q, v - u \rangle, \forall v \in L^2(\Omega) . \tag{1.58}$$

Here we have used the notation $BV(\Omega)^*$ to denote the dual space of $BV(\Omega)$, and $\langle \cdot, \cdot \rangle$ denotes the $L_2$ innerproduct. Note that for differentiable operators the subgradient becomes the usual derivative.

However, since this equation is a flow in $p$, and the relation between $p$ and $u$ is not explicitly given, it is not straight forward to solve this equation in practice. In Paper C, we study properties of a numerical tractable implementation of a PDE inverse scale space method, which approximates the inverse scale space flow. This flow was proposed by Burger et.al. They proved that the flow is convergent for linear regularization operators [11]. In our work, we follow this up and prove that the relaxed flow is convergent for a class of nonlinear convex regularization operators that are used in practice.

## 1.4   Level Set Formulation

In 1987 Osher and Sethian introduced a new method for interface evolvement in hyperbolic PDEs [61]. The method they invented is called the *level set method*, and it has revolutionized the way interfaces are handled in image processing and computer vision applications. Instead of representing a curve $\Gamma$ by an explicit parametrization, they embed the interface $\Gamma$ into the zero-level of a continuous higher dimensional function $\phi$ and thus made a very flexible framework for interface evolution. The concept of level set modelling has since the introduction in 1987 become an important tool in many branches of applied mathematics and computer science [71, 60]. The method is in particular widely used in imaging and vision science, see for example the book by Osher and Fedkiw, the book by Sethian or the book by Malladi and others [60, 71, 45].

In a series of papers, two of which appear in this thesis as Papers A and B, we have developed and analyzed new methods related to the level set method [40, 41, 42, 43]. We note that even though we refer to these methods as level set methods, we could equally well have called them labeling methods, as was mentioned by a reviewer of one of our papers. The labeling is however variational in the sense that we control the regularity of the resulting functions. Before we go into details on this model, we give an introduction to standard level set modeling, in particular level set image segmentation. Image segmentation is the task of dividing the set of pixels of the image into a set of classes or domains, where each distinct class corresponds to a specific property in the image. Thus we might consider segmentation as a higher level task than noise removal. We have considered an image segmentation model where the classes are based on the intensity values in the image. In our image segmentation algorithms, the aim is to label each pixel in the imaging domain, and identify the interfaces between the different classes. This is however a bit in contrast to shape driven segmentation where shapes are more prominent features than intensity values.

The level set method can be used to implicitly represent curves $\Gamma$ separating the image into a set of distinct regions [16, 83, 67, 3]. Instead of working with a
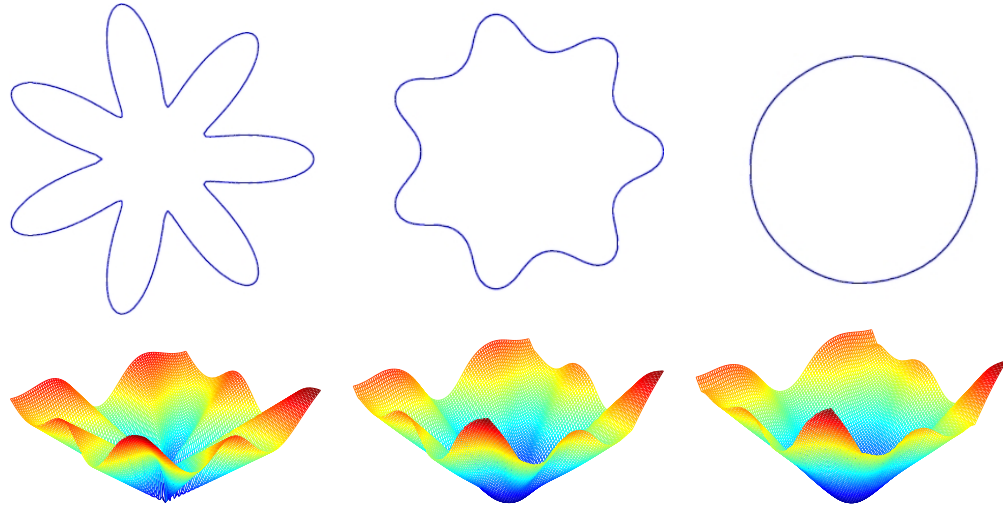
Figure 1.3: Levelset motion by mean curvature (MMC) flow. The leftmost column shows the configuration of the zero level of $\phi(x,0)$ and the full surface $\phi(x,0)$. The two next columns shows corresponding configurations $\phi(x,0.5)$ and $\phi(x,1.5)$. Eventually the circle will disappear in a point. The flow is simulated by the Matlab level set toolbox developed by Ian Mitchell [46, 79].

parameterization of the curves, the zero level of $\phi$ implicitly represents the curves. This is the rational idea of the level set method [61]. The curves are moved by solving a PDE in $\phi$. Which PDE that is solved depends on the application at hand. The implicit representation of the curves makes it possible to deal with topological changes like merging and splitting without any involvement of the difficulties that may occur when representing the curves by an explicit parameterization.

Osher and Sethian derived fundamental equations for interface movement in their 1987 paper [61]. One of these equations describe motion by mean curvature (MMC). In that case, the equation in $\phi(x,t)$ reads

$$\frac{\partial}{\partial t}\phi(x,t) - |\nabla\phi(x,t)|\left(\frac{\nabla\phi(x,t)}{|\nabla\phi(x,t)|}\right) = 0, \tag{1.59}$$

$$\phi(x,0) = \phi_0, \tag{1.60}$$

with appropriate boundary conditions. In Figure 1.3 we illustrate how the zero-level of a function $\phi(x,t)$ develops under the MMC-flow. This flow is fascinating in the sense that every initial curve will shrink to one or more circles, each that will eventually disappear in a point.

In level set modeling, it is customary to require the level set function to be a distance function, to avoid the function to become to flat or steep. One reason

is that a distance function is easier to handle numerically than a more general function. Another reason is that when the level set function is required to be a distance function the involved equations can be simplified [60]. The solution of the Eikonal equation

$$|\nabla \phi| = 1 \tag{1.61}$$

is a distance function. In practice, we can solve the equation

$$\frac{\partial}{\partial t}\phi(x,t) - sgn(\phi(x,t))(1 - |\nabla\phi(x,t)|) = 0, \tag{1.62}$$

$$\phi(x,0) = \phi_0 \tag{1.63}$$

with appropriate boundary conditions to steady state [60]. Note the similarity with the variational regularization formulation introduced earlier in this thesis; an equation of parabolic character is solved to steady-state in order to find the solution to an elliptic equation. Solving equation (1.62) to steady state is called redistancing. It is done at every few iteration of the equation for the evolvement of the level set function. We note that the redistancing procedure does not change the zero level of $\phi$. It merely replaces the function $\phi_0$ by a distance function $\phi$ which has the same zero level as $\phi_0$.

Chan and Vese proposed to use the level set model for image segmentation purposes in a variational setting [16, 83]. More precisely they minimize a binary version of the Mumford-Shah functional which in a simplified form reads

$$F^{MS}(u,\Gamma) = \int_\Omega |u - u_0|^2 \mathrm{d}x + \mu|\Gamma| + \nu \int_{\Omega\backslash\Gamma} |\nabla u|^2 \mathrm{d}x \tag{1.64}$$

with the level set function $\phi$ representing the evolution of the boundary curve $\Gamma$ [51]. In the functional (1.64) $|\Gamma|$ represents the length of the boundary surrounding the object, while $\mu$ and $\nu$ are scalar weighting parameters. This leads to a functional in $\phi$ that is minimized by solving the corresponding Euler-Lagrange equation. For a two phase segmentation (object and background), the idea is that inside the object, $\phi$ should be positive, and outside the object $\phi$ should be negative. Thus the zero level of $\phi$ represents the interface between the object and the background. From an initial configuration $\phi_0$, the goal is to end up with a function $\phi$ such that the zero level of $\phi$ coincides with the interface between the object and the background. Chan and Vese formulated the following functional

$$\begin{aligned} F(c_1, c_2, \Gamma) &= \mu \cdot \mathrm{Length}(\Gamma) + \nu \cdot \mathrm{Area}(\mathrm{inside}(C)) \\ &+ \int_{\mathrm{inside}\Gamma} |u_0 - c_1|^2 \mathrm{d}x + \int_{\mathrm{outside}\Gamma} |u_0 - c_2|^2 \mathrm{d}x, \end{aligned} \tag{1.65}$$

where $c_1 = \text{mean}(u_0)$ inside $\Gamma$ and $c_2 = \text{mean}(u_0)$ outside $\Gamma$. The edge $\Gamma$ is represented by the zero level of the level set function $\phi$. Expressed in terms of $\phi$ the functional (1.65) reads

$$
\begin{aligned}
F(c_1, c_2, \phi) &= \mu \cdot \int_\Omega \delta(\phi)|\nabla\phi|\mathrm{d}x + \nu \cdot \int_\Omega H(\phi)\mathrm{d}x \\
&+ \int_\Omega |u_0 - c_1|^2 H(\phi)\mathrm{d}x + \int_\Omega |u_0 - c_2|^2 (1 - H(\phi))\mathrm{d}x
\end{aligned}
\tag{1.66}
$$

where $H(\phi)$ denotes the Heaviside function

$$
H(\phi) = \begin{cases} 1, & \phi > 0, \\ 0, & \phi \le 0, \end{cases}
\tag{1.67}
$$

and $\delta(\phi)$ denotes the delta function

$$
\delta(\phi) = \begin{cases} 1, & \phi = 0, \\ 0 & \text{elsewhere.} \end{cases}
\tag{1.68}
$$

