NTNU
Norwegian University of
Science and Technology

Minimal Surfaces in Sub-Riemannian Geometries with Applications to Perceptual Completion

Per Martin Viddal

Master of Science in Physics and Mathematics
Submission date: June 2009
Supervisor: Brynjulf Owren, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

We will consider a perceptual completion model proposed by Citti and Sarti where a part of the visual cortex is modelled by using a sub-Riemannian geometry on the Lie group SE(2). A background study of sub-Riemannian geometries and its geodesics shall be done, and different techniques to solve the minimal surface problem shall be described and implemented.

Assignment given: 19. January 2009
Supervisor: Brynjulf Owren, MATH

Preface

This master thesis is written in my final semester of the Master's degree programme in Industrial Mathematics at the Norwegian University of Science and Technology (NTNU), Trondheim. The topic is geometric integration and is part of the GALA project[1], which is a specific targeted research project within the NEST (new and emerging science and technology) activity of the 6th framework of the European Union.

The work is a proceeding of my pre-master project. In connection to this subject, I have been privileged to attain several activities during the last year. I met several of the partners of the GALA-project when a guest lecture was held at NTNU last fall. The following discussions were an inspiration in the initial phase of my work. I also attended MaGIC[2] at Lillehammer this spring where I met many of the leading researchers within the field of geometric integration.

I would like to thank my supervisor Prof. Brynjulf Owren for helpful ideas and weekly discussions throughout this last year. His enthusiasm regarding this subject is an inspiration in itself, and he has always helped me with any problems I have run into.

I would also like to thank my live-in partner Åshild for support when I needed it. My mood has probably fluctuated according to any progress or setbacks in connection with this work, and her encouragement is deeply appreciated.

Trondheim, June 2009
Per Martin Viddal

---

[1] Sub-Riemannian geometric analysis in Lie Groups, "http://www.gala.unibo.it/".

[2] Manifolds and Geometric Integration Colloquia, organized by the Numerical Analysis groups at the University of Bergen and NTNU.

Abstract

A preliminary study of the papers [1] and [2] is done. The first one, written by Citti and Sarti, describe a perceptual completion model where a part of the visual cortex is modelled using a sub-Riemannian geometry on the Lie group SE(2). The second one, written by Hladky and Pauls, describe a model which completes the interior of a circular hole by spanning the lifted boundary by a minimal surface, presuming such a surface exists. These surfaces are solutions of occluded visual data as described by Citti and Sarti.

Based on the models above, we propose a new model. The lifted boundary of an arbitrary hole is spanned by a surface consisting of geodesics between points with matching Dirichlet boundary values.

All the three models are based on the sub-Riemannian geometry for the roto-translational space introduced by Citti and Sarti. The basic theory of sub-Riemannian geometries, including the derivation of some flows and operators in this degenerate space, is described. The models are implemented, and numerical results are presented.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background and historical remarks

Image analysis is the extraction of meaningful information from images. Digital images are mainly considered, as image processing techniques are solved using computers. Numerous numerical algorithms have been proposed, each with its own properties. The wide range of applications includes noise removal, reconstruction of holes and object recognition, among others. The presumably most famous one, the Perona-Malik-equation [3], result in noise removal and image enhancement by applying an anisotropic diffusion process on the original image.

In this thesis, we will consider algorithms which exploit main phenomenological properties described by psychology of Gestalt. Disocclusion or "amodal completion" is a very common process in human vision. Naturally, objects are rarely visible in its entirety, but our perception is often able to "reconstruct" the whole object by interpolating the missing parts. We call this phenomenon perceptual completion.

In Figure 1.1 we see two example images named after the psychologist G. Kanizsa. In the image on the left, we perceived a triangle in the centre even though it is not drawn. The other image shows two fishes which are blocked by, and at the same time blocking, the other. This last figure is a good example of disocclusion, where we perceive two distinct objects.


(a) Kanizsa's triangle.　(b) Kanizsa's fishes.

Figure 1.1: Two figures named after psychologist G. Kanizsa.

The continuation of object boundaries appears to play a central role of perceptual completion. The models in [4] and [5] consist of minimizing the functional

$$\int_{\Omega} |\nabla v| \left(1 + \left|\operatorname{div}\left(\frac{\nabla v}{|\nabla v|}\right)\right|^p\right) \mathrm{d}x\,\mathrm{d}y, \quad p > 1 \tag{1.1}$$

subject to the constraint $v|_{\Gamma^C} = u|_{\Gamma^C}$, where $u : \mathbb{R}^2 \to \mathbb{R}$ is the original image and Γ is the domain occupied by the occluding object. Further, p is a constant and the integral is computed over the image domain $\Omega \subset \mathbb{R}^2$. The first part of the functional (1.1) forces restored level lines inside Γ to have short length, while the second part forces them to have small curvature. Combined, this minimization produces level lines which are geodesics, or "shortest curves" in some sense, connecting two points on the boundary of Γ.

To maintain both occluded and occluding objects, a third dimension can be introduced. The objects are then present at the same time, ordered by depth as in [6] and [7]. When there is no clear depth ordering though, as in Figure 1.1(b), these models fail.

The model in [8] considers the perceptual completion phenomena as a problem of naturalizing phenomenological models on the basis of biological and neuropsychological evidence. The model mimics the functionality of the area V1 in the visual cortex, which is the place where visual stimuli are processed [9]. V1 consists of simple cells which have several properties, among them an orientation property [10]. The functional (1.1) is then replaced by a first order length minimizing functional in the natural contact structure induced in a three-dimensional space, more suitable for problems where a clear depth ordering is not present.

In the paper by G. Citti and A. Sarti [1], which is the main reference in this thesis, this point of view is developed further. It is based on the lifting of level lines, and the natural space in which the completion is performed is the roto-translational space, $\mathbb{R}^2 \times S^1$, denoted $\mathcal{RT}$. Both occluding and occluded objects are completed at the same time in this model. It is diffusion-driven, and the completed image has level lines which are geodesic curves in the metric of the space, making the completed surface in $\mathcal{RT}$ a minimal surface.

In the recent paper [2] by R. K. Hladky and S. D. Pauls, minimal surfaces of $\mathcal{RT}$ have been investigated further. From the model above, these surfaces are completions of occluded visual data. Assuming a circular hole in an image, the boundary of this hole is lifted to the three-dimensional space $\mathcal{RT}$. Then a minimal surface which spans this boundary is constructed using helix cords, presuming such a surface exist. Circular holes in images can in this way be completed using Dirichlet boundary conditions.

1.2 Outline of the thesis

In Chapter 2 we will clarify notations and definitions used throughout this thesis. The roto-translational space $\mathcal{RT}$ will be introduced, and we build up the underlying sub-Riemannian geometry which is essential in the models that follow.

Different flows, operators and partial differential equations will be described in Chapter 3. They all have a degenerate counterpart related to the sub-Riemannian geometry introduced by Citti and Sarti. This model consists of two partial differential equations, and we describe their relation to the Laplace-Beltrami flow and the mean curvature motion problem, respectively.

Chapter 4 addresses the model by Citti and Sarti [1]. In the pre-master [11] by P. M. Viddal, this model has been analysed in more detail, thus we will keep it brief here. The basic features of the model are explained and we state the numerical approximations used.

The model of Hladky and Pauls [2] to construct minimal surfaces in $\mathcal{RT}$ is described in Chapter 5. The theory is discussed with example figures.

In Chapter 6 we will introduce a new model to construct minimal surfaces. Much of the theory is based on [2], but instead of constructing helix cords, we will construct numerical approximations of geodesics. These geodesics are solutions of the geodesic equations found in Chapter 2, which is generated from the Hamiltonian function for the sub-Riemannian geometry on $\mathcal{RT}$. With this model, we can also consider arbitrary holes.

Numerical results are presented in Chapter 7. For some example images, results from the different models are compared. We also show some images which are problematic to complete using the two latter models.

Chapter 2

Preliminaries

We start by refreshing the basic definitions in Lie group theory in Section 2.1. For the roto-translational group, denoted $\mathcal{RT}$, we find an explicit representation of its Lie algebra. Next we define distributions and one-forms in Section 2.2 and Section 2.3, and relate this to $\mathcal{RT}$. In Section 2.4 we define a sub-Riemannian geometry. An explicit sub-Riemannian geometry on $\mathcal{RT}$ is found, and some of its characteristic properties are stated. Geodesics are the topic of Section 2.5. A geodesic is defined, and the "geodesic equations" are derived as a set of ordinary differential equations.

2.1 Lie groups and Lie algebras

The roto-translational group $\mathcal{RT}$ is a Lie group, see [1] or [11] for details. $\mathcal{RT}$ is homeomorphic to $\mathbb{R}^2 \times S^1$ and the equivalent matrix group $SE(2)$ (rigid motion).

Definition 2.1. *The* ***Lie algebra*** *of a Lie group G, denoted $\mathfrak{g}$, is the vector space of all left-invariant vector fields on G.*

Remark. The orthogonal vector fields

$$\begin{aligned} X_1 &= \cos\theta\partial_x + \sin\theta\partial_y, \\ X_2 &= \partial_\theta, \\ X_3 &= -\sin\theta\partial_x + \cos\theta\partial_y, \end{aligned} \tag{2.1}$$

constitute an explicit frame for the tangent space $T(\mathcal{RT})|_{\mathbf{x}}$ at any point $\mathbf{x} = (x, y, \theta) \in \mathcal{RT}$. These vector fields vary smoothly with $\mathbf{x}$, and have been shown to be left-invariant in [11]. Thus $\mathfrak{g} = \text{span}(X_1, X_2, X_3)$ is an explicit representation of the Lie algebra of $\mathcal{RT}$.

2.2 Distributions

Before we find a sub-Riemannian geometry on $\mathcal{RT}$ we need to define a distribution. The definitions below are collected from [12].

Definition 2.2. *A* ***distribution****, $\mathcal{H}$, is a linear subbundle of the tangent bundle of a manifold, $\mathcal{H} \subset T\mathcal{M}$. That is, for every point $\mathbf{x} \in \mathcal{M}$, the distribution is a subspace of the tangent space at $\mathbf{x}$, $\mathcal{H}|_{\mathbf{x}} \subset T\mathcal{M}|_{\mathbf{x}}$.*

Remark. Since $\mathcal{RT}$ is a Lie group, it carries the structure of a manifold. From the vector fields (2.1) on $\mathcal{RT}$, we define the distribution

$$\mathcal{H} = \operatorname{span}\left(X_1, X_2\right), \tag{2.2}$$

which will play a central role in later chapters.

The Lie bracket, or commutator, of X_1 and X_2 can be found through their action on a smooth function $u(x, y, \theta)$ as

$$\left[X_1, X_2\right] u = X_1(X_2 u) - X_2(X_1 u) = -\left(-\sin\theta u_x + \cos\theta u_y\right) = -X_3 u. \tag{2.3}$$

We see that X_1 and X_2 "generates" X_3 through the Lie bracket operation.

Definition 2.3. *A distribution $\mathcal{H}$ is called* ***bracket generating*** *if any local frame $\{X_i\}$ for $\mathcal{H}$, together with all of its iterated Lie brackets $[X_i, X_j], [X_i, [X_j, X_k]], \ldots$, spans the tangent bundle $T\mathcal{M}$.*

Definition 2.4. *For a distribution $\mathcal{H}$ on a manifold $\mathcal{M}$, let $\mathcal{H}^r$ be defined recursively as*

$$\mathcal{H}^r = \mathcal{H}^{r-1} + \left[\mathcal{H}, \mathcal{H}^{r-1}\right],$$

with $\mathcal{H}^1 = \mathcal{H}$, and where $\left[\mathcal{H}, \mathcal{H}^k\right] = \operatorname{span}\left\{[X, Y] : X \in \mathcal{H}, Y \in \mathcal{H}^k\right\}$. Moreover, we indicate $\mathcal{H}^k|_{\mathbf{x}}$ the vector space corresponding to $\mathcal{H}^k$ evaluated at $\mathbf{x} \in \mathcal{M}$. If m is the smallest natural number such that

$$\mathcal{H}^m|_{\mathbf{x}} = T\mathcal{M}|_{\mathbf{x}},$$

we say that $\mathcal{H}$ is bracket generating with ***step*** *m at $\mathbf{x}$.*

Remark. From above we have that $[X_1(\mathbf{x}), X_2(\mathbf{x})] = -X_3(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{RT}$, thus X_1, X_2 and $[X_1, X_2]$ spans $T(\mathcal{RT})$, and $\mathcal{H}$ (2.2) is a bracket generating distribution of step 2.

2.3 One-forms

Definition 2.5. *The* ***cotangent bundle*** *of a differentiable manifold $\mathcal{M}$, denoted $T^*\mathcal{M}$, is the dual of the tangent bundle. Elements in the cotangent bundle are linear functionals on each fiber of the tangent bundle.*

Definition 2.6. *A differential* ***one-form*** *Θ on a manifold $\mathcal{M}$ is a smooth section of the cotangent bundle $T^*\mathcal{M}$:*

$$T^*\mathcal{M} \ni \Theta : T\mathcal{M} \to \mathbb{R},$$

in the sense that at a point $\mathbf{x} \in \mathcal{M}$, $\Theta_{\mathbf{x}}$ is a linear functional on the tangent space:

$$T^*\mathcal{M}|_{\mathbf{x}} \ni \Theta_{\mathbf{x}} : T\mathcal{M}|_{\mathbf{x}} \to \mathbb{R}.$$

Let $\mathcal{H} = \operatorname{span}(Y_1, \ldots, Y_{n-1})$ be a distribution on the tangent bundle $T\mathcal{M} = \operatorname{span}(Y_1, \ldots, Y_n)$ with $Y_1, \ldots, Y_n$ mutually orthogonal. With the vector field $Y_n(\mathbf{x}) = f_1(\mathbf{x})\partial_{x_1} + \ldots + f_n(\mathbf{x})\partial_{x_n}$ in local coordinates, we can define the corresponding one-form $\Theta_{\mathbf{x}} = f_1(\mathbf{x})\mathrm{d}x_1 + \ldots + f_n(\mathbf{x})\mathrm{d}x_n$. The distribution $\mathcal{H}$ at $\mathbf{x}$ can then be defined as the kernel of $\Theta_{\mathbf{x}}$:

$$\mathcal{H}|_{\mathbf{x}} = \ker \Theta_{\mathbf{x}}. \tag{2.4}$$

Remark. From the vector field X_3 (2.1), we can define the corresponding one-form $\Theta_{(x,y,\theta)} = -\sin\theta \mathrm{d}x + \cos\theta \mathrm{d}y$. The distribution $\mathcal{H}$ (2.2) above can then be defined as

$$\mathcal{H}|_{(x,y,\theta)} = \ker\Theta_{(x,y,\theta)} = \{(v_1, v_2, v_3) : \ -\sin\theta v_1 + \cos\theta v_2 = 0\}. \tag{2.5}$$

This means that any vector $v = (v_1, v_2, v_3) \in T(\mathcal{RT})|_{(x,y,\theta)}$ lies in the distribution provided that $v_1 = \alpha\cos\theta, v_2 = \alpha\sin\theta$, where α is a constant.

The next definitions are collected from [13].

Definition 2.7. *A one-form Θ on a $(2n+1)$-dimensional smooth manifold is called a* ***contact form*** *if*

$$\Theta \wedge (\mathrm{d}\Theta)^n \neq 0,$$

where

$$(\mathrm{d}\Theta)^n = \underbrace{\mathrm{d}\Theta \wedge \ldots \wedge \mathrm{d}\Theta}_{n \text{ times}}.$$

Definition 2.8. *A* ***contact distribution*** *is a distribution defined as the kernel of a contact form.*

To check if our one-form is a contact form, we differentiate:

$$\omega = \mathrm{d}\Theta = -\cos\theta \mathrm{d}\theta \wedge \mathrm{d}x - \sin\theta \mathrm{d}\theta \wedge \mathrm{d}y = (\cos\theta \mathrm{d}x + \sin\theta \mathrm{d}y) \wedge \mathrm{d}\theta. \tag{2.6}$$

Since the underlying manifold is of dimension three we get

$$\begin{aligned} \Theta \wedge (\mathrm{d}\Theta)^1 &= \Theta \wedge \omega \\ &= (-\sin\theta \mathrm{d}x + \cos\theta \mathrm{d}y) \wedge (\cos\theta \mathrm{d}x + \sin\theta \mathrm{d}y) \wedge \mathrm{d}\theta \\ &= -\sin^2\theta\,(\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}\theta) + \cos^2\theta\,(\mathrm{d}y \wedge \mathrm{d}x \wedge \mathrm{d}\theta) \\ &= -\mathrm{d}x \wedge \mathrm{d}y \wedge \mathrm{d}\theta \neq 0. \end{aligned} \tag{2.7}$$

Thus $\Theta = -\sin\theta \mathrm{d}x + \cos\theta \mathrm{d}y$ is a contact form, and we conclude that $\mathcal{H}$ from (2.2) is a contact distribution. This property will be exploited below, when we introduce the geodesic equations.