This functional is minimized using a gradient descent method on $\phi$. Using the shorthand notation from the previous sections, this can be expressed by the equation

$$
\phi_t = -F'(\phi),
\tag{1.69}
$$

or more specifically as

$$
\phi_t = \delta_\epsilon(\phi)\left(\mu\nabla\cdot\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \nu - (u_0 - c_1)^2 + (u_0 - c_2)^2\right),
\tag{1.70}
$$
$$
\phi(x, 0) = \phi_0,
\tag{1.71}
$$

where $\delta_\epsilon(\phi)$ denotes a smooth approximation to $\delta(\phi)$. In order to segment an image into more than two different classes, a number of level set functions $\{\phi_i\}_{i=1}^N$ can be used, with $N = log_2(n)$, where $n$ is the number of classes. The model of Chan and Vese is successfully used in many applications.

Inspired by the work of Chan and Vese [16, 83] and Chan and Tai [21, 75, 32] we have proposed variants of the level set framework for representing domains and their interfaces [40, 41, 42, 43]. We have named these methods *piecewise constant level set methods*, to reflect the fact that at convergence, the level set function is piecewise constant. In the above mentioned papers, two of which appears in the second part of this thesis, we use the proposed methods for image segmentation purposes. As usual in the work in this thesis, we employ standard steepest descent methods. Tai, Christiansen and others have further developed these methods with respect to computational efficiency [76, 22, 77], and they have also generalized the methods from 2D to 3D. Tai, Nielsen, Li, Yao and others have extended the methods in other directions; to interface problems related to characterization of oil reservoirs [53, 57, 54, 56, 55, 39, 38].
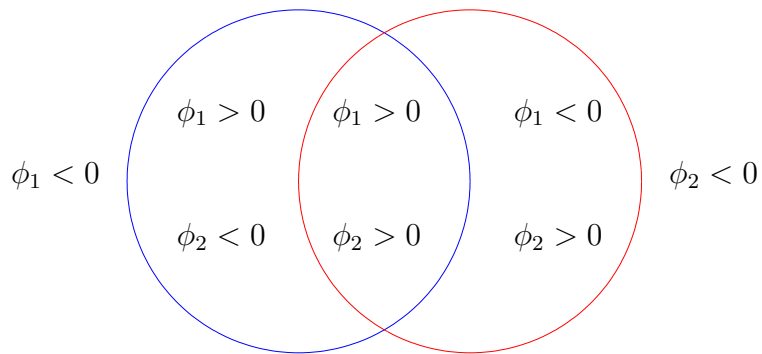
Figure 1.4: The Chan-Vese model can be extended to $n$ phase segmentation by the introduction of $log_2(n)$ level set functions. Here two level set functions are used for four phase segmentation.

# Chapter 2

# Diffusion Tensor Imaging

Matrix valued data arise in many applications, ranging from fluid dynamics, medical imaging and modeling of black holes in astrophysics [91]. In two of the papers in the second part of this thesis, we study regularization methods for matrix valued images, in particular from medical images of the human brain, via Diffusion Tensor Magnetic Resonance Imaging (DTMRI). In this chapter we give an introduction to important concepts of diffusion tensor imaging.

During the last 15 years, DTMRI has become a popular MR modality used both clinically and in research. The main advantage of DTMRI compared to other imaging modalities is the ability to measure direction specific (anisotropic) diffusion of water molecules in living tissue, e.g a human brain. Before we go into the details of DTMRI we give a brief overview of the basic principles of magnetic resonance imaging.

## 2.1 Magnetic Resonance Imaging

We will now give an overview of some of the concepts behind Magnetic Resonance Imaging (MRI) and DTMRI. For a more detailed description, we refer the reader to the papers of Mori, Barker et.al, Bammer and Natt and Frahm [47, 48, 4, 52].

Nuclear Magnetic Resonance (NMR) was discovered independently in 1946 by Edmund Purcell and Felix Block [8, 64]. For their discoveries they received the Nobel Price in 1952. In 1973 the first Magnetic Resonance Image (MRI) was acquired by Paul Lauterbur [37]. Lauterbur was together with Peter Mansfield rewarded the Nobel Price in 2003.

A nucleon with an odd nucleon number (number of protons + number of neutrons) spins around its own axis. The direction of the axis of rotation is randomly distributed in the absence of an external magnetic field. In the Magnetic

Resonance (MR) scanner a strong uniform magnetic field is applied to the nuclei, and the rotation axis of the nuclei gets aligned parallel or antiparallel to the magnetic field. Then a brief electromagnetic signal with the right resonance frequency interacts with the spin, and flips the proton into the opposite direction. When the electromagnetic signal disappear, the protons gradually align with the uniform magnetic field again. During the time of realignment to the uniform field, the nucleons emit energy, which can be measured by a receiver in the MR machine. The realignment time differs among the different tissue classes of the human brain.

In MRI the signal is measured in the Fourier domain. The MRI acquisition can be understood as a sampling of the discrete Fourier transform of the MRI signal. Thus the signal must be transformed to the spatial domain by the inverse discrete Fourier transform before the data can be interpreted a human readable image. Depending on the application at hand, the data can be sampled in more or less time efficient ways. We will not go into details on how the data is sampled and how the data is transformed from the Fourier domain to the spatial domain.

## 2.2   Diffusion Tensor Imaging

The principles behind diffusion NMR were discovered as early as in 1965 by Stejskal and Tanner [73, 72]. Note that this is eight years before the introduction of MRI. Diffusion tensor imaging was introduced as late as 1994 by Basser et. al [5], although diffusivity of water at microscopic length scales was measured by LeBihan et. al in 1985 [23].

Particles suspended in a fluid are not stationary objects. They oscillate and constantly collide with other particles, and with the molecules of the fluid. This phenomena is called Brownian motions, and it has a completely random nature. Depending on properties of the matter in which the particles are suspended, the diffusion can be isotropic (same in all directions) or anisotropic. If we suspend a droplet of ink at a sheet of cleenex paper, the colored molecules will diffuse isotopically, described by the equation

$$\frac{\partial C}{\partial t} = d\nabla^2 C, \tag{2.1}$$

where $C$ is the concentration of ink, and $d$ is a diffusion constant (a positive scalar). Recall that this equation is similar to the heat equation with temperature interchanged with concentration. In DTI, what we measure is the self-diffusion of water. This can be modelled by a probability distribution $P(x, t)$, where $P(x, t)$ describes the probability of finding a particle in a certain position at a particular time, given the initial distribution of particles $P(x, 0)$. In an isotropic medium the
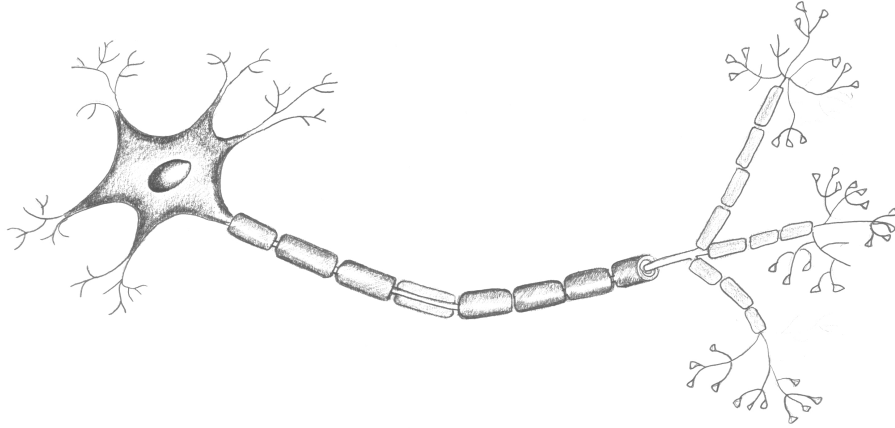
Figure 2.1: The axons are the messengers of the electrical signals between nerve cells in the body. They are surrounded by an insulating shield of myelin. The illustration is drawn by Heidi Lie.

equation which describes the diffusion in terms of the probability function is

$$\frac{\partial P(x,t)}{\partial t} = d\nabla^2 P(x,t). \tag{2.2}$$

We see that this is the same as equation (2.1). The solution to this equation is given by the Gaussian distribution

$$P(x,t) = \frac{1}{\sqrt{4d\pi t}} \exp^{\frac{-\|x\|^2}{4dt}}. \tag{2.3}$$

We can roughly divide the tissue of the human brain into three classes, *white matter*, *gray matter* and *cebro spinal fluid*. In gray matter and cebro spinal fluid water molecules can move more or less freely in all directions. Thus the diffusion is isotropic in these tissue classes. In white matter, the movement of the water molecules is more restricted. Along nerve fibers (axons) the water can move relatively freely. But the axons are surrounded by a shield of myelin (a lipid), which the molecules can not easily penetrate. Furthermore, the axons are bundled together. Thus the diffusion of water molecules becomes anisotropic in white matter. For illustrations of myelinated axons and bundles of them, see Figures 2.1 and 2.2.

In the highly structured tissue, water molecules can diffuse more freely along certain directions than others. This means that the diffusion is anisotropic, and we can describe it by an equation on the form

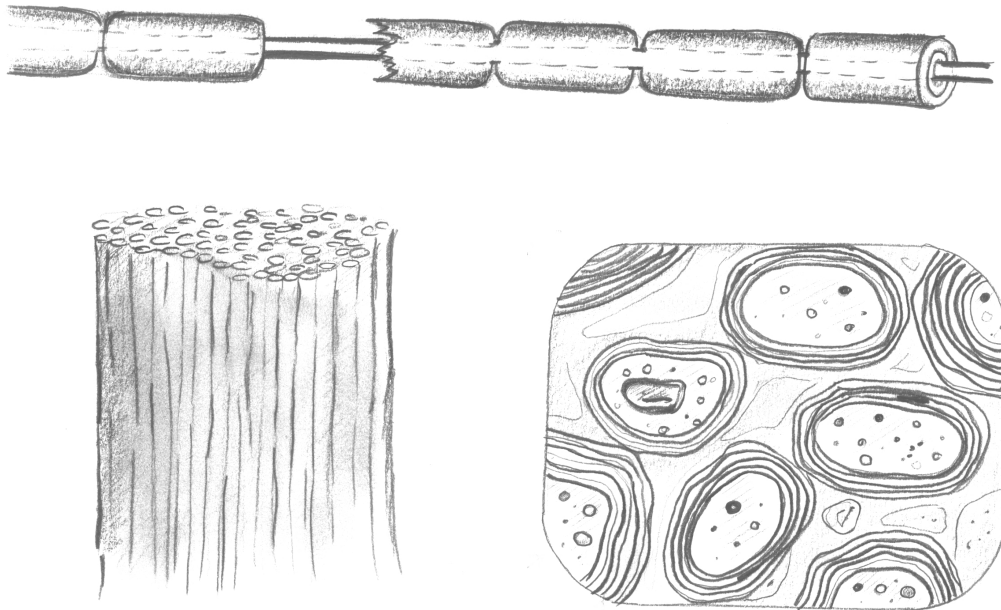$$\frac{\partial P(x,t)}{\partial t} = \nabla \cdot (D\nabla P(x,t)), \tag{2.4}$$

Figure 2.2: The upper figure shows one particular axon with a shield of myelin, which is damaged at one place. The lower leftmost figure shows a bundle of myelinated axons. The lower rightmost figure shows a drawing of a histological slide from the crossection of a bundle of axons. The illustrations are drawn by Heidi Lie.

where $D \in R^{3 \times 3}$ is the so-called *diffusion tensor*. The matrix $D$ contains structural information of the anisotropy of the average diffusion of water. This matrix varies from point to point in the brain. By Diffusion Tensor Imaging (DTI) we measure the local isotropy/anisotropy of the self diffusion of water inside the various types of tissue. We can make pointwise estimates of the matrix $D$ from a series of direction sensitive MR acquisitions. In the MR scanner we do as usual have a strong stationary field $B$. The axis of the spin of water molecules is parallel to the stationary field $B$. As in regular MR, an electromagnetic field pulse is applied to the molecules. The molecules emit signal as they realign with the field. Then a new electromagnetic field pulse flipping the molecules in the opposite direction is applied. The molecules will again emit signal as they realign with the magnetic field. If the molecules were stationary objects, i.e. if no Brownian movements took place, then the measurement of the MR signal after the two pulses would give similar results. But due to the Brownian motion (self diffusion) of the molecules in the period between the two MR pulses, a signal loss can be measured. By measuring this signal loss in a series of different directions, we can get information about the isotropy/anisotropy of the diffusion of water in tissue.
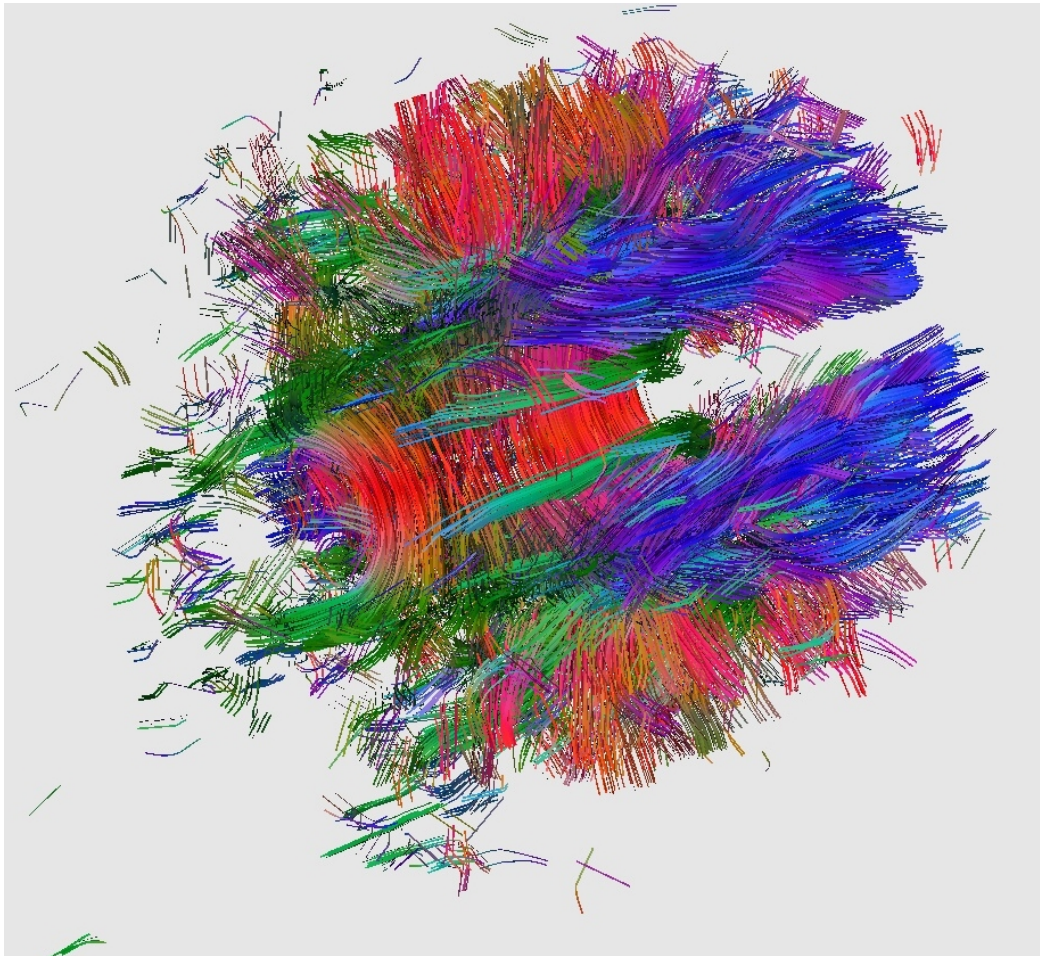
Figure 2.3: An example of a visualization of a human brain using the fibertracking. The data comes from a DTI scan with 25 diffusion directions. The fibertracking was done with the software MedINRIA [26].