2.4 Sub-Riemannian geometries

Now we are ready to define a sub-Riemannian geometry. In the rest of this chapter we follow the book "A Tour of Subriemannian Geometries, its Geodesics and Application" by R. Montgomery [12]. Proofs of any theorems are therefore omitted.

Definition 2.9. *A* ***sub-Riemannian geometry*** *is a manifold $\mathcal{M}$ endowed with a distribution together with a fiber inner product on this distribution.*

We will call the endowed distribution, $\mathcal{H}$, a *horizontal* distribution. Further, an object such as a vector field or curve will be called horizontal if it is tangent to $\mathcal{H}$.

The distribution $\mathcal{H}$ from (2.2) defines a sub-Riemannian geometry on $\mathcal{RT}$. We define the inner product between two vectors $v, w \in \mathcal{H}|_{\mathbf{x}}$ as

$$\langle v, w\rangle_{\mathbf{x}} = \langle \alpha, \beta\rangle, \tag{2.8}$$

where α and β are the coordinate-vectors of v and w with respect to $X_1(\mathbf{x})$ and $X_2(\mathbf{x})$, and $\langle\cdot,\cdot\rangle$ is the usual inner product on $\mathbb{R}^2$. Note that this inner product is defined only for horizontal vectors, otherwise it would define a degenerate norm on $\mathcal{RT}$. Considering the vector $v = \alpha X_3(\mathbf{x}) \neq 0$, $\|v\|^2_{\mathbf{x}} = \langle v, v\rangle_{\mathbf{x}} = 0$.

For any parameterized horizontal curve $\gamma(t) \in \mathcal{RT}$, the derivative, where it is defined, lies in the distribution:

$$\dot{\gamma}(t) \in \mathcal{H}|_{\gamma(t)} \quad \text{for almost every t.} \tag{2.9}$$

The length of a parameterized horizontal curve γ, denoted $\lambda(\gamma)$, can be expressed as

$$\lambda(\gamma) = \int \|\dot{\gamma}\| \mathrm{dt}, \tag{2.10}$$

where $\|\dot{\gamma}\| = \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}}$ is computed using the inner product on $\mathcal{H}|_{\gamma(t)}$, and the integral is computed over the domain of the curve.

Consequently it is possible to define the length d between two points $A, B \in \mathcal{RT}$, just as in Riemannian geometry:

$$d(A, B) = \inf\{\lambda(\gamma) : \gamma \text{ is a horizontal curve connecting } A \text{ and } B\}. \tag{2.11}$$

Note that $d(A, B)$ is infinite if no horizontal path connecting A and B exists.

2.5 Geodesics

A geodesic is basically the "shortest" curve connecting two points, or a generalization of the term "straight lines" to curved spaces.

Definition 2.10. *An absolutely continuous horizontal path that realizes the distance between two points is called a* ***minimizing geodesic*** *or simply* ***geodesic****.*

As with the horizontal paths, no geodesic connecting the points A and B exists if the length $d(A, B)$ is infinite.

Theorem 2.1. *Suppose that $\mathcal{M}$ is a connected manifold with a bracket-generating distribution. Then any two points of $\mathcal{M}$ can be joined by a horizontal path.*

This is in essence the famous theorem by Chow. Note that since a horizontal path exists between any two points, a geodesic between them also exists. The proof of this theorem is quite complex and therefore omitted, but an easy "verification" for our sub-Riemannian geometry on $\mathcal{RT}$ can be considered. First of all, we know that the distribution $\mathcal{H}$ from Section 2.2 is bracket-generating. Secondly, the manifold $\mathcal{RT} \simeq (x, y, \theta)$ is obviously connected, so the requirements of the theorem are fulfilled. Without loss of generality, we can assume a starting point $(0, 0, 0) \in \mathbb{R}^2 \times S^1$. An arbitrary point (x, y, θ) can then be reached by first following $X_2 = \partial_\theta$ to the θ-value ϕ, corresponding to the angle of the vector from $(0, 0)$ to (x, y) in $\mathbb{R}^2$. Next, follow X_1 from $(0, 0, \phi)$ to (x, y, ϕ), and finally follow X_2 again to the point (x, y, θ).

In Figure 2.1 the points $(0, 0, 0)$ and $(1, 1, 0)$ are connected using the procedure described above. If the curve is parameterized as $\gamma(t)$ for $t \in [a, b]$, then it satisfies

$$\dot{\gamma}(t) \in \operatorname{span}\{X_1(\gamma(t)), X_2(\gamma(t)\} \quad \text{for almost all } t \in [a, b]. \tag{2.12}$$

At the "problematic" points $(0, 0, \pi/2)$ and $(1, 1, \pi/2)$ the derivatives are not defined, so the curve is obviously horizontal.

Though we are not guaranteed that the curve in Figure 2.1 is a geodesic, we can conclude that $d\big((0, 0, 0), (1, 1, 0)\big)$ is finite, and thus a geodesic connecting these two points exists.

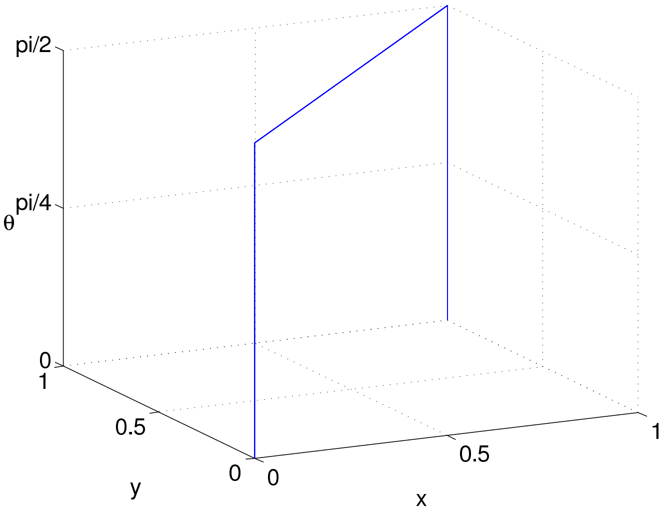


Figure 2.1: Visualization of Theorem 2.1 for the roto-translational space.

2.5.1 Geodesic equations

Geodesics can be found by minimizing the length between two points, and thus find the geodesic which realizes this minimum. A different approach is to minimize the *energy*,

$$E(\gamma) = \int_\gamma \tfrac{1}{2}\|\dot{\gamma}\|^2. \tag{2.13}$$

The curve minimizing the energy is also minimizing the length, which can be shown by the Cauchy-Schwarts inequality.

The definitions that follow are used to find a relation between the energy and the Hamiltonian function. This Hamiltonian in turn defines a system of ordinary differential equations which we will call the geodesic equations.

Definition 2.11. *A* ***cometric*** *is a section of the bundle* $S^2(T\mathcal{M}) \subset T\mathcal{M} \otimes T\mathcal{M}$ *of symmetric bilinear forms on the cotangent bundle.*

Remark. From Definition 2.5 we have that $T\mathcal{M}$ and $T^*\mathcal{M}$ are dual. This means that the cometric in Definition 2.11 defines a fiber-bilinear form (inner product on covectors) $((\cdot,\cdot)) : T^*\mathcal{M} \otimes T^*\mathcal{M} \to \mathbb{R}$. This form in turn defines a bundle map $\beta : T^*\mathcal{M} \to T\mathcal{M}$ taking covectors to tangent vectors by $\mathbf{p}(\beta_\mathbf{x}(\mu)) = ((\mathbf{p},\mu))_\mathbf{x}$ for $\mathbf{p}, \mu \in T^*\mathcal{M}|_\mathbf{x}$ and $\mathbf{x} \in \mathcal{M}$.

Definition 2.12. *The fiber-quadratic function* $H(\mathbf{x},\mathbf{p}) = \frac{1}{2}((\mathbf{p},\mathbf{p}))_\mathbf{x}$, *where* $((\cdot,\cdot))_\mathbf{x}$ *is the cometric on the fiber* $T^*\mathcal{M}|_\mathbf{x}$, *is called the* ***Hamiltonian****, or the* ***kinetic energy****.*

To relate the energy of a horizontal curve γ with the Hamiltonian, we note that $\dot{\gamma}(t) = \beta_{\gamma(t)}(\mathbf{p})$ for some covector $\mathbf{p} \in T^*\mathcal{M}|_{\gamma(t)}$. Thus

$$\tfrac{1}{2}\|\dot{\gamma}\|^2 = H(\mathbf{x},\mathbf{p}) \tag{2.14}$$

for this covector $\mathbf{p}$, and where $\mathbf{x}$ are the coordinates for $\gamma(t) \in \mathcal{M}$.

Definition 2.13. *Let X be a vector field on the manifold $\mathcal{M}$. The fiber-linear function on the cotangent bundle $P_X : T^*\mathcal{M} \to \mathbb{R}$ defined by $P_X(\mathbf{x}, \mathbf{p}) = \mathbf{p}(X(\mathbf{x}))$ is called the* ***momentum function*** *for X.*

To compute the Hamiltonian for our sub-Riemannian geometry we start with the vector fields X_1 and X_2. The momentum functions for these vector fields are

$$\begin{aligned} P_{X_1} &= \cos\theta p_x + \sin\theta p_y, \\ P_{X_2} &= p_\theta, \end{aligned} \tag{2.15}$$

where $\mathbf{p} = (p_x, p_y, p_\theta)$. Since the vector fields are orthonormal with respect to the sub-Riemannian inner product, we get an easy expression for the Hamiltonian:

$$H = \tfrac{1}{2}\left(P_{X_1}^2 + P_{X_2}^2\right). \tag{2.16}$$

In canonical coordinates, (x^i, p_i), the Hamiltonian generates the system of ordinary differential equations

$$\begin{aligned} \dot{x}^i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial x^i}. \end{aligned} \tag{2.17}$$

Definition 2.14. *The Hamiltonian differential equations* (2.17) *are called the* ***normal geodesic equations****.*

In certain geometries, there exist geodesics which do not solve the normal geodesic equations. These are called singular geodesics.

Theorem 2.2. *Any geodesic in a sub-Riemannian geometry whose distribution is a contact distribution, is normal. This means that all geodesics solve the geodesic equations above, and no so-called singular geodesics exist.*

Remark. Recalling above, $\mathcal{H}$ is a contact distribution on $\mathcal{RT}$, and only normal geodesics need to be considered. For simplicity, we will refer to (2.17) as the geodesic equations from here on.

Theorem 2.3. *Let $\xi(t) = (\gamma(t), \mathbf{p}(t))$ be a solution to the normal geodesic equations on $T^*\mathcal{M}$, and let $\gamma(t)$ be its projection to $\mathcal{M}$. Then every sufficiently short arc of γ is a minimizing sub-Riemannian geodesic. Moreover γ is the unique minimizing geodesic joining its end-points.*

In our case, a straight forward calculation yields:

$$\begin{aligned} \dot{x} &= \cos\theta\left(p_x\cos\theta + p_y\sin\theta\right), \\ \dot{y} &= \sin\theta\left(p_x\cos\theta + p_y\sin\theta\right), \\ \dot{\theta} &= p_\theta, \\ \dot{p}_x &= 0, \\ \dot{p}_y &= 0, \\ \dot{p}_\theta &= -\left(p_x\cos\theta + p_y\sin\theta\right)\left(-p_x\sin\theta + p_y\cos\theta\right). \end{aligned} \tag{2.18}$$

Remark. The projected curve $\gamma = (x, y, \theta) \in \mathcal{RT}$ is obviously horizontal since $\dot{\gamma} = (p_x \cos\theta + p_y \sin\theta)\, X_1(\gamma) + p_\theta X_2(\gamma) \in \mathcal{H}|_\gamma$.

The solution to this system of ordinary differential equations can not be expressed using elementary functions. Therefore we choose to approximate solutions using numerical techniques. If one would like to construct a geodesic, we start with boundary points A and B. The values of p_i at boundary points are free parameters. After a numerical solution is found, we project it to $(x, y, \theta) \in \mathbb{R}^2 \times S^1$, and we have an approximation of the geodesic joining A and B.

2.5.2 Example of calculated geodesic

When calculating geodesics from the set of ordinary differential equations (2.18), we use the MATLAB$^{\text{TM}}$ function "bvp4c". In addition to the equations, "bvp4c" takes as input a residual function imposing the boundary conditions and an initial guess for the solution. Further, the time interval for the curve has to be specified. Because of the specific form of (2.18), we can introduce a time scaling. Choosing the new variable $t = k\tau$, the form of the equations in $\mathcal{RT}$ stays the same. This means that the time interval is arbitrary, and the speed of the curve will adjust accordingly with $(p_x(\tau), p_y(\tau), p_\theta(\tau))$.

Figure 2.2 shows a curve calculated as described above. The end-points are $(0, 0, 0)$ and $(1, 1, 0)$ as in the example in Figure 2.1. When comparing the two figures, it is clear that the curve solving the geodesic equations has shorter length, as expected. This can also be verified with a numerical approximation of the length $\lambda(\gamma) = \int \|\dot{\gamma}(t)\| \mathrm{dt}$.

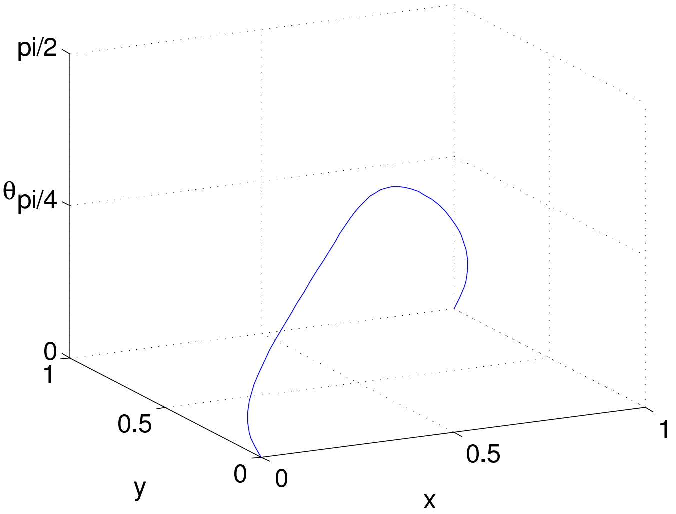


Figure 2.2: Geodesic joining $(0, 0, 0)$ and $(1, 1, 0)$ in $\mathcal{RT}$.

Chapter 3

Flows and operators

In this chapter we describe some flows and operators, and relate them to the sub-Riemannian geometry on $\mathcal{RT}$. The model of Citti and Sarti [1] involves two partial differential equations solved for a function $u(\mathbf{x}, t) : \mathcal{RT} \times [0, \infty) \to \mathbb{R}$. The initial function $u(\mathbf{x}, 0)$ is concentrated on a surface $\Sigma_0 \subset \mathcal{RT}$, and the solution $u(\mathbf{x}, t)$ defines a surface $\Sigma(t)$ for each t, as shown in Chapter 4. A sketch of proof that $\Sigma(t)$ is the mean curvature motion of Σ_0 and $u(\mathbf{x}, t)$ is the Laplace-Beltrami flow on $\Sigma(t)$ is given in [1].

We first define the sub-Laplacian and the Laplace-Beltrami operator in Section 3.1 and Section 3.2, respectively. We also find their explicit expressions in the sub-Riemannian geometry on $\mathcal{RT}$ defined in the previous chapter. As we will see, it is exactly these operators that define the two time evolutions in the model of Citti and Sarti. The relation to the mean curvature motion, as mentioned above, will become clear in the sections that follow. The mean curvature motion (MCM) problem in Euclidean and sub-Riemannian geometries will be addressed in Section 3.3 and Section 3.4, respectively. For each case, we will mention three different representation formulas for the solutions of the MCM problem.

3.1 Sub-Laplacian

The "horizontal" Laplace operator, or sub-Laplacian, is the Laplace operator in a sub-Riemannian geometry. This means that it is defined by the vector fields in a distribution. For the distribution $\mathcal{H} = \mathrm{span}(X_1, X_2)$, we denote the horizontal gradient as

$$\mathrm{grad}_R u = (X_1 u)X_1 + (X_2 u)X_2 \in \mathbb{R}^3, \tag{3.1}$$

where the subscript R indicates that it is adjusted to the sub-Riemannian geometry on $\mathcal{RT}$.

We will often use the coordinate-vector of the horizontal gradient with respect to X_1 and X_2, thus writing[1]

$$\nabla_R u = [X_1 u, X_2 u]^T \in \mathbb{R}^2. \tag{3.2}$$

Note that when finding the horizontal gradient of $u(\mathbf{x})$, we evaluate the vector fields at $\mathbf{x}$. Thus $\nabla_R = \nabla_R(\mathbf{x})$ is space-dependent, but for simplicity we ommit this dependency.

Since the vector fields X_1 and X_2 are orthonormal with respect to the inner product (2.8), we have

$$\|\mathrm{grad}_R u\|_{\mathbf{x}}^2 = (X_1 u)^2 + (X_2 u)^2 = \|\nabla_R u\|^2, \tag{3.3}$$

where $\|\cdot\|_{\mathbf{x}}$ is the norm associated $\mathcal{H}|_{\mathbf{x}}$, while $\|\cdot\|$ is the regular Euclidean norm. Because of this property, it is often favourable to use ∇_R.