From the information of the properties of the diffusion we can construct diffusivity maps which can be visualized directly or we can further use the information to construct a model of the pathways in the human brain from the MR acquisitions. This kind of reconstruction of pathways in tissue is called *fiber tracking* [94, 95, 47]. Fiber tracking is in the literature used for tracking both nerve fibers in white matter in the human brain and muscle fibers in the heart muscle. We show a fiber tracking visualization of the nerve fibers in a human brain in Figure 2.3.

To make a good model of e.g. a brain by fiber tracking, the DTMRI measurements must be of high quality. The main focus of the study of diffusion tensor images in this thesis is on how we can construct good regularization methods for diffusion tensor images. We have mainly focused on PDE models for regulariza-
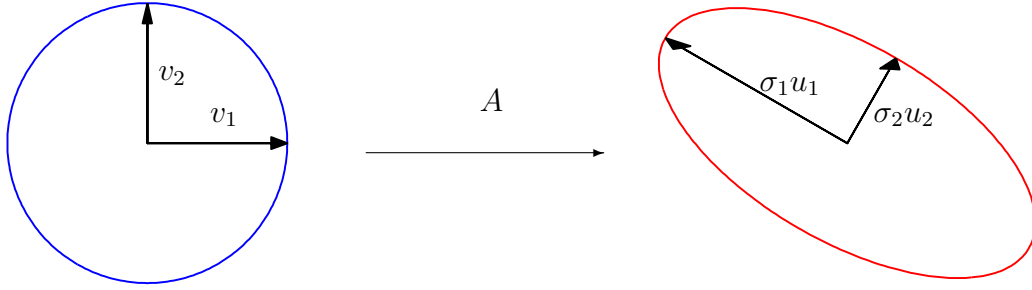
Figure 2.4: The action of any matrix A on the unit sphere is a hyper ellipsis. We write this as $A = U\Sigma V^*$ [82].

tion of diffusion tensor images (Paper D), but recently we have also studied how we can construct transform based methods for regularization of diffusion tensor images (Paper E).

## 2.3 Diffusion Tensor Model

The main ingredient of the diffusion tensor model is the symmetric positive definite (SPD) matrix $D \in R^{3 \times 3}$. This matrix models the self-diffusion coefficient of water-molecules in each voxel in the imaging domain. We will now motivate how we may interpret the diffusion matrix as a hyper ellipsis in $R^3$. Every matrix $A \in C^{M \times N}$ can be decomposed into the factors

$$A = U\Sigma V^*, \tag{2.5}$$

where $U \in C^{M \times M}$ and $V \in C^{N \times N}$ are unitary (orthogonal) matrices, and $\Sigma \in R^{M \times N}$ is a nonnegative diagonal matrix

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n), \qquad \sigma_1 \geq \sigma_2 \geq, \cdots, \geq \sigma_n > 0. \tag{2.6}$$

We denote the complex conjugate by $*$. The factorization (2.5) is called the singular value decomposition (SVD), and is one of the most fundamental matrix decompositions. The singular value decomposition motivates the following interpretation of a matrix; *Any matrix transforms the unit sphere to an hyper ellipsis* [82]. By a hyper ellipsis we mean a higher dimensional generalization of the concept of an ellipsis. A visualization of such a transformation is shown in Figure 2.4 for a matrix $A \in R^{2 \times 2}$. Another important matrix factorization, which we at least in the setting of diffusion tensor modeling can look at as a special case of the SVD, is the eigenvalue decomposition of a matrix $A \in C^{M \times M}$.

$$A = X\Lambda X^{-1} \tag{2.7}$$

where $\Lambda \in C^M$ is the diagonal matrix containing the *eigenvalues* of $A$ and $X \in R^{M \times M}$.x The columns $x_i$ of $X$ are called eigenvectors of $A$. The eigenvalues and eigenvectors obey the following equations

$$Ax_i = \lambda_i x_i \qquad i = 1, 2, \cdots, M. \tag{2.8}$$

This decomposition is however only applicable to quadratic ( $A \in C^{M \times M}$) matrices of full rank. We see that if the columns of $X$ are orthogonal to each other, we get a decomposition which is similar to the singular value decomposition.

In the case where $A = A^*$, that is when $A$ is symmetric, the singular values of $A$ are the absolute values of the eigenvalues of $A$. In addition, when all the eigenvalues are positive, the eigenvalue decomposition equals the singular value decomposition, modulo permutations of the eigenvectors/eigenvalues. In the diffusion tensor model, we deal with the special case of symmetric matrices with positive eigenvalues. Then the hyper ellipsis is called the *diffusion ellipsoid*. We will see that from a physical perspective, symmetry and positivity are crucial properties of an adequate diffusion tensor model.

The physical interpretation of the diffusion ellipsoid is that the length of each semi axis of the ellipsoid determines the ability of water molecules to diffuse along the direction of each particular semi axis. In a completely isotropic region, the hyper ellipsis is simply a scaled unit sphere. In anisotropic regions the hyper ellipsis is a deformed sphere. If one eigenvalue is much larger than the other two, the diffusion ellipsoid is cigar shaped. In this case the prominent part of the diffusion is along the direction of the eigenvector corresponding to the largest eigenvalue. If the two largest eigenvalues are of equal size, and the third is much smaller than the others, then the diffusion ellipsoid is disc-shaped. In this case the prominent part of the diffusion is restricted to a plane spanned by the two eigenvectors corresponding to the two largest eigenvalues. The sphere, cigar, and the disc comprises the three most characteristic ellipsoid shapes. In Figure 2.5 we show ellipsoids which are members of these three main categories.

Note that the diffusion tensor model as described in this thesis can be extended to more complex models, where not only one diffusion tensor is estimated for each voxel in the imaging domain. The voxels are macroscopic while the underlying diffusion process happens at a sub-voxel level. We are therefore only able to measure the average diffusion of molecules in each voxel. If two bundles of nerve fibers are crossing within a single voxel, the one-tensor model will not be able to correctly model diffusion in this voxel. In such voxels multi-tensor models might better characterize the diffusion, see for example the paper by Bergmann et. al [6].
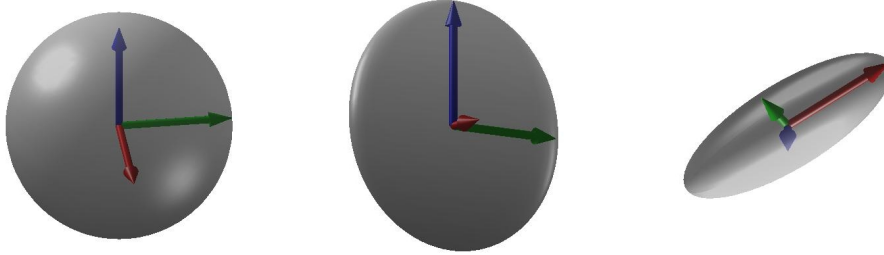
Figure 2.5: Three different types of diffusion ellipsoids, corresponding to spherical- (a), planar- (b) and cigar shaped (c) diffusion.

## 2.4 Diffusion Tensor Estimation

We have already mentioned that the diffusion tensor is estimated from a series of MR acquisitions. We explain this process in detail. We now have a series of $K$ gradient weighted MR measurements $\{S_k\}_{k=1}^{K} : R^3 \to R$, with the corresponding directions $x_k \in R^3$, $k = 1, 2 \cdots , K$. In addition we have the nonweighted image $S_0 : R^3 \to R$. All the images are sampled in the Fourier domain, transformed to the spatial domain on a quadratic mesh in 2D, and stacked together to a 3D image. However, to estimate the tensor we work with one voxel at the time. The Stejskal-Tanner equation

$$S_k = S_0 e^{-bg^T Dg} \tag{2.9}$$

relates the diffusion tensor $D$ to the measurements [73, 72]. Note the similarity with equation (2.3). The scalar $b$ is a machine parameter dependent on the strength of the magnetic field, the acquisition time and other MR parameters [81]. Since the only unknown in (2.9) is the tensor, and this tensor has six degrees of freedom, at least six measurements $S_k$ must be acquired in addition to the nonweighted image. Often the number of direction weighted images is increased in order to make better estimates of the diffusion tensor. It is customary to use 25 or even 50 gradient directions. Another way to increase the accuracy of the estimated tensor is by performing a large series of repeated measurements (NEX) along a relatively low number of gradient directions. Of course, the acquisition time does increase with both the number of gradient directions and the number of repeated measurements. However, with increased measurement time follows patient movement and increased cost per patient. In Paper D we investigate the possibility of decreasing the number of repeated measurements and instead post-process the data.