---

[1] The brackets are here used to denote vectors and matrices. When Lie brackets are used, this will be clear from the context.

Further, we find the sub-Laplacian as the R-divergence of the R-gradient:

$$\Delta_R = \nabla_R^T \cdot \nabla_R = X_1(X_1) + X_2(X_2) \equiv X_1^2 + X_2^2. \tag{3.4}$$

Applying this operator to a function $u(x, y, \theta)$, we get

$$\Delta_R u = \cos^2\theta\, u_{xx} + 2\cos\theta\sin\theta\, u_{xy} + \sin^2\theta\, u_{yy} + u_{\theta\theta}, \tag{3.5}$$

which can be expressed in compact form with the Hessian matrix of u:

$$\begin{aligned}\Delta_R u &= \operatorname{tr}\left(\begin{bmatrix}\cos\theta & \sin\theta & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}u_{xx} & u_{xy} & u_{x\theta}\\ u_{xy} & u_{yy} & u_{y\theta}\\ u_{x\theta} & u_{y\theta} & u_{\theta\theta}\end{bmatrix}\begin{bmatrix}\cos\theta & 0\\ \sin\theta & 0\\ 0 & 1\end{bmatrix}\right)\\ &= \operatorname{tr}\left(\sigma(\mathbf{x})\ \mathrm{Hess}(u)\ \sigma^T(\mathbf{x})\right).\end{aligned} \tag{3.6}$$

Here $\operatorname{tr}(\cdot)$ is the trace of a matrix, and $\sigma(\mathbf{x})$ is the matrix consisting of the vector coordinates of X_1 and X_2 at the point $\mathbf{x} = (x, y, \theta)$.

Remark. In n-dimensional Euclidean space, $\sigma(\mathbf{x})$ is the $n \times n$ identity matrix. Thus $\Delta u = \operatorname{tr}(\mathrm{Hess}(u))$ is the usual Laplace operator.

Using the three-dimensional gradient (3.1) above, we get

$$\begin{aligned}\Delta_R u &= \operatorname{div}(\operatorname{grad}_R u) = \partial_x\left((X_1 u)\cos\theta\right) + \partial_y\left((X_1 u)\sin\theta\right) + \partial_\theta\left(X_2 u\right)\\ &= \partial_x\left(\cos^2\theta\, u_x + \cos\theta\sin\theta\, u_y\right) + \partial_y\left(\cos\theta\sin\theta\, u_x + \sin^2\theta\, u_y\right) + \partial_\theta u_\theta\\ &= \cos^2\theta\, u_{xx} + 2\cos\theta\sin\theta\, u_{xy} + \sin^2\theta\, u_{yy} + u_{\theta\theta},\end{aligned} \tag{3.7}$$

which is the same expression, as expected.

The modified heat equation

$$u_t = \Delta_R u \tag{3.8}$$

is one of the partial differential equations in the model [1].

3.2 Laplace-Beltrami

The Laplace operator can be generalized to operate on functions defined on surfaces or Riemannian and pseudo-Riemannian manifolds. This more general operator is called the Laplace-Beltrami operator [14].

We start with a surface $\Sigma \subset \mathbb{R}^n$ which can be identified as the zero set of a continuous function $v : \mathbb{R}^n \to \mathbb{R}$:

$$\Sigma = \{\mathbf{x} \in \mathbb{R}^n : v(\mathbf{x}) = 0\}. \tag{3.9}$$

When assuming the non-degeneracy property $\|\nabla v\| \neq 0$, we know that Σ is a regular surface, and we are guaranteed the existence of a tangent plane at every point on Σ [15]. Next, $\nu = \nabla v / \|\nabla v\|$ is a unit vector normal to this surface, and we define the projection

$$P_\nu = I - \nu \otimes \nu, \tag{3.10}$$

or, component-wise as

$$(P_\nu)_{ij} = \delta_{ij} - \nu_i \nu_j, \tag{3.11}$$

where δ_{ij} is the Kronecker delta. This matrix will project any vector at $\mathbf{x}$ onto the tangent plane of Σ at $\mathbf{x}$.

For a function u defined on Σ, its Eulerian surface gradient is

$$\nabla_\nu u = \nabla u - (\nabla u \cdot \nu)\, \nu = P_\nu \nabla u, \tag{3.12}$$

where $\cdot$ is the usual scalar product.

The Laplace-Beltrami operator can then be defined in a natural way as the tangential divergence of the tangential gradient [16]:

$$\Delta_\nu u = \nabla_\nu \cdot \nabla_\nu u = (P_\nu \nabla)^T P_\nu \nabla u = (P_\nu \nabla)^T \nabla u, \tag{3.13}$$

where the last equality follows from the fact that P_ν is an orthogonal projector and thus $P_\nu = P_\nu^T = P_\nu^2$.

3.2.1 Horizontal Laplace-Beltrami

In a sub-Riemannian geometry we have a *horizontal* Laplace-Beltrami operator, related to the distribution $\mathcal{H}$. Here we will find the horizontal Laplace-Beltrami operator for $\mathcal{H} = \text{span}(X_1, X_2)$.

Assume, as above, that we have a surface Σ defined as the zero set of v. The normal unit-vector ν was found as the normalized gradient of v. As we saw in Section 3.1, we had two expressions for the horizontal gradient. When finding scalar values involving vector products of the gradient, it is arbitrary which one we use. Reasoning as above, we can set $\nu = \nabla_R v / \|\nabla_R v\|$ to be our normal unit-vector of Σ, and P_ν a projection matrix. The horizontal Laplace-Beltrami operator, $\Delta_{\nu,R}$, applied to a function u has the following form

$$\Delta_{\nu,R} u = (P_\nu \nabla_R)^T P_\nu \nabla_R u. \tag{3.14}$$

An explicit calculation with $\nabla_R = [X_1, X_2]^T$ yields

$$\begin{aligned} \Delta_{\nu,R} u &= \left(\begin{bmatrix} \left(1 - \frac{(X_1 v)^2}{\|\nabla_R v\|^2}\right) & -\frac{X_1 v X_2 v}{\|\nabla_R v\|^2} \\ -\frac{X_1 v X_2 v}{\|\nabla_R v\|^2} & \left(1 - \frac{(X_2 v)^2}{\|\nabla_R v\|^2}\right) \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right)^T \begin{bmatrix} X_1 u \\ X_2 u \end{bmatrix} \\ &= \frac{(X_2 v)^2 X_{11} u + (X_1 v)^2 X_{22} u}{(X_1 v)^2 + (X_2 v)^2} - \frac{X_1 v X_2 v \,(X_{12} u + X_{21} u)}{(X_1 v)^2 + (X_2 v)^2}. \end{aligned} \tag{3.15}$$

Here $\|\cdot\|$ is the usual Euclidean norm and $X_{ij} u = X_i(X_j u)$. As above, we assume the non-degeneracy condition $\|\nabla_R v\| \neq 0$.

The second partial differential equation in the model [1] is

$$u_t = \Delta_{\nu,R} u. \tag{3.16}$$

Thus we find the Laplace-Beltrami flow of u on some surface Σ defined by ν.

3.3 Euclidean mean curvature motion (MCM)

A particular problem which has attained a lot of attention the past decades is the mean curvature motion problem, denoted MCM from here on. In Euclidean spaces, MCM flow of a hypersurface is a geometrical evolution such that the normal velocity at each point is equal to the mean curvature at that point. This is a rather complicated evolution, and even smooth surfaces can develop singularities in finite time. For example, a sphere evolves under mean curvature flow by shrinking uniformly to a point. Other examples include shapes which split or merge, and when this happens, the notion of curvature is undefined at the point of intersection. Thus a weak notion of the solution is necessary.

Different techniques for solving the MCM problem have been proposed. The first one we mention, which was developed independently by Evans and Spruck [17] and Chen, Giga and Goto [18], involves assigning a partial differential equation to a level set function. The solution of this partial differential equation has level sets which evolve by MCM.

The second technique is to solve the heat equation, and continuously reinitialize after small time steps [19]. In the limit, when the time step tends to zero, it solves the MCM partial differential equation mentioned above, as proved by Evans [20].

Last, we are going to mention a stochastic approach to find a representation formula for the MCM partial differential equation described in [21]. This technique is based on stochastic sampling of Brownian motions in $\mathbb{R}^n$, which can be used to approximate other partial differential equations as well [22]. When finding numerical solutions to high-dimensional partial differential equations, a stochastic approach can drastically reduce computation time.

3.3.1 MCM partial differential equation

Following Evans and Spruck [17], we here introduce a notion of generalized MCM of hypersurfaces in $\mathbb{R}^n$. Given a smooth, bounded $(n-1)$-dimensional surface $\Gamma_0 \subset \mathbb{R}^n$, find a smooth function $g : \mathbb{R}^n \to \mathbb{R}$ whose zero set is Γ_0:

$$\Gamma_0 = \{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) = 0\}. \tag{3.17}$$

The idea now is to define a function $u \in \mathbb{R}^n$ as the solution to a certain partial differential equation (PDE) which has g as initial condition. The differential equation is called the mean curvature evolution PDE:

$$\begin{aligned} u_t &= \sum_{i,j=1}^{n} \left(\delta_{ij} - \frac{u_{x_i} u_{x_j}}{\|Du\|^2} \right) u_{x_i x_j} && \text{for } \mathbb{R}^n \times (0,\infty), \\ u &= g && \text{for } \mathbb{R}^n \times \{0\}. \end{aligned} \tag{3.18}$$

It is proved in [17] that (3.18) has a unique, weak solution. Further, the level sets of u evolve by mean curvature motion in regions where u is smooth and $\|Du\| \neq 0$. The evolution of the hypersurface Γ_0 in (3.17) can consequently be found as the zero set:

$$\Gamma_t = \{\mathbf{x} \in \mathbb{R}^n : u(\mathbf{x}, t) = 0\}. \tag{3.19}$$

Viscosity solutions

The weak solution of (3.18) is called a viscosity solution. To find such a solution, we follow the definitions introduced in [17].

Definition 3.1. *A function $u \in C(\mathbb{R}^n \times [0,\infty))$ is a **weak subsolution** of* (3.18) *provided that if*

for any $\phi \in C^\infty(\mathbb{R}^n \times (0,\infty))$ such that $u - \phi$ has a local maximum at a point $(\mathbf{x}_0, t_0)$, then

$$\begin{cases} \phi_t \leq \left(\delta_{ij} - \phi_{x_i}\phi_{x_j}/\|\nabla\phi\|^2\right)\phi_{x_i x_j}, & \text{at } (\mathbf{x}_0, t_0) \quad \text{if } \nabla\phi(\mathbf{x}_0, t_0) \neq 0, \\ \phi_t \leq \left(\delta_{ij} - \eta_i\eta_j\right)\phi_{x_i x_j}, & \text{at } (\mathbf{x}_0, t_0) \quad \text{if } \nabla\phi(\mathbf{x}_0, t_0) = 0, \end{cases}$$

where, in the last case, η is some vector in $\mathbb{R}^n$, with $\|\eta\| \leq 1$.

Definition 3.2. *A function $u \in C(\mathbb{R}^n \times [0,\infty))$ is a **weak supersolution** of* (3.18) *provided that if*

for any $\phi \in C^\infty(\mathbb{R}^n \times (0,\infty))$ such that $u - \phi$ has a local minimum at a point $(\mathbf{x}_0, t_0)$, then

$$\begin{cases} \phi_t \geq \left(\delta_{ij} - \phi_{x_i}\phi_{x_j}/\|\nabla\phi\|^2\right)\phi_{x_i x_j}, & \text{at } (\mathbf{x}_0,t_0) \quad \text{if } \nabla\phi(\mathbf{x}_0,t_0) \neq 0, \\ \phi_t \geq \left(\delta_{ij} - \eta_i\eta_j\right)\phi_{x_i x_j}, & \text{at } (\mathbf{x}_0,t_0) \quad \text{if } \nabla\phi(\mathbf{x}_0,t_0) = 0, \end{cases}$$

where, in the last case, η is some vector in $\mathbb{R}^n$, with $\|\eta\| \leq 1$.

Definition 3.3. *A function $u \in C(\mathbb{R}^n \times [0,\infty)) \cap L^\infty(\mathbb{R}^n \times [0,\infty))$ is a **weak solution** of* (3.18) *if u is both a weak subsolution and a weak supersolution.*

From the definitions 3.1 - 3.3, we see that weak solutions to the MCM problem exists even if singularities arise.

3.3.2 Approximating MCM by diffusion

Bence, Merriman and Osher [19] have proposed a computational algorithm for tracking the evolution in time of a set in $\mathbb{R}^n$ whose boundary moves by mean curvature motion. Given a compact set $C_0 \subset \mathbb{R}^n$ we solve the heat equation

$$\begin{aligned} u_t &= \Delta u && \text{in } \mathbb{R}^n \times (0,\infty), \\ u(\mathbf{x},0) &= \begin{cases} 1 \text{ if } \mathbf{x} \in C_0 \\ 0 \text{ if } \mathbf{x} \notin C_0 \end{cases} && \text{on } \mathbb{R}^n \times \{t=0\}. \end{aligned} \tag{3.20}$$

For a time $t > 0$, we define the set

$$C_t = \left\{\mathbf{x} \in \mathbb{R}^n : u(x,t) \geq \tfrac{1}{2}\right\}. \tag{3.21}$$

In [19], the authors show that the evolution C_t approximates the MCM for the boundary Γ_0 of C_0 for small times, when Γ_0 is smooth. Let us now replace the initial condition in (3.20) with the indicator function of C_t, and solve the heat equation again. By continuously reinitializing over short time steps, this should yield an approximation to MCM flow even for large times.

Let $\mathcal{H}(t)C_0$ denote the reinitializing procedure (3.20, 3.21) applied to an initial compact set C_0. Then denote $\mathcal{M}(t)C_0$, the solution of the MCM evolution PDE (3.18) with the indicator function of C_0 as initial data. Then the following is proved by Evans [20]:

$$\lim_{m\to\infty} H(t/m)^m g = M(t)g, \tag{3.22}$$

where H and M are the operators $\mathcal{H}$ and $\mathcal{M}$ applied to *functions* with initial value g, respectively.

So by fixing $m \gg 1$, and applying the renormalizing diffusion scheme (3.20, 3.21) m times, we get an approximation of the solution of the MCM problem at time t.

3.3.3 Stochastic representation formula for MCM

Another approach to solve the MCM problem, is to consider a stochastic optimal control problem. The following theory follows a paper by Buckdahn, Cardaliaguet and Quincampoix [21].

For a problem in $\mathbb{R}^n$, define a control set $\mathcal{V}$ as

$$\mathcal{V} = \left\{v \in \mathcal{S}_n : v \geq 0, I - v^2 \geq 0 \text{ and } \operatorname{tr}(I - v^2) = 1\right\}, \tag{3.23}$$

where $\mathcal{S}_n$ is the set of symmetric $n \times n$ matrices.

Now fix a $T > 0$. For a $v(s) \in \mathcal{V}\ \forall\ s$, solve the stochastic differential equation:

$$\begin{cases} dX^{\mathbf{x},t,v(\cdot)}(s) = \sqrt{2}v(s)dW(s) & \text{for } s \in (t,T], \\ X^{\mathbf{x},t,v(\cdot)}(t) = \mathbf{x}. \end{cases} \tag{3.24}$$

Here $W(s)$ is an n-dimensional Brownian motion where for each component i, $W_i(s+t) - W_i(s)$ is normally distributed with expectation 0 and standard deviation $\sqrt{t}$. This means that each control $v(s)$ defines a path in $\mathbb{R}^n$ starting from $\mathbf{x}$ at time t denoted $X^{\mathbf{x},t,v}(s)$ for $s \in [t,T]$.

The following theorem is proved in [21]:

Theorem 3.1. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a bounded uniformly continuous function. Let $T > 0$ be fixed, and let us, for any initial position $(\mathbf{x},t) \in \mathbb{R}^2 \times [0,T]$, set*

$$V(\mathbf{x},t) = \inf_{v \in \mathcal{V}} \left(\textit{ess-sup}_{\Omega} g \left(X^{\mathbf{x},t,\nu(\cdot)}(T) \right) \right),$$

where $X^{\mathbf{x},t,\nu(\cdot)}(\cdot)$ is the solution to (3.24)*. Then V is the solution, in the viscosity sense, of the MCM evolution PDE* (3.18) *(written here with a terminal condition):*

$$\begin{cases} V_t &= -\sum_{i,j=1}^{n} \left(\delta_{ij} - \frac{V_{x_i}V_{x_j}}{\|DV\|^2} \right) V_{x_i x_j} & \textit{for } \mathbb{R}^n \times (0,\infty), \\ V &= g & \textit{for } \mathbb{R}^n \times \{T\}. \end{cases}$$

From this representation formula, it is clear that approximations to the solution of (3.18) can be found by stochastic sampling. Since we do not know the optimal control $\nu(\cdot) \in \mathcal{V}$ though, a good way to approximate the infimum over every such $\nu(\cdot)$ must be found.

3.4 Sub-Riemannian MCM

As with the operators mentioned above, the MCM problem can also be defined in sub-Riemannian geometries. The horizontal curvature is naturally expressed using the horizontal gradient. Further, a similar MCM evolution PDE (3.18) can be derived using vector fields instead of the usual partial differentiation.