The estimation of the tensor can for example be done by a linear least squares approach. Let $G$ be the $K \times 3$ matrix with rows $g_k$. From $G$ we derive the matrix

$$\hat{G} = \begin{pmatrix} g_1^2 & 2g_1g_2 & 2g_1g_3 & g_2^2 & 2g_2g_3 & g_3^2 \end{pmatrix}. \tag{2.10}$$

A specific choice of $G$ gives

$$G = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix} \qquad \hat{G} = \begin{pmatrix} 1 & 0 & 2 & 0 & 0 & 1 \\ 1 & 0 & -2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & -2 & 0 & 1 & 0 & 0 \end{pmatrix}. \tag{2.11}$$

By minimizing the functional

$$\min_d \|\hat{G}d - f\|, \tag{2.12}$$

where $d$ is the vector containing the six unique elements of the tensor $D$, and $f$ is the vector with elements $f_k = \frac{ln(S_0) - ln(S_k)}{b}$, we make an estimate of the diffusion tensor. This problem has the formal solution

$$d = (\hat{G}^T \hat{G})^{-1} \hat{G}^T f. \tag{2.13}$$

In the case where $G$ has more than six rows, $\hat{G}$ will also have more than six rows, and the inverse must be interpreted as a pseudo-inverse. We see that in this case, the solution is a least squares solution to the tensor estimation problem.

In the numerical experiments in this thesis we use two different imaging protocols, with six and 25 gradient directions respectively.

## 2.5   A Handful of Anisotropy Measures

We have now estimated the diffusion tensor. What we need next is a way to measure the anisotropy of diffusion tensors. A large set of anisotropic measures are used in the literature [14, 92, 31]. Most of these measures exclusively depend on the eigenvalues of the diffusion tensor, and hence they are rotational invariant. We will show a few of the most commonly used measures of anisotropy [92, 31]. In each voxel of the imaging domain, we can define the *mean diffusivity*

$$\text{MD} = \frac{\text{tr}(D)}{3}. \tag{2.14}$$

This measure differentiates between isotropic and anisotropic tissue. But since it does not contain any information about the relative size of the eigenvalues of $D$,

some information is thrown away if we use this measure. Better measures which takes into account the relative size of the eigenvalues in fact exists. We assume that the eigenvalues are sorted in decreasing order, i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The *relative anisotropy*, which is defined by

$$\text{RA} = \sqrt{\frac{1}{6} \frac{(\bar{\lambda} - \lambda_1)^2 + (\bar{\lambda} - \lambda_2)^2 + (\bar{\lambda} - \lambda_3)^2}{\lambda_1 + \lambda_2 + \lambda_3}}. \tag{2.15}$$

is one such measure. Another commonly used anisotropy measure is the *fractional anisotropy* defined by

$$\text{FA} = \sqrt{\frac{3}{2} \frac{(\bar{\lambda} - \lambda_1)^2 + (\bar{\lambda} - \lambda_2)^2 + (\bar{\lambda} - \lambda_3)^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{2.16}$$

where $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$. Both the fractional anisotropy and the relative anisotropy are used in practical applications. However, it seems like the fractional anisotropy is the measure which is most robust to noise in the measurements [31]. The scaling factors are chosen such that the range of both $RA$ and $FA$ is $[0, 1]$.

The anisotropy of the diffusion tensor can also conveniently be expressed via barycentric coordinates, as done by Westin et. al [92, 36]. They use the three diffusivity parameters

$$c_l = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}, \quad c_p = \frac{2(\lambda_2 - \lambda_3)}{\lambda_1 + \lambda_2 + \lambda_3}, \quad c_s = \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}. \tag{2.17}$$

By construction we always have

$$c_l + c_p + c_s = 1. \tag{2.18}$$

We observe that in regions of completely isotropic diffusion, the index $c_s = 1$, and $c_p = c_l = 0$. In regions where one single diffusion direction is dominant we have $c_l = 1$ and $c_p = c_s = 0$. In the third extreme case the diffusion is restricted to a plane spanned by the eigenvectors corresponding to the two dominant eigenvalues. Then $c_p = 1$ and $c_l = c_s = 0$. We can interpret the triplet $(c_l, c_p, c_s)$ as coordinates in a barycentric coordinate system, as shown in Figure 2.6.

We note that the scalar fractional anisotropy (FA) measure is not as precisely determining the structure of the tensor as the barycentric triplet. But in practice, we are often merely looking for a measure of anisotropy rather than a detailed description of the diffusion tensor. In clinical studies, the fractional anisotropy seems to be the prominent measure of anisotropy. It is commonly used together with the field of eigenvectors corresponding to the largest eigenvalues in order to make RGB direction encoding visualizations of 3D data [62]. This has become a standard way for visualization of anisotropy in diffusion tensor imaging. The
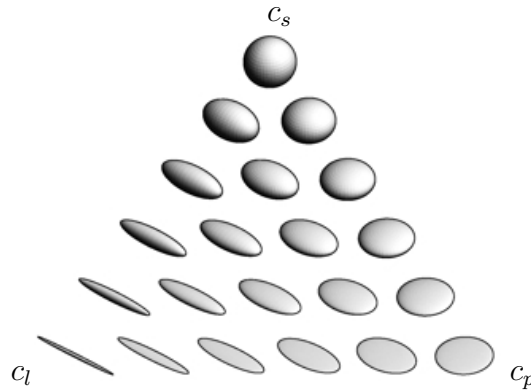
Figure 2.6: A barycentric coordinate system can be used to classify diffusion matrices according to anisotropy. The figure is generated by the software package *teem* written by Gordon Kindlmann [35, 36].
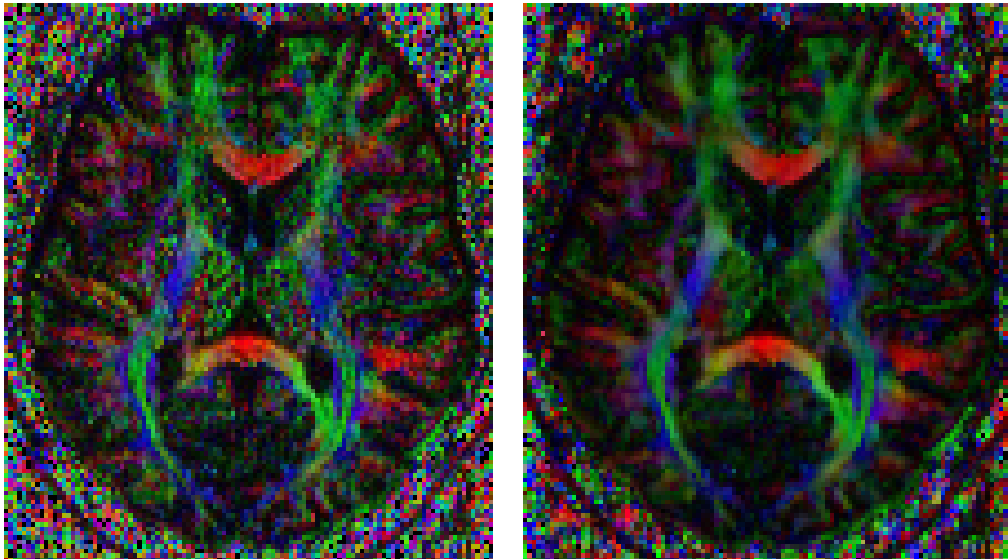


Figure 2.7: An example of a slice of direction encoded DTI. In the leftmost image we show raw data from four measurements and in the rightmost we show the same data regularized by the total variation regularization method developed in Paper D in the second part of this thesis.

strength of the technique is that it displays the fractional anisotropy and directional information about the diffusion at the same time. Another way to visualize diffusion tensor fields is by volume rendering of the diffusion tensor as ellipsoids or more general by volume rendering of glyphs [36, 35].

# Chapter 3

# Summaries of Papers

We have arranged the papers chronologically, based on when the work on the papers was done. The first three papers are already published in journals or conference proceedings. Complete bibliographic details are written on the cover page of each paper. We will now give a short summary of each paper.

# Summary of Paper A

## A Binary Level Set Model and some Applications to Mumford-Shah Image Segmentation

*Johan Lie, Marius Lysaker and Xue-Cheng Tai*

In this paper we develop a new method for representation of domains in for image segmentation applications. As illustrated in the Figure 1.4 and explained in Chapter 1, it is possible to represent a set of domains in an image by the sign of level set functions. At most $log_2(N)$ level set functions are required to represent $N$ domains. It is however possible to represent the domains in other related ways. If we assume that we have an object and the background, we instead represent this by a function which takes the value $\phi = 1$ inside the object and the value $\phi = -1$ at the background. Then we can write an equation for an image consisting of the object and the background as

$$u = \frac{c_1}{2}(\phi + 1) - \frac{c_2}{2}(\phi - 1). \tag{3.1}$$

From this formulation, it is clear that the interface between the object and the background is implicitly defined as points where the function $\phi$ has jumps.

This approach can be generalized in the following way. Assume that the image consists of $N$ different regions $\{\Omega_i\}_{i=1}^{N}$. Then we construct a set of $N$ basis functions $\{\psi_i\}_{i=1}^{N}$ such that

$$\psi_i(x) = \begin{cases} 1 & : & x \in \Omega_i \subset \Omega, \\ 0 & : & x \in \Omega \setminus \Omega_i. \end{cases} \tag{3.2}$$

Using this representation, we may write a function $u$ with $N$ different constant levels as the sum

$$u = \sum_{i=1}^{n} c_i \psi_i. \tag{3.3}$$

We now have a way to represent a piecewise constant image $u$, which we may use for image segmentation purposes by solving a minimization problem. This problem is related to the problem (1.66) in the Chan-Vese method. However, at convergence we require the level set function to have a piecewise constant nature. The basis functions $\{\psi_i\}_{i=1}^{N}$ corresponding to the different regions $\{\Omega_i\}_{i=1}^{N}$ in the image are at convergence only allowed to take the values $1$ or $0$. Therefore we

need to solve a constrained variational problem. We do this by an augmented Lagrangian formulation [7, 58]. Define the segmentation functional

$$F(\phi, c) = \frac{1}{2} \int_\Omega |u - u_0|^2 dx + \beta \sum_{i=1}^N \int_\Omega |\nabla \psi_i| dx, \tag{3.4}$$

and the set of constraints $K(\psi_i)$ such that $K(\psi_i) = 0$ in $\Omega_i$ for all $i = 1, 2, \cdots, N$. We want to solve the constrained problem

$$\min_{\phi, c} \{F(\phi, c)\} \quad \text{subject to} \quad K(\psi_i) = 0. \tag{3.5}$$

The augmented Lagrangian functional which corresponds to the segmentation functional $F(c, \phi)$ is

$$L_\mu(\phi, c, \lambda) = F(\phi) + \sum_{i=1}^N \int_\Omega \lambda_i K_i dx + \frac{\mu}{2} \sum_{i=1}^N K_i^2 dx. \tag{3.6}$$

Here $\mu$ is a scalar, $c$ is a vector of $N$ components and $\lambda$ is a function of the same dimension as $\phi$. Note that the segmentation functional (3.4) have similarities with the functional in the Chan-Vese functional (1.66). The first integral in the functional (3.4) is a fidelity functional, while the second term is a regularization functional which measures the total variation of each basis function. By recalling the properties of the total variation regularizer, it is clear that the regularizer in some sense measures the perimeter of the basis functions.

Following the standard augmented Lagrangian minimization approach [7, 58], we minimize the functional (3.4) by searching for a saddle point of the augmented functional (3.6), i.e. a point where

$$\frac{\partial L_\mu(\phi, c, \lambda)}{\partial \phi} = 0, \quad \frac{\partial L_\mu(\phi, c, \lambda)}{\partial \lambda} = 0, \quad \frac{\partial L_\mu(\phi, c, \lambda)}{\partial c} = 0. \tag{3.7}$$

This means that to find a minimum of the segmentation functional (3.4) we basically need to solve a set of three coupled equations to steady state. However, because of the sometimes severe illposedness of the problem, care must be taken in the optimization process.

In the paper we show numerical experiments on synthetical and semi-realistic images.

# Summary of Paper B

## Piecewise Constant Level Set Methods and Image Segmentation

*Johan Lie, Marius Lysaker and Xue-Cheng Tai*

In this paper we unify the methods from Paper A and another related paper [43] to a more general framework for image segmentation. We give an overview of the methods developed in Paper A and [43]. What is new compared to Paper A is that we do not specify a particular way how to generate the basis functions, but rather try to focus on the basic framework of piecewise constant level set modeling.

# Summary of Paper C

### Inverse Scale Space Methods for
### Nonlinear Regularization

*Johan Lie and Jan M. Nordbotten*

Burger, Osher, Gilboa and Xu introduced a computationally effective way how to compute an inverse scale space flow [11, 12]. This flow is called the *relaxed inverse scale space flow*. It is formulated as a modification of a standard regularization flow, by

$$
\begin{align}
u_t &= -p(u) + \lambda(u - f + v) \tag{3.8} \\
v_t &= \alpha(f - u), \tag{3.9} \\
u(x,0) &= \text{mean}(u), \tag{3.10} \\
v(x,0) &= 0, \tag{3.11}
\end{align}
$$

where $\lambda$ and $\alpha$ are positive scalars, and $v$ is a function of the same dimensions as $u$. The function $p(u)$ is the derivative of a regularization functional. If we set $\alpha > 0$, it is plausible to believe that the second equation will drag the solution from the initial state $u(x,0) = 0$ towards the forcing data $f$. And since the first equation is merely a modification of a standard regularization flow, we might also believe that the flow of $u$ will in some sense smooth this flow. This is in fact what happens in practice.

By inspection, we see that the steady state solution ($u_t = v_t = 0$) of this flow must obey

$$
\begin{align}
u &= f, \tag{3.12} \\
v &= \frac{p(f)}{\lambda}. \tag{3.13}
\end{align}
$$

In the paper we discuss properties of this relaxed scale space flow, and introduce a theoretical foundation for the nonlinear inverse scale space flow as a regularization flow. In particular, we introduce an energy $e(t)$ measuring how far we are from the steady state solution.

We show that the time derivative of the energy is always negative, hence

$$
e(t) < e(0) \quad \forall t. \tag{3.14}
$$

Using this property of the energy , we prove that the relaxed inverse scale space flow will always converge to the steady state solution (3.12) as long as the regularization functional corresponding to $p(u)$ is convex.

We perform some numerical experiments in the paper.

# Summary of Paper D

### Total Variation Regularization of
### Matrix Valued Images

*Oddvar Christiansen, Tin-Man Lee, Johan Lie,*
*Usha Sinha and Tony F. Chan*

In the paper we develop a new regularization method for matrix valued images, in particular for diffusion tensor images of the human brain. We do this by generalizing the model introduced by Blomgren and Chan for vector valued images to matrix valued images [9].

By the model Blomgren-Chan model, a channel with a high total variation norm is stronger regularized than a channel with low total variation norm. Our extension has two essential ingredients; the information flow between the different imaging channels is ensured by the Blomgren-Chan coupling and the positive definiteness of the diffusion matrix $D$ is ensured via the implicit representation $D = LL^T$, where $L$ is a lower triangular matrix. Instead of working directly with the tensor elements $D_{ij}$, we work with the factors $L_{ij}$. The Cholesky factorization ensures that the matrix $D$ is both symmetric and positive definite.

We formulate a minimization problem for matrix valued functions that is very similar to the vector valued minimization problem of Blomgren and Chan

$$
\min_{d_{ij}} \left\{ \mathrm{TV}(D) + \frac{\lambda}{2} \sum_{ij} ||d_{ij} - f_{ij}||_2^2 \right\} \tag{3.15}
$$

We perform numerical experiments on synthetic data and data from DTI acquisitions from a human brain.

# Summary of Paper E

### Shape Adaptive DCT for Denoising of
### Tensor Valued Images

*Ørjan Bergmann, Oddvar Christiansen,*
*Johan Lie and Arvid Lundervold*

In the paper we generalize the so-called Shape Adaptive Discrete Cosine Transform (SADCT) methods [27, 33] from 2D scalar valued images to 3d matrix valued images.

In the SADCT methodology, a region $\Omega_x$ is constructed around each point $x$ in the dataset. This region grows outwards from each data point $x$ until the region "collides" with an edge. Whether or not we have reached an edge is controlled by a statistical criteria called the Intersection of Confidence Interval (ICI) criteria [34]. The data inside each region $\Omega$ is used to construct estimates of the denoised data.

We perform preliminary numerical experiments, and indicate that the SADCT methodology is well suited for denoising of matrix valued data like diffusion tensor images.

# Chapter 4

# Conclusions and Further Work

In this thesis we have touched upon mathematical problems in imaging science. Our focus has mainly been PDE driven methods, and the application area has mainly been medical imaging. We have tried to be open-minded with respect to solution methodology/strategy.

We have developed a framework for segmentation of scalar images which can be seen as a variant of the standard level set methods. The model has been studied to some extent. The model has its own peculiarities, like illposedness. This makes it a nontrivial task to solve the corresponding minimization problems. Work has been done to speed up convergence of these methods, see for example the work by Tai and others.