Contrary to the Euclidean case, the sub-Riemannian MCM problem is a fairly new research topic and little is known about the solutions of these. In addition to singular points where the gradients vanish, we can encounter so-called characteristic points. These are points where the Euclidean gradient is orthogonal to the horizontal space, thus the horizontal gradient vanishes at both singular and characteristic points.

In this section we are going to treat the explicit sub-Riemannian geometry on $\mathcal{RT}$ defined by the distribution $\mathcal{H} = \operatorname{span}(X_1, X_2)$.

3.4.1 Horizontal MCM partial differential equation

From (3.18), it is easy to show that the MCM partial differential equation is equivalent to

$$u_t = \operatorname{tr}(\operatorname{Hess}(u)) - \left\langle \operatorname{Hess}(u) \frac{\nabla u}{\|\nabla u\|}, \frac{\nabla u}{\|\nabla u\|} \right\rangle, \tag{3.25}$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^n$.

Since the vector fields X_1 and X_2 do not commute, the Hessian matrix related to our sub-Riemannian geometry has to be symmetrized. With the notation $X_{ij} = X_i(X_j)$, we define the horizontal Hessian of a function u as

$$\operatorname{Hess}_R(u) = \begin{bmatrix} X_1^2 u & \frac{1}{2}(X_{12}u + X_{21}u) \\ \frac{1}{2}(X_{12}u + X_{21}u) & X_2^2 u \end{bmatrix}. \tag{3.26}$$

As we pointed out in Section 3.1, we do not have to modify the inner product $\langle\cdot,\cdot\rangle$ when using the gradient ∇_R. We find the following MCM evolution PDE for our sub-Riemannian geometry:

$$\begin{aligned} u_t &= \operatorname{tr}\left(\operatorname{Hess}_R(u)\right) - \left\langle \operatorname{Hess}_R u \frac{\nabla_R u}{\|\nabla_R u\|}, \frac{\nabla_R u}{\|\nabla_R u\|} \right\rangle \\ &= X_1^2 u + X_2^2 u - \frac{1}{\|\nabla_R u\|^2} \left\langle \begin{bmatrix} X_1^2 u & \frac{1}{2}\left(X_{12}u + X_{21}u\right) \\ \frac{1}{2}\left(X_{12}u + X_{21}u\right) & X_2^2 u \end{bmatrix} \begin{bmatrix} X_1 u \\ X_2 u \end{bmatrix}, \begin{bmatrix} X_1 u \\ X_2 u \end{bmatrix} \right\rangle \\ &= X_1^2 u + X_2^2 u - \frac{1}{\|\nabla_R u\|^2} \left(X_1^2 u (X_1 u)^2 + X_2^2 u (X_2 u)^2 + (X_{12}u + X_{21}u) X_1 u X_2 u \right) \\ &= \sum_{i,j=1}^{2} \left(\delta_{ij} - \frac{X_i u X_j u}{\|\nabla_R u\|^2} \right) X_{ij} u. \end{aligned} \tag{3.27}$$

We see that the sub-Riemannian MCM evolution PDE has the same form as (3.18), where the partial differentiations $\partial/\partial x_i$ are replaced by the vector fields X_i. Further, the upper summation index is lowered from the dimension of our problem to the number of vector fields in our distribution. We also notice that in addition to singular points, characteristic points where $\nabla u \perp \operatorname{span}(X_1, X_2)$ will produce a zero denominator.

3.4.2 Approximating horizontal MCM by horizontal diffusion

In Section 3.3.2 we saw that solutions to the MCM evolution PDE could be approximated by solving the heat equation for short time steps and continuously reinitialize. In the model of Citti and Sarti [1], one of the equations is the horizontal heat equation $u_t = \Delta_R u$. After this equation is solved for a small time step, the time-evolution is set to evolve according to the Laplace-Beltrami operator described in Section 3.2.1. This will play the role of the reinitialization procedure (3.21). Then these equations are solved alternately. A sketch of proof of the convergence to the MCM flow of the initial surface is given in [1], and it can be completed using nonlinear semigroup theory as in the paper of Evans [20] mentioned above.

3.4.3 Stochastic representation formula for horizontal MCM

The stochastic approximation of MCM can also be generalized to sub-Riemannian geometries as described in the relatively new article by Dirr, Dragoni and Renesse [23].

As in Section 3.1, we denote the 2×3 matrix $\sigma(\mathbf{x})$ as the matrix consisting of the vector coordinates of $X_1(\mathbf{x})$ and $X_2(\mathbf{x})$ as rows. As argued in [23], we have to replace the Itô by a Stratonovich stochastic differential equation (see [24] for details) which reflects the fact that we do not work in an Euclidean space:

$$\begin{cases} dX^{\mathbf{x},t,\nu(\cdot)}(s) = \sqrt{2}\,\sigma^T\left(X^{\mathbf{x},t,\nu(\cdot)}(s)\right) \circ \nu(s) dW(s), & s \in (t, T], \\ X^{\mathbf{x},t,\nu(\cdot)}(t) = \mathbf{x}, \end{cases} \tag{3.28}$$

where $\circ\nu(s)dW(s)$ denotes the integral with respect to $\nu(s)W(s)$ in the sense of Stratonovich. As in (3.24), ν is an element in the control set $\mathcal{V}$ of $n \times n$ matrices. In our case $n = 2$, which is the dimension of $\mathcal{H}$. Further, W is a 2-dimensional Brownian motion and $X^{\mathbf{x},t,\nu(\cdot)}(s)$ defines a *horizontal* path in $\mathbb{R}^2 \times S^1$ for each control.

Note that by Itô's formula, replacing $\sigma(\mathbf{x})$ with the identity matrix will produce the same stochastic differential equation (3.24).

In [23] it is shown that the value function in Theorem 3.1, modified to the stochastic differential equation above, solves the horizontal MCM evolution PDE (3.27) in the viscosity sense.

Chapter 4

A perceptual completion model

In this chapter we are going to describe the model of Citti and Sarti [1]. Their paper is titled "A Cortical Based Model of Perceptual Completion in the Roto-Translation Space", which refers that the model is inspired by the functionality of the Visual Cortex. As mentioned in Chapter 3, the model consists of two horizontal partial differential equations.

In Section 4.1 we briefly describe how the model is inspired by psychological and biological evidence. The initialization of the model with an image-input is described in Section 4.2, and Section 4.3 is devoted to the flows which evolves the image.

Finally, we discuss numerical approximations and boundary conditions in Section 4.4 and Section 4.5, respectively.

4.1 Introduction

By an occlusion we here mean holes, objects blocking other objects, or other situations resulting in loss of visual information. A good example is the "blind spot" in the human visual field. The blind spot corresponds to the lack of light-detecting photoreceptor cells on the optic disc of the retina where the optic nerve passes through it [9]. In everyday life though, the blind spots are rarely noticed. This is because the human brain fills in the blind spot using information in the surroundings and information from the other eye, and this first part is what we are going to model.

The Visual Cortex consists of several layers or areas, and the layer where information is first processed is called V1. V1 consists of simple cells which have different properties. Among these are direction, polarity and elongation [10]. The orientation is related to brightness gradients, and each simple cell alter their orientation after what input arrives. It is then natural to associate a circle of orientations, S^1, above each point $(x, y) \in \mathbb{R}^2$. When there is "missing information", the cells which represent this part of the visual field choose their direction according to neighbouring cells. This is called local association fields [25].

Remark. For more discussions about the relation to psychological and biological models, we refer to [1] and [11].

4.2 Extraction of existing information

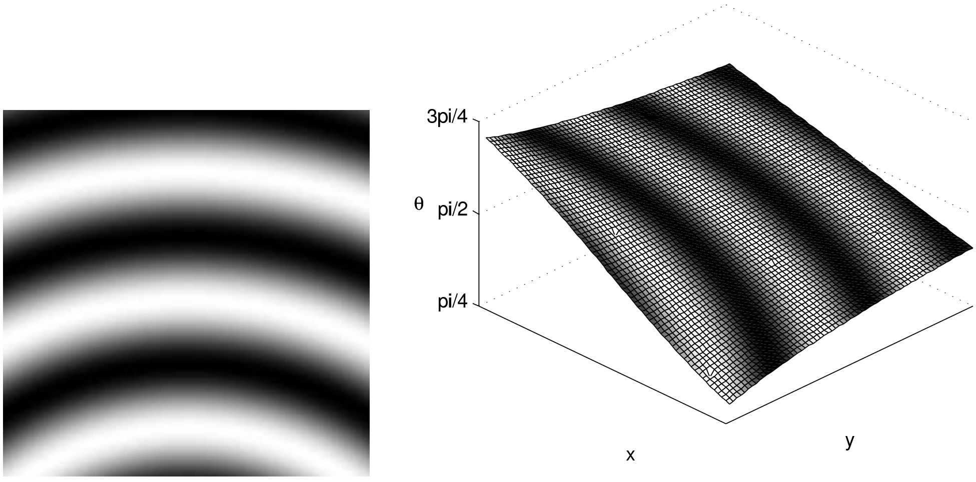
Assume $I : (x, y) \to \mathbb{R}$ a grayscale image containing some sort of occlusion. Following [1], we "lift" I to three dimensions:

$$u_0(x, y, \theta) = \sin\theta I_x - \cos\theta I_y = -X_3(\theta)I, \tag{4.1}$$

where X_3 is the vector field on $\mathcal{RT}$ orthogonal to the distribution $\mathcal{H}$ in Section 2.2. Let Σ_0 be defined as

$$\Sigma_0 = \left\{ (x, y, \bar{\theta}) : |X_3(\bar{\theta})I| = \max_{\theta} |X_3(\theta)I| \right\}. \tag{4.2}$$

Σ_0 is now a surface in $\mathcal{RT}$, or an image-oriented manifold. This initialization process is related to the functionality of the simple cells in V1. The “height”, or orientation θ is chosen to attain maximum activity in I.



(a) Original image.

(b) Lifted image.

Figure 4.1: Image lifted in $\mathcal{RT}$.

Figure 4.1 shows an example of a lifted image. We see that the height is related to the gradient direction at each point.

Remark. Some information about the original image is lost with this initialization procedure, since u_0 is constructed by taking directional derivatives of $I(x, y)$. The image in Figure 4.1 (b) is the original image lifted to Σ_0, and not the corresponding u_0 constrained on Σ_0.

4.3 Completion

The next step is to evolve the function $u_0\delta_{\Sigma_0}$, the function u_0 constrained on the surface Σ_0 above. The evolutions are found using differential equations on the sub-Riemannian geometry defined by the distribution $\mathcal{H} = \text{span}(X_1, X_2)$. The equations are the sub-Laplacian

$$u_t = \Delta_R u = X_1^2 u + X_2^2 u, \tag{4.3}$$

and the Laplace-Beltrami flow

$$u_t = \Delta_{\nu,R} u = (P_\nu \nabla_R)^T \cdot \nabla_R u. \tag{4.4}$$

These two time evolutions are applied in an alternating fashion. First we diffuse with the sub-Laplacian for a time t_1, then evolve according to the Laplace-Beltrami flow for a time t_2. Then we do these two processes repeatedly N times, resulting in a total time $N(t_1 + t_2)$.

The explicit expression of the sub-Laplacian for our sub-Riemannian geometry is shown in (3.5). The Laplace-Beltrami operator is a bit more complex as it is dependent on a surface whose normal vector is ν.

Remember that we started off with the function u_0 constrained on the surface Σ_0. After the diffusion process, this surface has expanded since the sub-Laplacian includes the vector field $X_2 = \partial_\theta$. We now have a "brick"-like shape in $\mathcal{RT}$. As in the initialization process, we here model the simple cells functionality by choosing the orientation θ which results in the maximum activity in u. Following this approach, let the new surface at time t_1 be defined as

$$\Sigma(t_1) = \left\{ (x, y, \bar{\theta}) : u(x, y, \bar{\theta}, t_1) = \max_{\theta} |u(x, y, \theta, t_1)| \right\}. \tag{4.5}$$

Assuming a smooth u, we know that for each spatial point $(x, y) \in \mathbb{R}^2$, the maximum along θ of u occurs where $\partial_\theta u = 0$ and $\partial_\theta^2 u < 0$. By assuming only maxima, $\Sigma(t_1)$ can be expressed as

$$\Sigma(t_1) = \left\{ (x, y, \theta) : v = 0, X_2 v \neq 0 \right\}, \quad v = X_2 u. \tag{4.6}$$

From [15], we know that $\Sigma(t_1)$ is a regular surface, and a normal vector exists at each point on this surface. Reasoning as in Section 3.1, we can construct our normal vector as

$$\nu = \frac{\nabla_R v}{\|\nabla_R v\|} = \frac{(X_1 v, X_2 v)}{\sqrt{(X_1 v)^2 + (X_2 v)^2}}. \tag{4.7}$$

The expression of the Laplace-Beltrami operator applied to u on this surface is

$$\begin{aligned} (P_\nu \nabla_R)^T \cdot \nabla_R u &= \frac{1}{\|\nabla_R v\|^2} \left(\begin{bmatrix} (X_2 v)^2 & -X_1 v X_2 v \\ -X_1 v X_2 v & (X_1 v)^2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right)^T \begin{bmatrix} X_1 u \\ X_2 u \end{bmatrix} \\ &= \frac{X_{11} u (X_2 v)^2 + X_{22} u (X_1 v)^2}{(X_1 v)^2 + (X_2 v)^2} - \frac{(X_{12} u + X_{21} u)\, X_1 v X_2 v}{(X_1 v)^2 + (X_2 v)^2}, \end{aligned} \tag{4.8}$$

where P_ν is calculated as in (3.10), and $X_{ij} u = X_i (X_j u)$.

Post processing

After a certain time t', we can build a resulting image from the function $u(x, y, \theta, t')$. The procedure we will use in the results of Chapter 7 is the following:

$$I(x, y) = \max_{\theta} |u(x, y, \theta, t')|. \tag{4.9}$$

4.4 Discretization

We denote u^n_{lmq} the function u evaluated at the discrete node $(x, y, \theta) = (l\Delta x, m\Delta y, q\Delta\theta)$ at time $t = n\Delta t$. Digital images are already discretized, due to its pixel representation. Thus for an input image, l and m must be in the intervals $[0, ..., L-1]$ and $[0, ..., M-1]$, respectively, where $L \times M$ is the pixel resolution. The third dimension θ can be discretized by dividing the interval $[0, 2\pi)$ into Q equidistant points, so $q \in [0, ..., Q-1]$.

The various derivatives are approximated as:

$$\begin{aligned}
D_1 u^n_{lmq} &= \cos\theta_q \delta_x u^n_{lmq} + \sin\theta_q \delta_y u^n_{lmq},\\
D_2 u^n_{lmq} &= \delta_\theta u^n_{lmq},\\
D_{11} u^n_{lmq} &= \cos^2\theta_q \delta^2_x u^n_{lmq} + 2\sin\theta_q \cos\theta_q \delta_x \delta_y u^n_{lmq} + \sin^2\theta_q \delta^2_y u^n_{lmq},\\
D_{22} u^n_{lmq} &= \delta^2_\theta u^n_{lmq},\\
D_{12} u^n_{lmq} &= \cos\theta_q \delta_x \delta_\theta u^n_{lmq} + \sin\theta_q \delta_y \delta_\theta u^n_{lmq},\\
D_{21} u^n_{lmq} &= -\sin\theta_q \delta_x u^n_{lmq} + \cos\theta_q \delta_x \delta_\theta u^n_{lmq} + \cos\theta_q \delta_y u^n_{lmq} + \sin\theta_q \delta_y \delta_\theta u^n_{lmq},
\end{aligned} \tag{4.10}$$

where $D_i u$ is the approximation of $X_i u$ and $D_{ij} u$ is the approximation of $X_i(X_j u)$. Further, we approximate the partial derivatives of u with second-order accurate centred finite differences. For example

$$\begin{aligned}
\delta_y u^n_{lmq} &= \frac{1}{2\Delta_y}\left(u^n_{lm+1q} - u^n_{lm-1q}\right),\\
\delta^2_x u^n_{lmq} &= \frac{1}{\Delta^2_x}\left(u^n_{l+1mq} - 2u^n_{lmq} + u^n_{l-1mq}\right).
\end{aligned} \tag{4.11}$$

An explicit Euler method is used as time integrator, which means that we approximate the time derivatives with a first-order accurate forward difference. The alternating combined scheme is then:

```
FOR $n = 0, ..., N_1 - 1$, DO:
    $u^{n+1}_{lmq} = u^n_{lmq} + \Delta t \left(D_{11} u^n_{lmq} + D_{22} u^n_{lmq}\right)$
END FOR
$v^{N_1}_{lmq} = D_2 u^{N_1}_{lmq}$
FOR $n = N_1, ..., N_1 + N_2 - 1$, DO:
    $u^{n+1}_{lmq} = u^n_{lmq} + \Delta t \Big[ \frac{(D_2 v^n_{lmq})^2 D_{11} u^n_{lmq} + (D_1 v^n_{lmq})^2 D_{22} u^n_{lmq}}{(D_1 v^n_{lmq})^2 + (D_2 v^n_{lmq})^2} - \frac{D_1 v^n_{lmq} D_2 v^n_{lmq} \left(D_{12} u^n_{lmq} + D_{21} u^n_{lmq}\right)}{(D_1 v^n_{lmq})^2 + (D_2 v^n_{lmq})^2} \Big]$      (4.12)
    $v^{n+1}_{lmq} = D_2 u^{n+1}_{lmq}$
END FOR.
```

We see that after every sub-Laplacian diffusion step (the first for-loop), we have a function v that defines the new surface as in (4.6). v is found using the numerical approximation of $X_2 u$.