Furthermore, we have studied properties of relaxed inverse scale space methods, and established a proof that the flow proposed by Burger, Osher and coworkers in fact behaves as intuitively expected. In the relaxed inverse scale space formulation we use concepts from both scale space methods and the regularization methods. It is clear that it should be possible to generalize the relaxed inverse scale space methods also to matrix valued images. This has become clear after we have been working on the regularization flow for matrix valued images. More work must however be done in this direction.

Then we have developed a new mathematical model for regularization of matrix valued images via total variation modeling. This model has been studied to some detail. The model is a clean and clear generalization of the concept of total variation regularization of matrix valued images. The results of the work seems promising. We have in the work on total variation regularization shown that we can decrease the number of required MR acquisitions by postprocessing of the data. However, to really make the model practical, more research must be done to speed up the computations of the solution. We do believe that it is possible to use a dual formulation like the one Chambolle used for the scalar valued total variation regularization problem [15].

Last, we have generalized the SADCT methods from 2D scalar valued images to 3D matrix valued images. As far as we know this is the first attempt to regularize matrix valued data with SADCT methods. We find the results of this work to be very promising. We observe that the methods give at least as good results as the total variation regularization formulation. Concerning the SADCT methods it is clear that these methods can be further developed with respect to computational time. These methods are direct methods, i.e. not iterative methods. And in addition, the methods are quite local in nature. This makes it easy to paralleling the methods; they are "embarrassingly parallelizable". This would make the methods extremely quick.

In all the papers related to PDE methods, we have used the method of steepest descent for the solution of the minimization problems. It is clear that this is not an optimal solution approach with respect to computational time. The piecewise constant level set methods from paper A and B have already been further developed by Tai, Christiansen and others to overcome the slow convergence behaviour. For the regularization of matrix valued images by PDE methods, it would certainly be very interesting to see if this could be done in the same way as Chambolle has done for scalar valued images [15]. This is a very interesting topic for further research.

# Acronyms

**PDE** Partial Differential Equation

**TV** Total Variation

**ROF** Rudin, Osher and Fatemi Model for image denoising

**PM** Perona-Malik

**FA** fractional anisotropy

**SPD** symmetric positive definite

**DTI** Diffusion Tensor Imaging

**DTMRI** Diffusion Tensor Magnetic Resonance Imaging

**MR** Magnetic Resonance

**MRI** Magnetic Resonance Image

**NMR** Nuclear Magnetic Resonance

# Bibliography

[1] Luis Alvarez, Frédéric Guichard, Pierre-Louis Lions, and Jean-Michel Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mech. Anal.*, 123(3):199–257, 1993.

[2] Luis Alvarez, Pierre-Louis Lions, and Jean-Michel Morel. Image selective smoothing and edge detection by nonlinear diffusion. II. *SIAM J. Numer. Anal.*, 29(3):845–866, 1992.

[3] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[4] Roland Bammer. Basic principles of diffusion-weighted imaging. *European Journal of Radiology*, 45:169–184, 2003.

[5] P.J Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.

[6] Ørjan Bergmann, Gordon L. Kindlmann, Arvid Lundervold, and Carl-Fredrik Westin. Diffusion k-tensor estimation from q-ball imaging using discretized principal axes. In *MICCAI*, pages 268–275, 2006.

[7] Dimitri P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.

[8] F. Block. Nuclear induction. *Physical Review*, 7 and 8:460–474, 1946.

[9] P. Blomgren and T.F Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[10] Haim Brezis and Felix Browder. Partial differential equations in the 20th century. *Advances in Mathematics*, 135:76–144, 1998.

[11] Martin Burger, Guy Gilboa, Stanley Osher, and Jinjun Xu. Nonlinear inverse scale space methods. *Communications in Mathematical Sciences*, 4(1):179–212, 2006.

[12] Martin Burger, Stanley Osher, Jinjun Xu, and Guy Gilboa. Nonlinear inverse scale space methods for image restoration. *Lecture Notes in Computer Science*, 3752:25–36, 2005.

[13] Francine Catté, Pierre-Louis Lions, Jean-Michel Morel, and Tomeu Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.

[14] Mara Cercignani, Matilde Inglese, Elisabetta Giancarlo Pagani, Massimo Comi, and Massimo Filippi. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *AJNR Am J Neuroradiol.*, 22(5):952–958, 2001.

[15] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

[16] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[17] T.F. Chan, S. Esedoglu, F. Park, and A. Yip. *Handbook of Mathematical Models in Computer Vision*, chapter Recent Developments in Total Variation Image Restoration. Springer, 2006.

[18] Tony F. Chan, Ke Chen, and Xue-Cheng Tai. Nonlinear multilevel schemes for solving the total variation image minimization problem. *Proc. 1st Int. Conf. PDE-Based Image Processing and Related Inverse Problems*, 2006.

[19] Tony F. Chan and Selim Esedoglu. Aspects of total variation regularized $L_1$ function approximation. *SIAM Journal of Applied Mathematics*, 65(5):1817–1837, 2005.

[20] Tony F. Chan and Jianhong Shen. *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*. SIAM, 2005.

[21] Tony F. Chan and Xue-Cheng Tai. Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM J. Sci. Comput.*, 25:881–904, 2003.

[22] Oddvar Christiansen and Xue-Cheng Tai. Fast implementation of piecewise constant level set methods. In *Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems, CMA, Oslo, August 8-12, 2005*, Mathematics and Visualization. Springer, 2006.

[23] LeBihan D. and Breton E. Imagerie de diffusion in-vivo par résonance magnétique nucléaire (french). *C R Acad Sci (Paris)*, 301(15):1109–1112, 1985.

[24] Jerome Darbon and Marc Sigelle. Image restoration with discrete constrained total variation part i: Fast and exact optimization. *Journal of Mathematical Imaging and Vision*, 26(3), 2006.

[25] L. C. Evans. *Partial Differential Equations*. Number 19 in Graduate Studies in Mathematics. AMS, 1998.

[26] Pierre Filliard and Nicolas Toussaint. MedINRIA: Medical image navigation and research tool by inria.

[27] A. Foi, K. Dabov, V. Katkovnik, and K. Egiazarian. Shape-adaptive dct for denoising and image reconstruction. In *Proc. SPIE Electronic Imaging 2006, Image Processing: Algorithms and Systems V, San Jose*, 2006.

[28] I. M. Gelfand and S. V. Fomin. *Calculus of variations*. Revised English edition translated and edited by Richard A. Silverman. Prentice-Hall Inc., Englewood Cliffs, N.J., 1963.

[29] Gene H. Golub, Per C. Hansen, and Dianne P. O'Leary. Tikhonov regularization and total least squares. *Siam J. Matrix Analysis and Applications*, 21(1):185–194, 1999.

[30] Per Christian Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Number 4 in SIAM Monographs on Mathematical Modeling and Computation. SIAM, 1997.

[31] Khader M. Hasan, Andrew L. Alexander, and Ponnada A. Narayana. Does fractional anisotropy have better noise immunity characteristics than relative anisotropy in diffusion tensor mri? an analytical approach. *Magnetic Resonance in Medicine*, 51:413417, 2004.

[32] B. Heimsund, T. F. Chan, T. K. Nilssen, and X. C. Tai. Level set methods and augmented lagrangian for a parameter identification problem. In V. Barbu, I. Lasiecka, D. Tiba, and C. Varsan, editors, *Analysis and optimization of differential systems*, pages 189–200. Kluwer Academic Publishers, Boston/Dordrecht/London, 2003.

[33] IEEE SP/CAS and Tampere International Center for Signal Processing. *POINTWISE SHAPE-ADAPTIVE DCT AS AN OVERCOMPLETE DENOISING TOOL*, number 5 in The International Workshop on Spectral Methods and Multirate Signal Processing, Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland, June 2005.

[34] Vladimir Katkovnik, Karen Egiazarian, and Jaakko Astola. Adaptive window size image de-noising based on intersection of confidence intervals (ICI) rule. *J. Math. Imaging Vision*, 16(3):223–235, 2002. Nonlinear image processing and pattern recognition, XII (San Jose, CA, 2001).

[35] Gordon Kindlmann. Teem: Tools to process and visualize scientific data and images. http://teem.sourceforge.net/.

[36] Gordon Kindlmann. Superquadric tensor glyphs. 2004.

[37] P.C Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

[38] Hongwei Li and Xue-Cheng Tai. Piecewise constant level set method for interface problems. CAM-UCLA 06-05, 2006.

[39] Hongwei Li and Xue-Cheng Tai. Piecewise constant level set methods for multiphase motion. CAM-UCLA 06-17, 2006.

[40] Johan Lie, Marius Lysaker, and Xue-Cheng Tai. A piecewise constant level set framework. In P. Neittaanmaki, T. Rossi, S. Korotov, E. Onate, , and D. Knorzer, editors, *European Congress on Computational Methods in Applied Sciences and Engineering ECCOMAS 2004*, 2004.