This scheme will diffuse according to the sub-Laplacian for a time $N_1 \Delta t$, followed by the Laplace-Beltrami flow for a time $N_2 \Delta t$. Then this scheme is iterated N times.

Remark. In [11], the first part of the scheme (4.12), the horizontal heat equation, has been analysed more extensively. An upper bound on the time step Δt has been found applying the Courant-Friedrich-Lewy-condition. Numerical experiments imply that this is a conservative estimate, valid also for the second part of the scheme. Further, different combinations of the parameters N_1 and N_2 have been compared. Though small variations occur in test images, we will use the same combination for all the results in Chapter 7.

4.5 Boundary conditions

4.5.1 Spatial boundary conditions

Stated in [1], Neumann boundary conditions should be used. In [11], homogeneous Neumann boundary conditions were first used by imposing

$$\partial u/\partial \vec{n} = 0 \tag{4.13}$$

on the boundaries. Here $\vec{n}$ is the outward normal direction of the boundaries. This produces a change of character in images near the boundaries. Level lines are "straightened out", which seems natural since we impose, for example, $u_x = 0$ at the left boundary.

Next, a different approach is considered. Let us replace the outward normal vector with the vector $(\cos\theta, \sin\theta)$, which corresponds to the two first vector coordinates of X_1. This leads to a "roto-translational" Neumann boundary conditions:

$$X_1 u = \cos\theta u_x + \sin\theta u_y = 0 \quad \text{along the boundaries.} \tag{4.14}$$

This gives better results, as can be seen in Chapter 7 where we compare the two different boundary conditions.

4.5.2 Boundary conditions in the θ-direction

In the θ-direction, we use periodic boundary conditions. With a discretized θ-dimension, periodic boundary conditions can be imposed by using the fictive nodes

$$\begin{aligned} u^n_{l,m,Q} &= u^n_{l,m,0}, \\ u^n_{l,m,-1} &= u^n_{l,m,Q-1}, \end{aligned} \tag{4.15}$$

in the expressions of $\delta_\theta u$ and $\delta_\theta^2 u$.

Chapter 5

Minimal surfaces in $\mathcal{RT}$

In this chapter we will describe the model of Hladky and Pauls [2]. This model, which is based on [1], is more of a "brute-force" model. Instead of completing an entire image with the diffusion process described in the previous chapter, the content of a circular hole is reconstructed by creating a minimal surface which spans the lifted boundary of the hole. As stated in [1], their surface $\Sigma(t) \subset \mathcal{RT}$ converges to a minimal surface when $t \to \infty$, in the sense that the horizontal mean curvature is zero.

We start by introducing the basic aspects of the model in Section 5.1. In Section 5.2 we explain how the model is initialized, by lifting the circular boundary to $\mathcal{RT}$. We also classify points with certain properties on this lifted boundary. Points on the lifted boundary shall be connected, and the criterion for connectability will be defined in Section 5.3. In Section 5.4 we construct the curves which span the minimal surface. Finally, we describe how an image can be completed with this method, and we make some concluding remarks in Section 5.5 and Section 5.6, respectively.

5.1 Introduction

Let us assume that a grayscale image I contains a hole D which is an open circle. The values on ∂D are prescribed by Dirichlet boundary conditions. The basic idea of the model is the following. We "lift" the circular boundary ∂D to $\mathcal{RT}$. Then we find all points on this lifted circle which are connectable by rules which we define below. Next, rules between every pair of connectable points are constructed, and these rules span a minimal surface as described in [2]. When projecting down to $\mathbb{R}^2$ again, we have contour-curves covering D, and the image is completed.

The model is closely related to the sub-Riemannian geometry on $\mathcal{RT}$. Remembering from above, we called curves horizontal if they lie in the distribution $\mathcal{H} = \operatorname{span}(X_1, X_2)$. The curves we are going to consider in this model are horizontal curves with constant curvature. The authors of [2] call these horizontal curves (and connected subsets of them) *rules*, and we will adapt this notation. Rules are helix-cords, or when projected to $\mathbb{R}^2$, circle-cords.

Rules can be found explicitly as follows. Let $\mathbf{x}(t) = \exp(t, X_1 + kX_2)(x_0, y_0, \theta_0)$. That is, $\mathbf{x}(t)$ is the flow map of the linear combination $X_1 + kX_2$ applied to the initial point $(x_0, y_0, \theta_0) \in \mathcal{RT}$. In local coordinates, $\mathbf{x}(t)$ is the solution of the system of ordinary differential equations $\dot{x}(t) = \cos\theta(t)$, $\dot{y}(t) = \sin\theta(t)$, $\dot{\theta}(t) = k$. With $k \neq 0$, we get:

$$\begin{aligned} x(t) &= x_c + R\sin\theta(t), \\ y(t) &= y_c - R\cos\theta(t), \\ \theta(t) &= \theta_0 + kt, \end{aligned} \tag{5.1}$$

where $R = 1/k$, $x_c = x_0 - R\sin\theta_0$ and $y_c = y_0 + R\cos\theta_0$. If $k = 0$, we get the straight line,

$$\begin{aligned} x(t) &= x_0 + (\cos\theta_0)\, t, \\ y(t) &= y_0 + (\sin\theta_0)\, t, \\ \theta(t) &= \theta_0. \end{aligned} \tag{5.2}$$

Below, we will discuss how ∂D is lifted, and how we find points which are connectable by rules. Some theorems and lemmas from [2] are stated, and the omitted proofs of any theorems can be found there.

5.2 Configuration

The algorithm exploits the setup in [1]. Denote the parameterized curve which corresponds to the boundary of the hole, ∂D, by $\beta(t)$. Since we assume circular holes, we can express β as

$$\beta(t) = (x_0 + R\cos t, y_0 + R\sin t)\,, \quad \text{for } t \in [0, 2\pi). \tag{5.3}$$

In three dimensions, we denote the lifted boundary $\gamma(t) = (\beta(t), \theta(t)) \in \mathcal{RT}$, where $\theta(t)$ corresponds to the normal gradient direction of the image I along $\beta(t)$. θ can be found as follows:

$$\theta(t) = \arctan\left(-\frac{I_x \circ \beta(t)}{I_y \circ \beta(t)}\right), \tag{5.4}$$

where the branch of arctan is chosen to get a continuous $\theta(t)$. This corresponds to the lifting procedure (4.2) in the previous chapter.

In addition to θ, we define ϕ_β as

$$\frac{\dot\beta(t)}{\|\dot\beta(t)\|} = (-\sin\phi_\beta(t), \cos\phi_\beta(t))\,. \tag{5.5}$$

Because of the specific parameterization chosen for the circular boundary, we see that we can identify $\phi_\beta(t) = t$.

We now have two angle functions. $\theta(t)$ is related to ∇I along the circle $\beta(t)$, and $\phi_\beta(t)$ corresponds to the normal gradient of the circle $\beta(t)$ itself. The algorithm uses these angle functions to identify points with different properties, and they are also vital in the process of finding points which can be connected with rules. Define the transversality function Q as

$$Q(t) = \theta(t) - \phi_\beta(t). \tag{5.6}$$

With the simplification above, we get $Q(t) = \theta(t) - t$. The transversality function "measures" the angle between $\dot\beta$ and ∇I.

Example. Let us consider the intensity function $I(x, y) = x - y$. This gives a constant ∇I, and $\theta(t) = \pi/4\ \forall t$. In Figure 5.1 we see a contour plot of the intensity function, together with the circle β centred at $(0, 0)$ with radius 1.

The figure also shows the points where the transversality function is a multiple of $\pi/2$. The red squares correspond to the points where $Q(t) = k\pi$, and the blue correspond to $Q(t) = \pi/2 + k\pi$, where $k \in \mathbb{Z}$. At the blue squares, the contours are parallel to $\dot\beta$, that is ∇I is orthogonal to $\dot\beta$.

Definition 5.1. *A point $p = \gamma(t)$ such that $\dot\gamma(t) \in \mathcal{H}|_{\gamma(t)}$, where $\mathcal{H} = span\{X_1, X_2\}$ is the contact distribution from* (2.2)*, is called* ***Legendrian****. Furthermore, a point $p = \gamma(t)$ such that $\dot\gamma(t) \in span\{X_2(\gamma(t)), X_3(\gamma(t))\}$ is called* ***orthogonal****.*

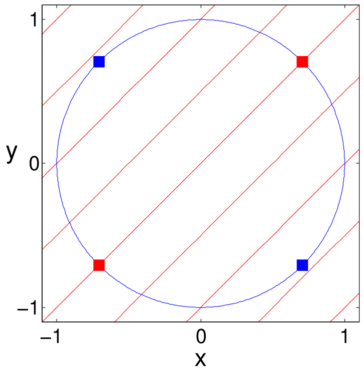


Figure 5.1: Countour plot of $I(x, y) = x - y$ together with the circle $\beta(t)$.

Lemma 5.1. *The Legendrian points of γ occur precisely where*

$$Q(t) = \pi/2 + k\pi, \quad k \in \mathbb{Z}.$$

The orthogonal points occur precisely where

$$Q(t) = k\pi, \quad k \in \mathbb{Z}.$$

Proof. From the parameterization of β (5.3) we get

$$\dot{\gamma}(t) = \frac{\partial}{\partial t}(x_0 + R\cos t, y_0 + R\sin t, \theta(t)) = -R\sin t\frac{\partial}{\partial x} + R\cos t\frac{\partial}{\partial y} + \dot{\theta}\frac{\partial}{\partial \theta}.$$

With the explicit expressions of X_1, X_2 and X_3, together with the trigonometric addition formulas we get

$$\dot{\gamma}(t) = R\sin(\theta - t)X_1 + \dot{\theta}X_2 + R\cos(\theta - t)X_3,$$

and this completes the proof. ☐

5.3 Finding points to connect

We will now use the transversality function, together with the Lemma above, to find points on γ which we can connect with rules.

When considering a point p, we define the accessible set, $\mathcal{A}(p)$, as all points which can be connected from p by a rule. $\mathcal{A}(p)$ is given by the implicit equation

$$\frac{y - y_0}{x - x_0} = \tan\left(\frac{\theta + \theta_0}{2}\right), \tag{5.7}$$

where $p = (x_0, y_0, \theta_0)$. This is easy to prove using trigonometric identities together with the expressions of the rules (5.1, 5.2).

The points we are interested in lie entirely on γ, the lifted circular boundary. We denote by $\mathcal{A}(p, \gamma)$ the set of points on γ which can be connected from p, that is $\mathcal{A}(p, \gamma) = \gamma \cap \mathcal{A}(p)$.

Lemma 5.2. *For $p = \gamma(t)$, the non-trivial part of $\mathcal{A}(p, \gamma)$ is given implicitly by*

$$Q(t) + Q(u) = (2k + 1)\pi, \quad k \in \mathbb{Z}.$$

Proof. From (5.7) we see that $\gamma(u) = (x, y, \theta) \in \mathcal{A}(\gamma(t), \gamma)$ if and only if

$$\frac{y - y_0}{x - x_0} = \frac{y_c + R\sin u - (y_c + R\sin t)}{x_c + R\cos u - (x_c + R\cos t)} = \frac{\sin u - \sin t}{\cos u - \cos t} = \tan\left(\frac{\theta(u) + \theta(t)}{2}\right).$$

Further, the differences in the fraction can be simplified:

$$\frac{\sin u - \sin t}{\cos u - \cos t} = \frac{-\cos((u+t)/2)\sin((u-t)/2)}{\sin((u+t)/2)\sin((u-t)/2)} = \frac{-\cos((u+t)/2)}{\sin((u+t)/2)}.$$

Denoting $a = (\theta(u) + \theta(t))/2$ and $b = (u + t)/2$, we get

$$\begin{aligned} 0 &= \cos a \cos b + \sin a \sin b = \cos(a - b) \\ &\Rightarrow a - b = (Q(t) + Q(u))/2 = (2k+1)\pi/2, \end{aligned}$$

which completes the proof. □

Remark. Any point can be connected to itself with the trivial zero-radius rule.

Using Lemma 5.2, we can find $\mathcal{A}(p, \gamma)$ for all points p on the circular hole. With certain restrictions, we are also guaranteed that the accessible set $\mathcal{A}(p, \gamma)$ only contains one non-trivial point $\gamma(u)$ for each $p = \gamma(t)$. In this way we have a set of pairs which we can connect with rules.

Definition 5.2. *We call a completion problem on the form above* ***non-degenerate*** *if β and the region β encloses contain only a finite number of critical points for I, and the critical points on β are such that the θ-function in* (5.4) *can be extended continuously through these points. Further, a problem is called* ***completely non-degenerate*** *if no critical point of I lies on β.*

Theorem 5.3. *Suppose we have a completely non-degenerate occlusion problem as above, where there are no critical points inside D. If $Q'(t) \neq 0$ for all $t \in [0, 2\pi]$ then there exist a minimal spanning surface of γ denoted Σ, and any rule connecting two points on γ projects to a curve in the interior of D. Moreover, if $Q'(t) < 0$ for all $t \in [0, 2\pi]$, the projection of Σ to D is surjective.*

The proof can be found in [2]. The results in Chapter 7 will not satisfy the strict requirements of this theorem. However, below we will show an example of a problem where this theorem applies.

With the stronger criterion in Theorem 5.3 that $Q'(t) < 0$, we can verify that $\mathcal{A}(\gamma(t), \gamma)$ only consists of one non-trivial point $\gamma(u)$. In this way we can construct a function $u(t) : S^1 \to S^1$. We find the derivative of the expression in Lemma 5.2 implicitly with the chain rule:

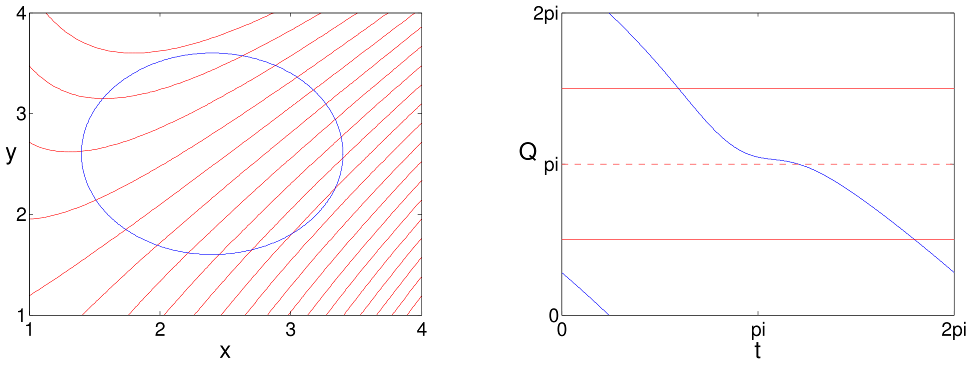
$$\begin{aligned} \frac{\partial}{\partial t}\left(Q(t) + Q(u(t))\right) &= \theta'(t) - 1 + \theta'(u)u'(t) - u'(t) = 0 \\ \Rightarrow u'(t) &= -\frac{\theta'(t) - 1}{\theta'(u) - 1} = -\frac{Q'(t)}{Q'(u)}. \end{aligned} \tag{5.8}$$

This means that $u'(t) < 0$ and u is a strictly decreasing function. Further, u is one-to-one and onto as stated in [2].

Example. An easy example is presented in Figure 5.2. A contour plot of $I(x,y) = x(y-x)$ is shown together with a circle of radius 1 representing the boundary of a hole in (a). The parameterized circle starts at $(2.4+1, 2.6)$ and is oriented counterclockwise. We notice that no critical points are on or inside this boundary.

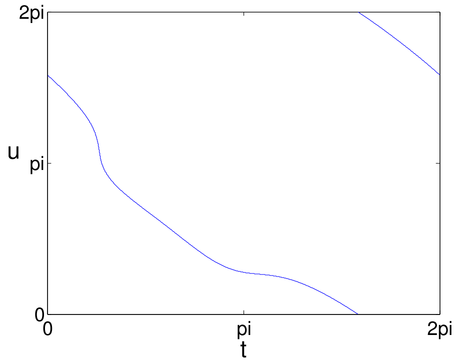
In (b), we see a plot of the transversality function $Q(t)$ for this problem. The points where Q crosses the red lines at $\pi/2$ and $3\pi/2$ are Legendrian points (Lemma 5.1). The points where Q is a multiple of π represented as the red dashed line together with the upper and lower boundaries, are orthogonal. The Legendrian and orthogonal points can easily be observed in (a) where the contour lines are parallel and orthogonal to the circle, respectively.

Since there are no critical points involved, and since $Q(t)$ is strictly decreasing (when neglecting the modulo 2π), we know that Theorem 5.3 applies. In the last figure (c) we see the function $u(t)$ constructed using Lemma 5.2. $u(t)$ is one-to-one and maps S^1 onto S^1, as expected.



(a) Contours of $I = x(y-x)$ and a circular hole centered at $(2.4, 2.6)$ with radius 1.

(b) The transversality function $Q(t)$.



(c) Plot of $u(t)$ using Lemma 5.2.

Figure 5.2: Example problem: Contours of I and the functions $Q(t)$ and $u(t)$.

Remark. In the general case, the accessible set $\mathcal{A}(p,\gamma)$ can contain more than one nontrivial point. This means that the implicit plot of $Q(t) + Q(u) = (2k+1)\pi$, as in figure (c) above, fails to be a graph over either t or u.

5.4 Constructing minimal surfaces

Now we are ready to construct a minimal surface. When assuming that Theorem 5.3 applies, we can construct the function $u(t)$. Since u is one-to-one, there are only two Legendrian points located at some parameter values t_1 and t_2. When connecting the points $\gamma(t)$ to $\gamma(u(t))$, it is enough to consider the interval $t \in T = (t_1, t_2)$. By symmetry, $u(T) \in S^1 \setminus T$, and every non-Legendrian point has been connected to another.

When connecting the points $\gamma(t)$ and $\gamma(u(t))$, the rule connecting them are found by solving

$$\begin{bmatrix} 1 & 0 & \sin\theta(t) \\ 0 & 1 & -\cos\theta(t) \\ 1 & 0 & \sin\theta(u(t)) \\ 0 & 1 & -\cos\theta(u(t)) \end{bmatrix} \begin{bmatrix} x_s \\ y_s \\ R_s \end{bmatrix} = \begin{bmatrix} x_0 + R\cos t \\ y_0 + R\sin t \\ x_0 + R\cos u(t) \\ y_0 + R\sin u(t) \end{bmatrix} \tag{5.9}$$

for x_s, y_s and R_s. Here x_0, y_0 and R are the parameters for the circular hole, and θ is the lifting defined in (5.4). From the solution (x_s, y_s, R_s), we construct the rule

$$\begin{bmatrix} x(\tau) \\ y(\tau) \\ z(\tau) \end{bmatrix} = \begin{bmatrix} x_s + R_s\sin\tau \\ y_s - R_s\cos\tau \\ \tau \end{bmatrix}, \tag{5.10}$$

where we use the time variable τ to distinguish it from the parameter t.

Proposition 5.4. *The curve* (5.10) *will pass through the two points* $\gamma(t)$ *and* $\gamma(u(t))$ *as a helix.*

Proof. From Lemma 5.2 we know that the two points are connectable by a rule, which is a helix, thus the matrix equation (5.9) has a unique solution (x_s, y_s, R_s). Further, the solution is given by

$$\begin{aligned} x_s &= x_0 + R\cos t - R_s\sin\theta(t), \\ y_s &= y_0 + R\sin t + R_s\cos\theta(t), \\ R_s &= \frac{R\,(\cos u(t) - \cos t)}{\sin\theta(u(t)) - \sin\theta(t)} = -\frac{R\,(\sin u(t) - \sin t)}{\cos\theta(u(t)) - \cos\theta(t)}. \end{aligned}$$

Inserting this into (5.10), we get

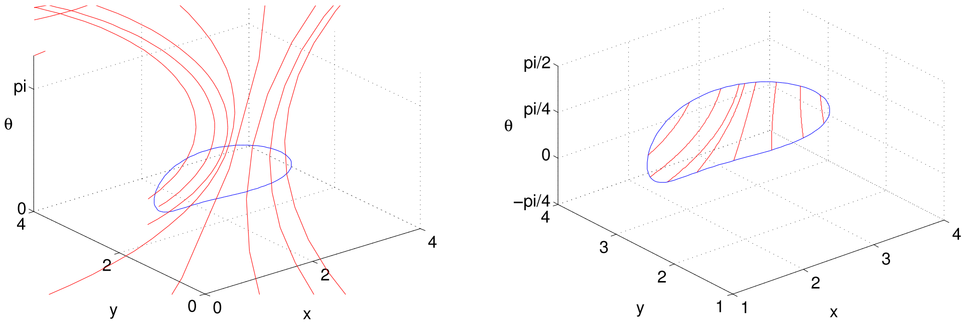
$$\begin{bmatrix} x(\theta(t)) \\ y(\theta(t)) \\ z(\theta(t)) \end{bmatrix} = \begin{bmatrix} x_0 + R\cos t \\ y_0 + R\sin t \\ \theta(t) \end{bmatrix} = \gamma(t), \qquad \begin{bmatrix} x(\theta(u(t))) \\ y(\theta(u(t))) \\ z(\theta(u(t))) \end{bmatrix} = \begin{bmatrix} x_0 + R\cos u(t) \\ y_0 + R\sin u(t) \\ \theta(u(t)) \end{bmatrix} = \gamma(u(t)),$$

which completes the proof. □

Next, the curve is restricted, as in the proof of Proposition 5.4 above, by setting the parameter for the solution curve $\tau \in [\theta(t), \theta(u(t))]$. When Theorem 5.3 applies, we are also guaranteed that the projected solution lies inside D.

Example. Figure 5.3 shows an example of rules obtained by solving (5.9). The problem is the same as in the example of Figure 5.2 described above.

Remark. When the accessible set $\mathcal{A}(p, \gamma)$ contains more than one non-trivial point, rules from p to all these points are constructed. Some points will then have multiple connections, and projected rules can cross each other. In Chapter 7, we will only consider cases where no projected rules are allowed to cross. For some problems, $u = u(t) : S^1 \to S^1$ even if the requirements of Theorem 5.3 are violated.



(a) Full rules (b) Part of rules inside γ.

Figure 5.3: Example showing the lifted boundary γ together with rules.

5.5 Completion

When applying the model above to a grayscale digital image, many approximations have to be considered. First of all, a circle cannot be realized in a pixel-based image. Next, the rules will not match the pixels either. In this section we will briefly describe how the results in Chapter 7 are found.

In all the results we start with a continuous intensity function rather than an image. A chosen circle is lifted, and a finite number of rules are constructed. Then we assign each pixel value in an image to be the intensity function at discrete points (x_i, y_i). The pixels which represent points inside the circular hole shall now be replaced using the rules we have constructed. For each discrete point (x_i, y_i) inside the hole, we find the projected rule which passes closest, and this pixel value is set to the Dirichlet value this rule represents.

Note that when we construct rules, the intensities $I(x, y)$ are never considered. The connections are found exclusively based on the gradient direction of I and the circular hole β. This can lead to ambiguous curves when considering the Dirichlet boundary condition; the two connectable points $\gamma(t)$ and $\gamma(u(t))$ can have different boundary values.

5.6 Conclusion

The assumption of circular holes is related to existence and uniqueness aspects. When finding the connection $t \sim u$ from Lemma 5.2, the authors of [2] describe a number of cases ranging from cases where one can always find such a u, cases with multiple u and cases where no u exists. The model can be generalized to arbitrary closed curves, but much of the theory above will become significantly more complex. For instance, $\phi_\beta(t)$ (5.5) can not be identified with t, and connectable points can not be found using Lemma 5.2.

Secondly, we assume an underlying continuous intensity function $I(x, y)$. If we adjust the model to only consider a digital image input, many problems can occur. The lifting θ must be approximated, and the lifted curve will be discrete. The discrete version of the implicit equation $Q(t) + Q(u) = (2k + 1)\pi$ (Lemma 5.2) may have no solutions. This problem can be overcome by interpolating the discrete points of the lifted boundary γ. For simplicity however, we will always consider the "continuous" problem here.

When comparing this model with the model described in the previous chapter, it is clear that the latter is more suited for arbitrary images. It assumes no extra knowledge about the image, in particularly no information about the situation of any hole. On the other hand, the model [2] is significantly less computationally intensive.

Chapter 6

Completion using geodesics

In Chapter 5, a model to complete circular holes was presented. Making rules or horizontal curves with constant curvature, was the essential ingredient in that model. In this chapter we are going to exploit much of the same theory, but instead of calculating rules, we are going to construct geodesics that solve the geodesic equations (2.17).

The model is based on finding points with equal boundary values, and the transversality function Q (5.6) is helpful when considering circular holes. However, arbitrary holes whose boundaries can be parameterized can also be considered.

In Section 6.1 we will explain the motivation of this model. In Section 6.2 we describe how we find points on the boundary which shall be connected by geodesic. Problems with both circular and arbitrary holes are discussed, together with some example figures. The procedure of completing an image when the geodesics are found is explained in Section 6.3, together with a conclusion.

6.1 Introduction

By the Chow theorem (Theorem 2.1), we know that any two points in $\mathcal{RT}$ can be connected by horizontal curves. Thus the length between any two points is finite, and a geodesic connecting them exists. By considering circular holes, as in the model described in the previous chapter, any point on the lifted circle can be connected to any other point on the lifted circle by a minimizing geodesic. This is not what we want to do. First of all this would produce ambiguous curves; the two end-points can have completely different boundary values. We want to produce geodesics which represent specific values, and when projected to $\mathbb{R}^2$ are completion-contours. Secondly, geodesics between arbitrary points on the boundary could produce curves which go outside D, the domain of the hole. In a worst case scenario, a projected geodesic can live exclusively outside D.

6.2 Connecting points

Let us denote by $\beta(t)$ the closed boundary of a hole. It is topologically equivalent to a circle, and we can parameterize it over S^1. The boundary is lifted to $\gamma(t) = (\beta(t), \theta(t)) \in \mathcal{RT}$, where $\theta(t)$ (5.4) is the lifting procedure as in the previous chapter. For each point on β a certain Dirichlet boundary value f is given:

$$f(t) = I(x, y) \circ \beta(t). \tag{6.1}$$

When only assuming grayscale images, we have that $f(t) : S^1 \to \mathbb{R}$. Further, for an underlying function $I \in C^1(\mathbb{R}^2)$, we have that $f(t) \in C^1(S^1)$ and both $f(t)$ and $f'(t)$ are periodic.

As described above, we are interested in connecting points whose boundary values are equal. For arbitrary problems, this is not straight forward. Considering a point $\beta(t)$, there may be many other points which have the same boundary value $f(t)$. Thus we make the restriction that a point $\beta(t)$ can only be connected to one other. Further, this choice can not be arbitrary. When every point has been connected to another, none of the geodesics connecting them can, when projected to $\mathbb{R}^2$, cross each other. We collect this in a definition below, stating when we, for our model, have a completed surface.

Definition 6.1. *Let every point $\gamma(t_1^i)$, whose boundary value $f(t_1^i)$ is not a global maximum or minimum of f, be connected to one other point $\gamma(t_2^i)$ with a geodesic ξ^i. Then a surface is completed if and only if*

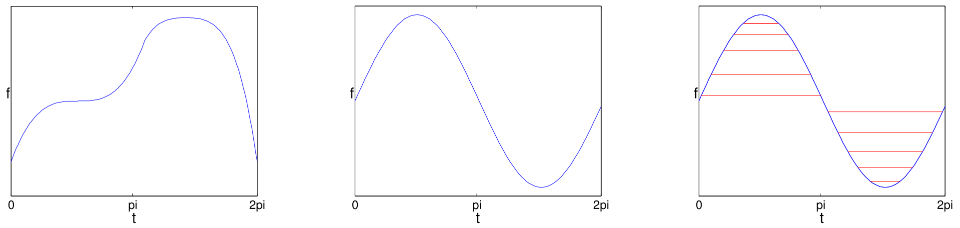
$$\begin{aligned}
(i) &\quad f(t_1^i) = f(t_2^i) \quad \forall\, i,\\
(ii) &\quad \textit{no geodesics } \xi^i,\ \xi^j \textit{ projected to } \mathbb{R}^2 \textit{ cross each other,}\\
(iii) &\quad \textit{every projected } \xi^i \textit{ lives inside } D, \textit{ the region } \beta \textit{ encloses.}
\end{aligned}$$

When $f(t)$ has values which appear in pairs, the completion process is rather easy. For every point, there is one other (unique) point where f has the same value and we construct the geodesic connecting them.

Proposition 6.1. *Let f be non-constant and smooth. If $f'(t)$ is zero at two points, the values $f(t)$ appear in pairs elsewhere.*

Proof. The proof follows from the fact that both f and f' are periodic. □

Proposition 6.1 can also be verified by observing some possible cases in Figure 6.1. In (a) we see a function with one saddle point and one global maximum. At the end point, the derivative is not defined when assuming f periodic. Since we also have periodicity in f', this case can not occur when building our boundary value function from a smooth $I(x, y)$. Further, local minima or maxima or two saddle points can not be present either. This means that the non-constant f has one global maximum and one global minimum as in (b). In (c) a visualization of the pairing is shown. The extremal points are the only ones which cannot be connected to another.

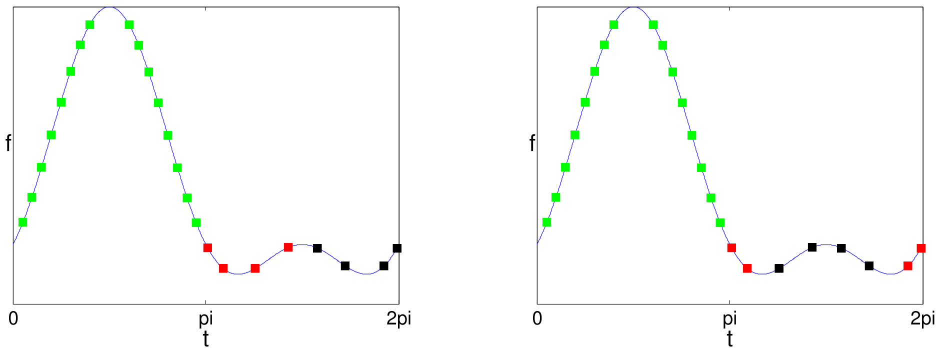


(a) Periodic f, non-periodic f'. (b) Periodic f and f'. (c) $f(t)$ appear in pairs.

Figure 6.1: Example verifying Proposition 6.1.

As we saw above, the case where f' has two zeros is trivial. We consider all points on an interval between the two extremal points, and match them with f on the other interval. By symmetry, every point has been connected to another.

When f' has more than two zeros, a trial-and-error procedure is used. Consider the case in Figure 6.2 below. f' has four zeros, and we see two different scenarios in (a) and (b). These figures show coloured squares representing the connection between points. The green squares are situated at the same places for both cases, as there are no alternatives. For $t \in [\pi, 2\pi]$ there are two alternatives, represented by the red and black squares in the figures. Now we construct the completed surface for both these cases. Afterwards we compare them, and hopefully one of the results is preferable.



(a) A combination of connections.

(b) A different combination of connections.

Figure 6.2: Connections when $f'(t) = 0$ at more than two points.

6.2.1 Circular holes

We will here consider a circular boundary $\beta(t)$. We can now exploit some of the theory from Chapter 5. First of all, if Theorem 5.3 is satisfied, we can conclude that only two points on the lifted boundary are Legendrian. Further, there are only pairs of points which can be connected by rules (5.1). The next lemma relates this to our boundary value function $f(t)$.

Proposition 6.2. *Consider a problem where Theorem 5.3 applies. Then there are two Legendrian points on $\gamma(t)$, and $f'(t) = 0$ only at these two points.*

Proof. Since $u(t)$ from (5.8) is one-to-one from S^1 onto S^1, we know that there are two points where $u(t)$ crosses the diagonal $u = t$. At these points $Q(t) + Q(u) = 2Q(t) = (2k+1)\,\pi$, for $k \in \mathbb{Z}$, thus they are Legendrian by Lemma 5.1. Further, let

$$f'(t) = \partial/\partial t \; I \circ \beta(t) = \; - R\sin(t) I_x \circ \beta(t) + R\cos(t) I_y \circ (t) = 0,$$

where R is the radius of the circle β. Since there are no critical points on β, there are no points where $I_x \circ \beta(t) = I_y \circ \beta(t) = 0$ simultaneously. Then we can insert the expression above into the transversality function (5.6) and get

$$Q(t) = \arctan\left(-\frac{I_x \circ \beta(t)}{I_y \circ \beta(t)}\right) - t = \arctan\left(-\frac{1}{\tan(t)}\right) - t = \pi/2 \mod \pi.$$

So every point where $f'(t) = 0$ is a Legendrian point, thus there can only be two zeros of $f'(t)$. $\square$

There is an important consequence of this proposition. For a problem where Theorem 5.3 from the previous chapter applies, we are also guaranteed that Proposition 6.1 above applies. This means that when a given problem is "trivial" using the model of Hladky and Pauls [2], then it is also "trivial" using this model.

6.2.2 Arbitrary holes

Let $\beta(t)$ be the parameterized boundary enclosing the hole D. Assuming a smooth underlying function I, the boundary function $f(t) = I \circ \beta$ will be smooth even if β is not. In this way, we can consider squares, triangles or other complex boundaries.