[41] Johan Lie, Marius Lysaker, and Xue-Cheng Tai. Piecewise constant level set methods and image segmentation. In Ron Kimmel, Nir Sochen, and Joachim Weickert, editors, *Scale Space and PDE Methods in Computer Vision: 5th International Conference, Scale-Space 2005*, volume 3459 of *Lecture Notes in Computer Science*. Springer-Verlag GmbH, 2005.

[42] Johan Lie, Marius Lysaker, and Xue-Cheng Tai. A binary level set model and some applications to Mumford-Shah image segmentation. *IEEE Transactions on Image Processing*, 15(5):1171–1181, 2006.

[43] Johan Lie, Marius Lysaker, and Xue-Cheng Tai. A variant of the level set method and applications to image segmentation. *AMS Mathematics of Computation*, 75:1155–1174, 2006.

[44] Marius Lysaker, Stanley Osher, and Xue-Cheng Tai. Noise removal using smoothed normals and surface fitting. *IEEE Trans. Image Processing*, 13(10):1345–1357, 2004.

[45] Ravikanth Malladi, editor. *Geometric Methods in Bio-Medical Image Processing*. Mathematics and Visualization. Springer, 2002.

[46] Ian Mitchell. A toolbox of level set methods. Technical Report TR-2004-09, UBC Department of Computer Science, 2004.

[47] Susumi Mori and Peter C.M. van Zijl. Fiber tracking: principles and strategies - a techical review. *NMR Biomed*, 15:468–480, 2002.

[48] Susumo Mori and Peter B. Barker. Diffusion magnetic resonance imaging: Its principle and applications. *The Anatomical Record (NEW ANAT.)*, 257:102–109, 1999. Feature Article.

[49] P. Mrazek. Monotonicity enhancing nonlinear diffusion. *Journal of Visual Communication and Image Representation*, 13(1):313 – 323, 2002.

[50] Pavel Mrázek. *Nonlinear Diffusion for Image Filtering and Monotonicity Enhancement*. PhD thesis, Center for Machine Perception, Department of Cybernetics Faculty of Electrical Engineering, Czech Technical University, 2001.

[51] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and assosiated variational problems. *Comm. Pure Appl. Math*, 42, 1989.

[52] Oliver Natt and Jens Frahm. In vivo magnetic resonance imaging: insights into structure and function of the central nervous system. *Measurement Science and Technology*, 16(4), 2005.

[53] Lars Kristian Nielsen. *Reservoir Characterisation by a Binary Level Set Method and Adaptive Multiscale Estimation*. PhD thesis, University of Bergen, 2006.

[54] L.K. Nielsen, X.-C. Tai, S.I. Aanonsen, and M. Espedal. A binary level set model for elliptic inverse problems with discontinuous coefficients. Submittet to Inverse Problems,2005, 2006.

[55] L.K. Nielsen, X.-C. Tai, S.I. Aanonsen, and M. Espedal. Identifying reservoir facies using a binary level set method. Sumbitted to Society of Petroleum Engineers Journal, 2006.

[56] L.K. Nielsen, X.-C. Tai, S.I. Aanonsen, and M. Espedal. Reservoir description using a binary level set model. Submitted to Computation and Visualization in Science, 2006.

[57] L.K. Nielsen, X.-C. Tai, S.I. Aanonsen, and M. Espedal. Reservoir description using a binary level set approach with additional prior information about the reservoir model. In Conference Proceedings of PDE-Based Image Processing and Related Inverse Problems, Oslo, August, 2005, 2007.

[58] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[59] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. Using geometry and iterated refinement for inverse problems (1): Total variation based image restoration. *UCLA Computational and Applied Mathematics Report*, 04-13, 2004.

[60] Stanley Osher and Ronald Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2003.

[61] Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.

[62] Sinisa Pajevic and Carlo Pierpaoli. Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain. *Magnetic Resonance in Medicine*, 42:526–540, 1999.

[63] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.

[64] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69:37–38, 1946.

[65] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica D.*, 60:259–268, 1992.

[66] Leonid I. Rudin. *Images, Numerical Analysis of Singularities and Shock Filters*. 1987.

[67] C. Samson, Blanc L. Feraud, G. Aubert, and J. Zerubia. A level set model for image classification. *IJCV*, 40(3):187–198, 2000.

[68] Guillermo Sapiro and Dario L. Ringach. Anisotropic diffusion of multival-ued images with applications to color filtering. *IEEE Transactions on Image Processing*, 5(11):1582–1586, 1996.

[69] Otmar Scherzer and Chuck Groetsch. Inverse scale space theory for inverse problems. In M. Kerckhove, editor, *Scale-Space and Morphology in Computer Vision : Third International Conference, Scale-Space 2001*, volume 2106 of *Lecture Notes in Computer Science*. Springer-Verlag GmbH, 2001.

[70] Otmar Scherzer and Joachim Weickert. Relations between regularization and diffusion filtering. *J. Math. Imaging Vision*, 12(1):43–63, 2000.

[71] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1999.

[72] E. O. Stejskal. Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow. *The Journal of Chemical Physics*, 43(10):3597–3603, 1965.

[73] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *The Journal of Chemical Physics*, 42(1):288–292, 1965.

[74] Walter A. Strauss. *Partial Differential Equations: An Introduction*. Whiley, 1992.

[75] Xue-Cheng Tai and Tony F. Chan. A survey on multiple level set methods with applications for identifying piecewise constant functions. *International J. Numerical Analysis and Modelling*, 1(1):25–48, 2004.

[76] Xue-Cheng Tai, Oddvar Christiansen, Ping Lin, and Inge Skjaelaaen. Image segmentaion using some piecewise constant level set methods with mbo type of projection. *CAM-UCLA 05-24, to appear in International Journal of Computer Vision*, 2005.

[77] Xue-Cheng Tai and Chang-Hui Yao. Image segmentation by piecewise constant mumford-shah model without estimating the constants. CAM-UCLA 06-18, 2006.

[78] T. Tasdizen, R. Whitaker, P.Burchard, and S.Osher. Geometric processing via normal maps. *ACM Trans. on Graphics*, 22:1012–1033, 2003.

[79] The Mathworks. MatLab, The Language of Technical Computing.

[80] A. Tikhonov and V.A Arsenin. *Solutions of ill-posed problems*. VH Winston, 1977.

[81] Paul Tofts, editor. *Quantitative MRI of the Brain*. John Wiley and Sons, 2005.

[82] Lloyd N. Trefethen and III David Bau. *Numerical Linear Algebra*. SIAM, 1997.

[83] L. A. Vese and T. F. Chan. A multiphase level set framevork for image segmentation using the mumford and shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[84] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.

[85] Curtis A. Vogel. *Computational Methods for Inverse Problems*. Frontiers in Applied Mathematics. SIAM, 2002.

[86] Z. Wang, B.C Vemuri, and Y Chenand T Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex dwi. *IEEE Transactions on Medical Imaging*, 23(8):930 – 939, 2004.

[87] J. Weickert. Applications of nonlinear diffusion in image processing and computer vision. *Acta Math. Univ. Comenian. (N.S.)*, 70(1):33–50, 2000.

[88] Joachim Weickert. A review of nonlinear diffusion filtering. *Lecure Notes in Computer Science*, 1252:3 – 28, 1997. Proceedings of the First International Conference on Scale-Space Theory in Computer Vision.

[89] Joachim Weickert and Brahim Benhamouda. Why the perona-malik filter works. Technical Report DIKU-RT-97/22, Dep. Comp. Sci. University of Copenhagen, 1997.

[90] Joachim Weickert and Thomas Brox. Diffusion and regularization of vector- and matrix-valued images. Universitat des Saarlandes, Fachrichtung 6.1 Mathematik, Preprint No. 58, 2002.

[91] Joachim Weickert and Hans Hagen (Eds.). *Visualization and Processing of Tensor Fields*. Mathematics and Visualization. Springer, 2005.

[92] C.-F Westin, S.E. Maier, H. Mamata, A. Nabavi, F.A Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor mri. *Medical Image Analysis*, 6:93–108, 2002.

[93] Wotao Yin, Donald Goldfarb, and Stanley Osher. Image cartoon-texture decomposition and feature selection using the total variation regularized l1 functional. cam05-47, 2005.

[94] Leonid Zhukov and Alan H. Barr. Oriented tensor reconstruction: Tracing neural pathways from diffusion tensor mri. IEEE Visualization 2002, 2002.

[95] Leonid Zhukov and Alan H. Barr. Heart-muscle fiber reconstruction from diffusion tensor mri, 2003.