When dealing with arbitrary boundaries, we cannot apply the theory from Chapter 5. We have to rely on Proposition 6.1 or go with the trial-and-error procedure described above. Either way, we find points with matching boundary values. The boundary is lifted to $\mathcal{RT}$, and the points are connected with geodesics.

6.3 Completion

When all pairs of connectable points $\beta(t_1)$ and $\beta(t_2)$ are found, they are lifted to $\gamma(t_1), \gamma(t_2) \in \mathcal{RT}$ according to (5.4). These are now the boundary conditions for the geodesic equations (2.18). Numerical solutions $\psi(\tau) = (\xi(\tau), \mathbf{p}(\tau)) \in \mathbb{R}^6$ are found as described in the example of Section 2.5.2. From Theorem 2.3, the solution of the geodesic equations projected from $T(\mathcal{RT})^*$ to $\mathcal{RT}$ is the unique geodesic joining its endpoints. The projection $\xi(\tau) = (x(\tau), y(\tau), \theta(\tau)) \in \mathcal{RT}$ is then a numerical approximation to the unique geodesic joining $\gamma(t_1)$ and $\gamma(t_2)$. When only the completed image is of interest, this solution is again projected to $(x(\tau), y(\tau)) \in \mathbb{R}^2$. Remember that every such projected geodesic has a value $f(t_1)\,(= f(t_2))$ associated to it.

When finding a completed image using this method, the same procedure as described in Section 5.5 is used where we replace the rules with our geodesics.

Chapter 7

Numerical results

In this chapter we will present results from the three models described above. For ease of notation, we will denote CS and HP the models of Citti and Sarti (Chapter 4) and Hladky and Pauls (Chapter 5), respectively. The model in Chapter 6 will be denoted the geodesic model.

We start by comparing the two different boundary conditions for the CS model in Section 7.1. Then we compare results for some example images in Section 7.2, where we mainly consider the HP model and the geodesic model. Figures showing the results in three dimensions will also be included, and we compare some contour plots with the original ones. In Section 7.3 an easy example using the geodesic model on a rectangular hole is presented. Finally, in Section 7.4 we show a few examples which can only be completed with the CS model. These are images where objects are occluding other objects, and no defined "hole" is present.

7.1 Boundary conditions for the CS model

The numerical scheme in the CS model involves several parameters, for example the number of iterations associated to each of the two equations and the step sizes of spatial-, θ- and time dimensions. For experiments regarding these parameters, we refer to [11]. We are here merely interested in the visual key properties of completed images, and we will therefore only refer to the number of iterations we apply the numerical scheme (4.12).

In Figure 7.1 we see an original image with a hole in the centre.

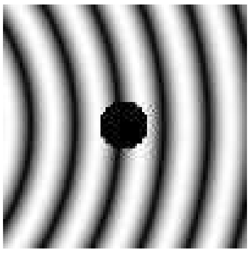
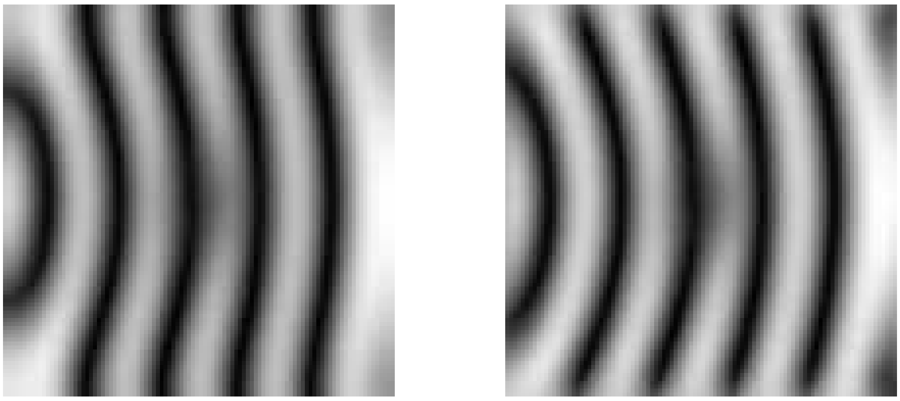


Figure 7.1: Original image used to compare boundary conditions for the CS model.

Figure 7.2 shows the completed images with (a) regular- and (b) "roto-translational" Neumann boundary conditions after 80 iterations. We see that level-lines of the completed image with regular Neumann boundary conditions are "straightened out" near the edges. Applying additional iterations will make this "error" even more prominent. By imposing the other boundary conditions, the character near the edges seems to be more correct. The completed hole in the centre looks equal for both cases. This means that when a hole in an image is relatively small, and when only the missing part of the image is interesting, it is arbitrary which boundary conditions we impose. When the hole is relatively large though, more iterations are necessary to complete it. In this case the "roto-translational" boundary conditions have to be imposed to prevent the reflection from the boundaries to distort inside the hole. In the rest of the results we impose the "roto-translational" boundary conditions when using the CS model.



(a) Regular Neumann boundary conditions.

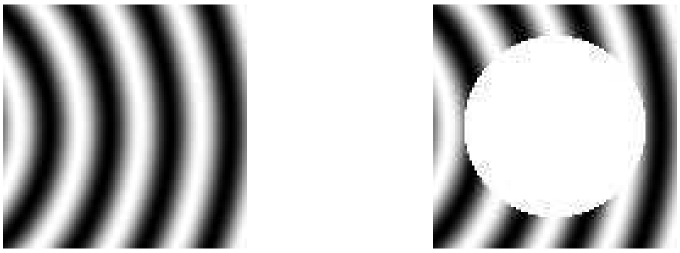
(b) "Roto-translational" Neumann boundary conditions.

Figure 7.2: Example showing the difference between the two boundary conditions in the CS model.

7.2 Comparisons of test images

7.2.1 Example 1

The first example we are going to consider is the image on the left of Figure 7.3, with an underlying intensity function $I(x, y) = \sin \sqrt{x^2 + y^2}$. The image is constructed by setting each pixel value equal to I at discrete points in the range $x \in [4\pi, 12\pi]$ and $y \in [-4\pi, 4\pi]$. In Section 7.1 above we saw results when applying the CS model to a similar image, so we will here consider the other two models. On the right of the same figure, we see the image with a circular hole centred at $(8\pi, 0)$ with radius 3π.



(a) Original image.

(b) Original with hole.

Figure 7.3: Example 1: Original image and original with circular hole.

In the HP model, the process of finding points to connect is based on the transversality function (5.6). Figure 7.4 shows the transversality function Q for the image above when the branch of arctan from (5.4) is chosen to get a continuous θ along the circular hole.

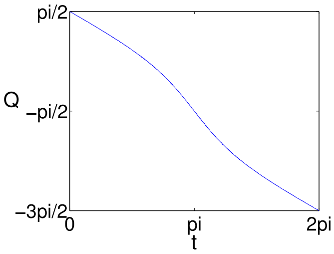
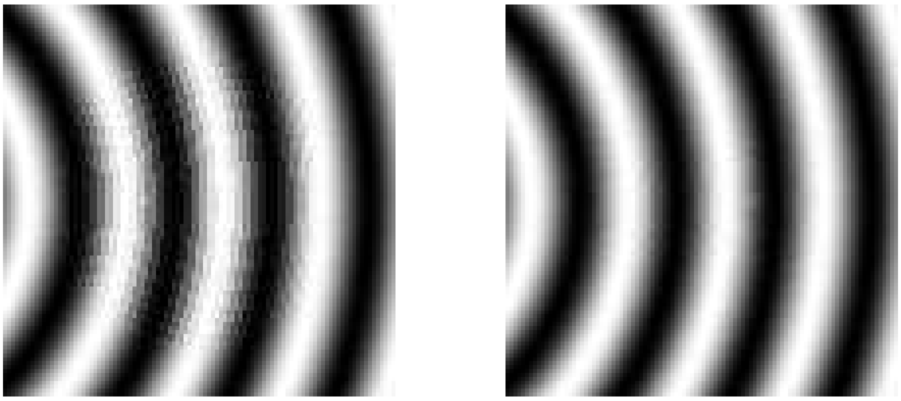


Figure 7.4: The transversality function Q for example 1.

We note that there are two Legendrian points (Lemma 5.1) at $t = 0$ and $t = \pi$, and that $Q'(t) < 0$. Further, two points denoted by their parameter value t and u on the circular boundary shall be connected if $Q(t) + Q(u) = (2k + 1)\pi$, for a $k \in \mathbb{Z}$. In this case we get an easy expression for points which are connectable, namely $u = u(t) = 2\pi - t$. This means that the connectable points are symmetric around the x-axis, which is natural when considering the original image. Even though there are critical points everywhere where $\cos\sqrt{x^2 + y^2} = 0$, we have a unique minimal surface.

In Figure 7.5 we see the result after applying the HP model with (a) 50 rules and (b) 200 rules (the resolution is 100 × 100 pixels). Using 50 rules is not adequate, but the result using 200 rules looks identical to the original. A reason that justifies the success using the HP model is that the original image consists of circular contours.



(a) 50 rules. (b) 200 rules.

Figure 7.5: Example 1: Completed using the HP model.

For the geodesic model, connecting points is based on the boundary value function f. For this example, f will oscillate between -1 and 1 because of the underlying intensity function. When only considering f, there are therefore several different combinations we could have used. However, we will exploit additional information we have, namely the situation of the Legendrian points. Since there are only two of them, we can start with either one. Then we follow the circular boundary a short distance in either direction, and find the boundary value here. When searching from the Legendrian point in the other direction, we find a point with matching boundary value, and in this specific example it is the symmetric point equally distanced from the Legendrian point we started from. Following this procedure all the way to the other Legendrian point, every point has been connected to another.

In Figure 7.6 we see the boundary value function f (6.1) together with line specifications indicating the connections as described above. For instance, a point on one of the solid green lines shall be connected to a point on the other solid green line. We see that the connection process starts at $t = 0$ or $t = \pi$, which correspond to the Legendrian points. Note that elsewhere, the points where $f' = 0$ corresponds to critical points of I on the circular boundary. The symmetric pattern can be recognized in the figure, and will produce exactly the same pairing as the HP model for the same image.

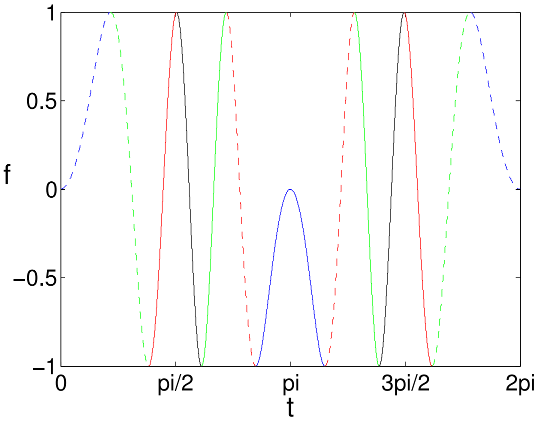


Figure 7.6: Boundary function f for example 1. Equal line specification indicate connections.

Figure 7.7 shows the result by constructing geodesics between points as explained above. 200 geodesics are used. We see that the image is continuous, but we do not get the desired behaviour inside the hole. The contours are approximately vertical in the middle. We know that geodesics in $\mathcal{RT}$ are the shortest curves connecting two points when measuring length as in (2.11). It seems like these curves are "too short".

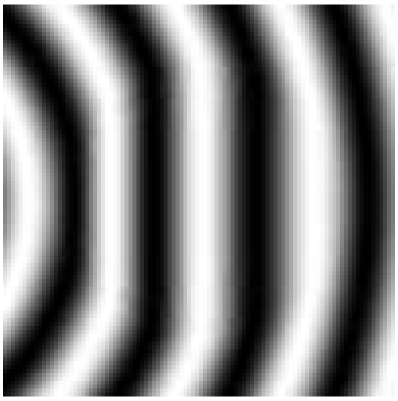


Figure 7.7: Example 1: Completed using the geodesic model.

In Figure 7.8 we see the calculated geodesics in three dimensions. In the centre, parts of the geodesics lie at the same height with a θ-value near $\pi/2$. At this height the vector $\vec{X}_1$ points in the y-direction, and this is the reason for the vertical contours in the completed image.

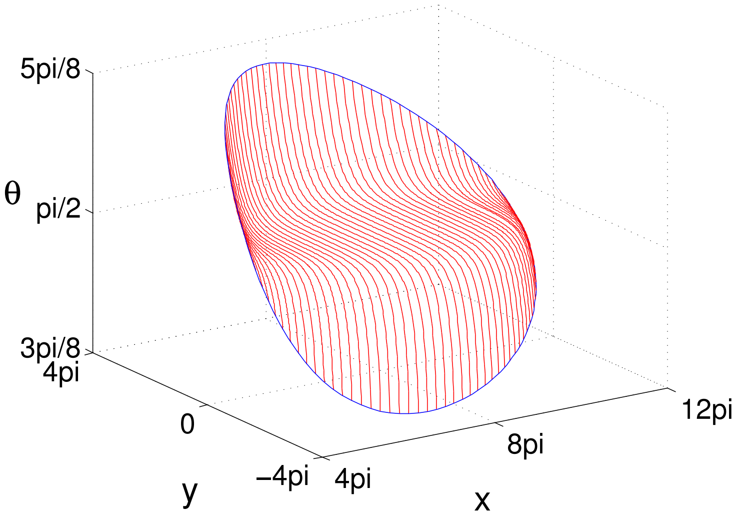
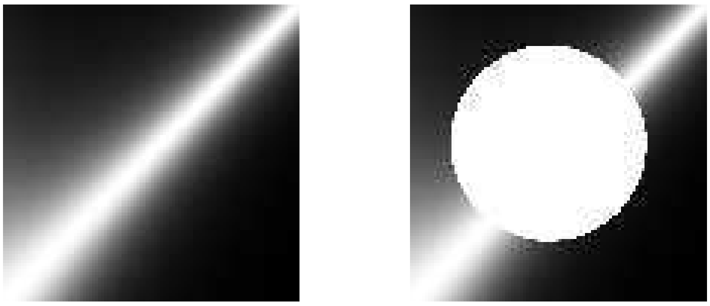


Figure 7.8: 50 geodesics in three dimensions for example 1.

7.2.2 Example 2

The next example we are going to consider is the image on the left of Figure 7.9. It is constructed from the intensity function $I(x, y) = 1/(1 + x^2(y - x)^2)$ with range $x, y \in [1, 4]$. On the right, we see the original image with a hole centred at $(2.4, 2.6)$ with radius 1.



(a) Original. (b) Original with hole.

Figure 7.9: Example 2: Original image and original with hole.

The completed image using the HP model is shown in Figure 7.10 below. We see that the white diagonal is preserved, and the overall features look similar to the original image. Above the diagonal in the upper right corner we have some distortions though. It looks as if the rules ending up in this region have wrong Dirichlet boundary values compared to the colours outside the hole.

Figure 7.10: Completed using the HP model.

The boundary values for this example are approximately between 0 and 1. Figure 7.11 shows the rules, or helix cords, in three dimensions. Further, the rules are coloured green if the two points they connect have boundary values which differs by more than 0.1, or about 10% of the maximum difference. Remember that the criterion of two points being connectable was related to the image gradient along the boundary, so differences in boundary values will occur. A slight improvement can be observed if we let each helix represent the average boundary value of the two points they connect.

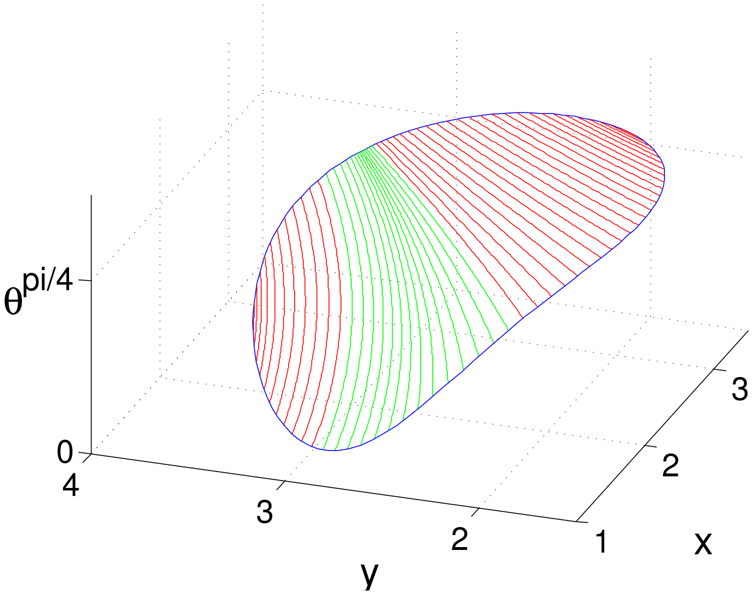


Figure 7.11: Rules constructed with the HP model. Green colour shows the curves whose end points have boundary values which differ by more than 0.1.

The geodesic model, which were designed to connect points with equal boundary values, will now be considered. In Figure 7.12, the boundary value function is shown.

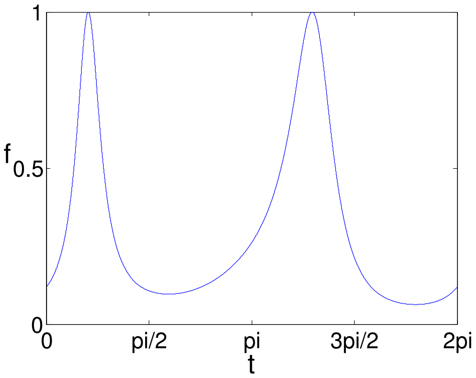
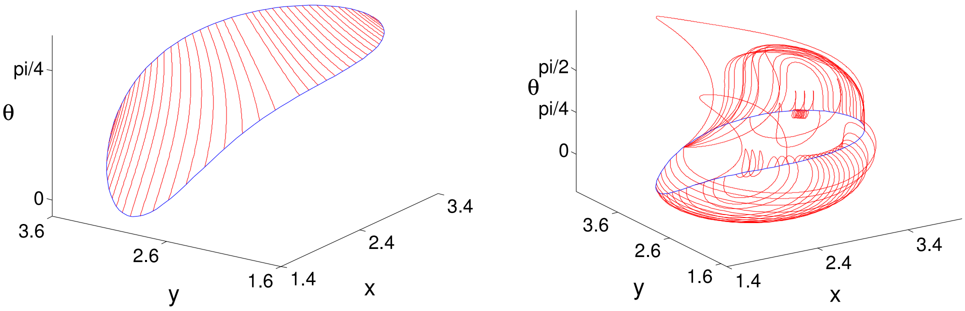


Figure 7.12: Boundary value function f for example 2.

There are four points where $f'(t) = 0$, and there are two different combinations of connecting points when no projected geodesics between them are allowed to cross each other. One possible combination is to connect points on each side of the two "valleys". The other possibility is to connect the points on each side of the "peaks". Below, we will show results using both these combinations.

Figure 7.13 shows how the calculated geodesics look like for the two different combination of connecting points. In the first case, when we connect points in the two "valleys", we see a systematic pattern and the geodesics constitute a surface spanning the hole. This is not the case when connecting points on the two "peaks", where the geodesics have strange behaviours. In addition, some of them lives partly outside the circular hole when projected to $\mathbb{R}^2$.



(a) 1st combination.

(b) 2nd combination.

Figure 7.13: The geodesics in 3 dimensions for the two combinations of connecting points with equal boundary value.

The completed images for the two cases are shown in Figure 7.14. As expected, the result on the right cannot be the correct solution. First of all, the geodesics in the figure above can not span a minimal surface, and secondly, since each of them represent a specific value, the result cannot be continuous.

The result on the left looks very promising though. As with the HP model, the diagonal feature is preserved. Further, the colours near the boundary coincides with the region outside the hole, as expected.

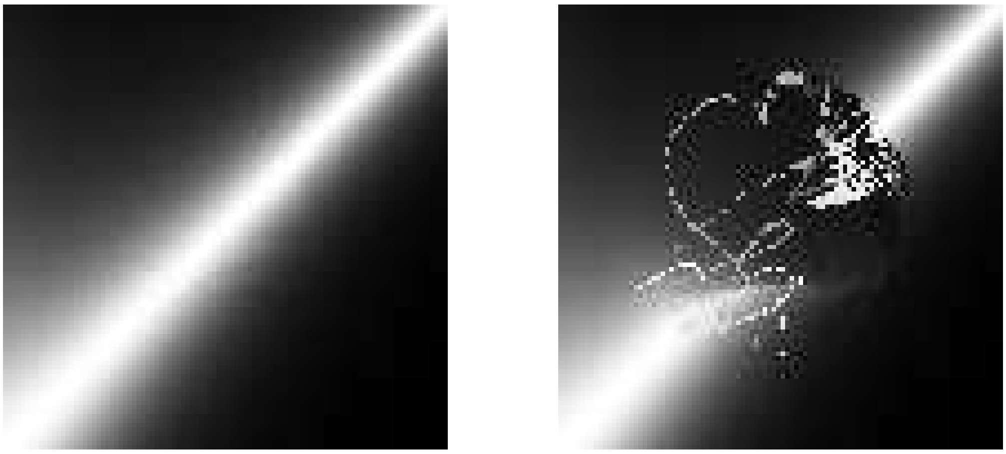
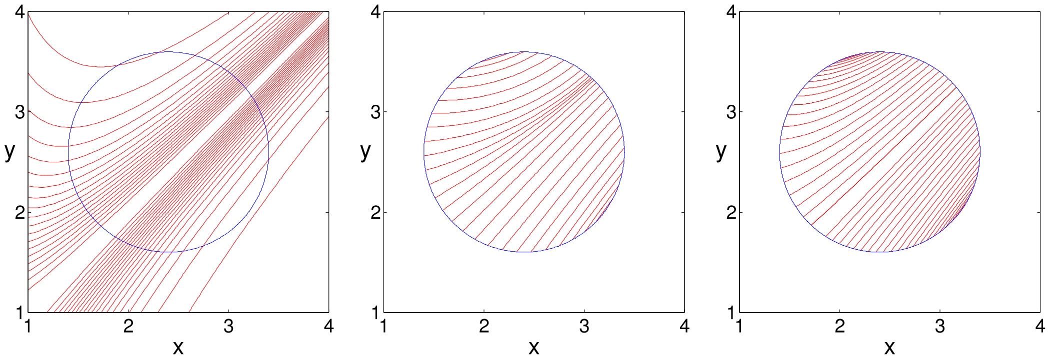

(a) 1st combination. (b) 2nd combination.

Figure 7.14: The completed images for the two combinations.

In Figure 7.15 we see (a) the exact contours of $I(x, y)$ compared with (b) the projected rules and (c) the projected geodesics. The preferable combination above is chosen for the geodesic model. Looking at the exact contours on the left, it is obvious that they are not circle cords, especially not in the upper left corner. Thus the projected rules cannot represent them perfectly. Further, we see a heap of rules ending up in the upper right. This corresponds to the place where we saw the distortions in the completed image in Figure 7.10 above. The projected geodesics on the right have the desired shape.



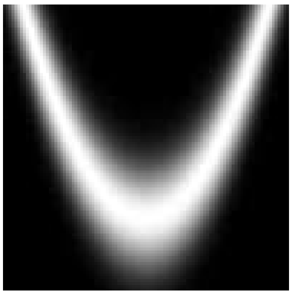
(a) Exact contours of $I(x, y)$. (b) Projected rules. (c) Projected geodesics.

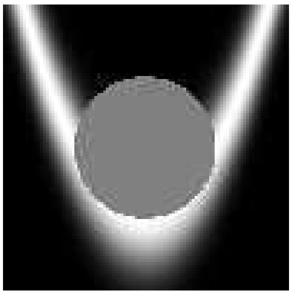
Figure 7.15: Comparison of exact contours versus projected rules and geodesics.

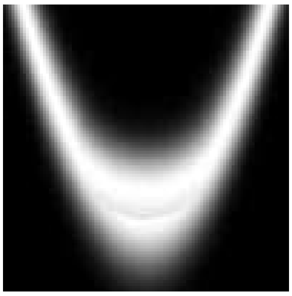
7.2.3 Example 3

The last example we are going to consider is the image (a) of Figure 7.16. It is constructed from the intensity function $I(x, y) = \exp(-2(y - x^2)^2)$ with $x \in [-2, 2]$ and $y \in [-1, 3]$. In

(b) we see the original image with a circular hole centred at $(0, 1)$ with radius 1. In (c) and (d) we see the completions using the HP model and the geodesic model, respectively. Both models preserves concavity, and the results are approximately equal. In the lower part of the hole we see the results have wrong colours. In this part of the image the exact contours coincides with the circle itself, and both models have difficulties constructing curves in this region.


(a) Original image.


(b) Original with hole.


(c) HP model.

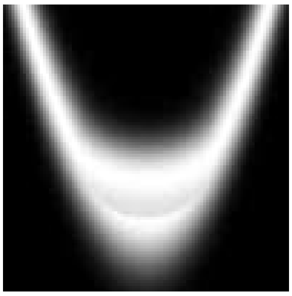

(d) Geodesic method.

Figure 7.16: Example 3: Original and completed.

7.3 Arbitrary boundary

In this section we will apply the geodesic model on an image with a non-circular hole. We will consider the image in Figure 7.17. In (a) we see the original, and in (b) the original with a rectangular hole.

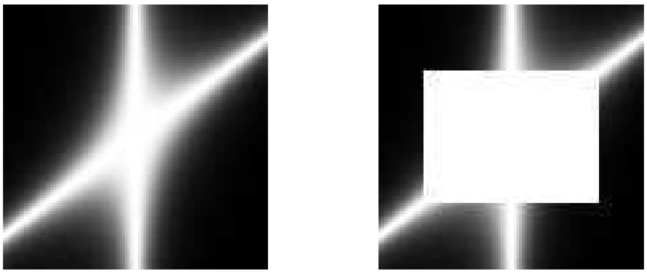

(a) Original image. (b) Original with hole.

Figure 7.17: Original image and original with rectangular hole.

The image is constructed from the intensity function $I = 1/(1 + x^2(2y - x)^2)$, with $x \in [-2.5, 2.5]$, $y \in [-1.5, 1.5]$. The rectangular hole is the part of the image where $-2 < x < 2$ and $-1 < y < 1$. In Figure 7.18 we see a contour plot of I. We note that the there are critical points along the rays $x = 0$ and $y = x/2$, and that the index of the rectangular curve is -1. This means that when we follow the curve counterclockwise, the gradient direction of I runs clockwise through S^1. Thus when considering modulo 2π, the lifted boundary will be continuous.

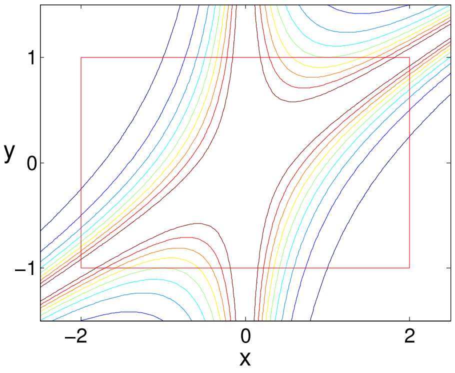


Figure 7.18: Exact contours with rectangular hole.

In Figure 7.19 we see the lifted boundary with calculated geodesics. Since we are only interested in completing a non-circular hole, the pairing is chosen to correspond with the exact contours shown above.

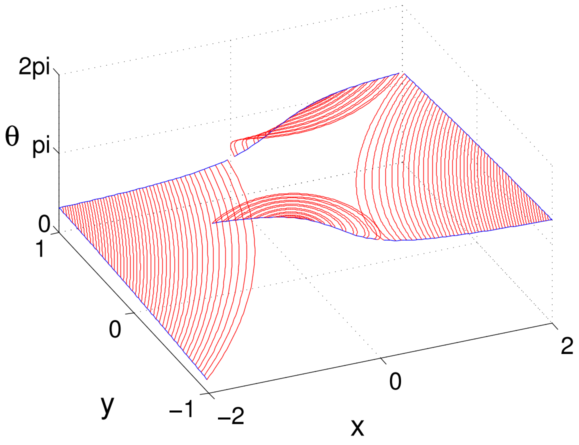


Figure 7.19: Lifted rectangular boundary with calculated geodesics.

When we project the same geodesics in the figure above, we get the result in Figure 7.20. The completed image is a good replica of the original, and the concavity is preserved.

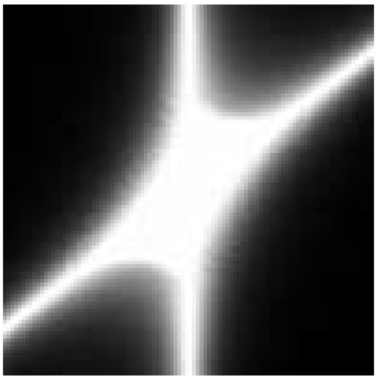


Figure 7.20: Completion of image with rectangular hole.

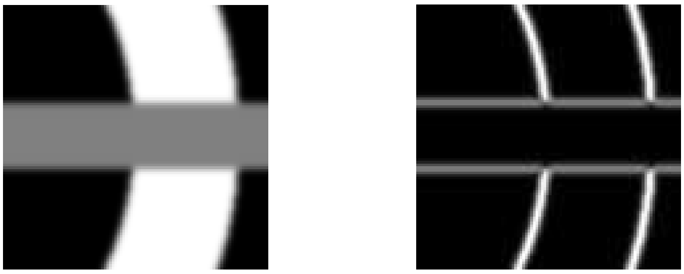
7.4 Advantages of the CS model

In this section we will show some examples where the "direct"-methods cannot be used. These are images where a defined "hole" is not present. For the HP model and the geodesic model we assume knowledge of the location of a hole, and the purpose of these models is to complete only this hole. In many natural circumstances, we are interested in all the objects in a particular image. The CS model assumes no extra information other than the original image itself. Though a long computation time compared to the other models, it only needs an image as input.

We start with an easy example, and finally we show the completion of Kanizsa's fishes mentioned in the introduction.

7.4.1 Occluded and occluding objects

In Figure 7.21 we see an original image on the left. The two white sections are perceived as



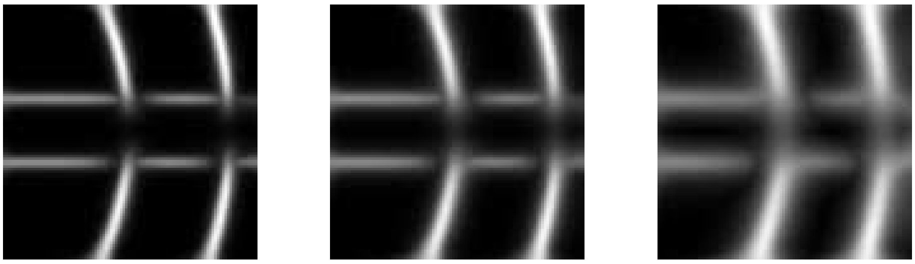
(a) Original image.

(b) Original image represented in u.

Figure 7.21: Original image and how it is represented in $\mathcal{RT}$.

one object extending behind the gray occluder. On the right, we see how this image is represented as the initial function u_0 (4.1) constrained on the initial surface Σ_0 (4.2). It is constructed from derivatives of the original image, so only the boundaries of the two objects can be seen. Since gray colours are represented as scalar values, the magnitude of the gradients of the two objects differ. The gray bar, which is the occluding object, has smaller gradient magnitudes which in turn gives darker boundary lines in the right image.

In Figure 7.22 we see the results after different iteration lengths. We see that the boundaries of the occluded object are diffusing through the occluder. Finally, the boundaries meets halfway through. Since this object have gradients of larger magnitude, it is the prominent one, and the gray bar seems to be lost. However, in Figure 7.23 below, a slice plot of the three-dimensional array u is shown. The different objects lives at different heights in u. The height corresponds to the gradient direction, and the gray bar can be extracted if it is of interest.



(a) 20 iterations. (b) 40 iterations. (c) 80 iterations.

Figure 7.22: Completion using the CS model.

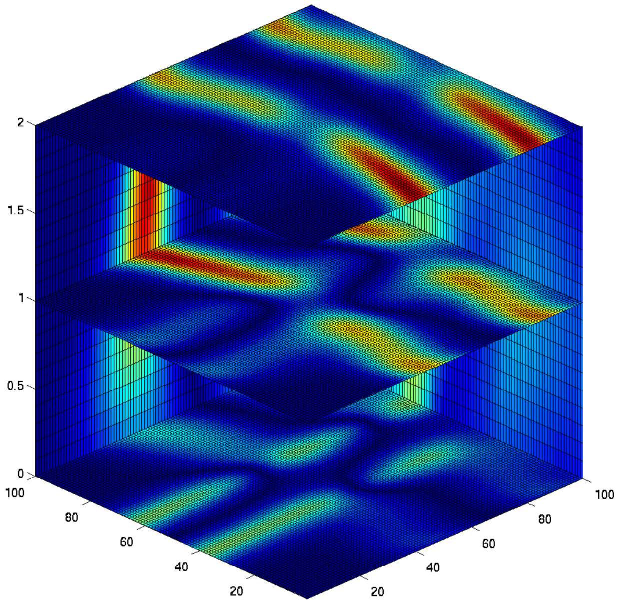
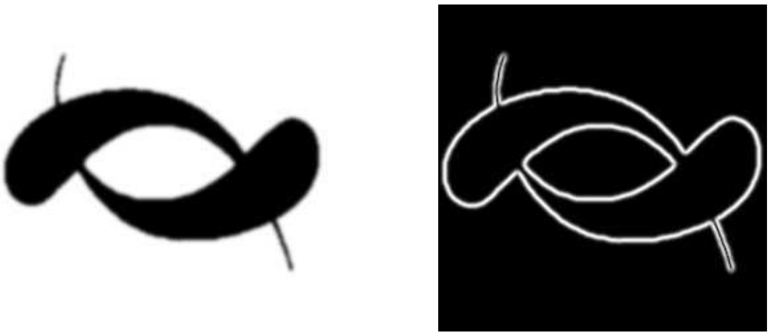


Figure 7.23: Representation in three dimensions.

7.4.2 Kanizsa's fishes

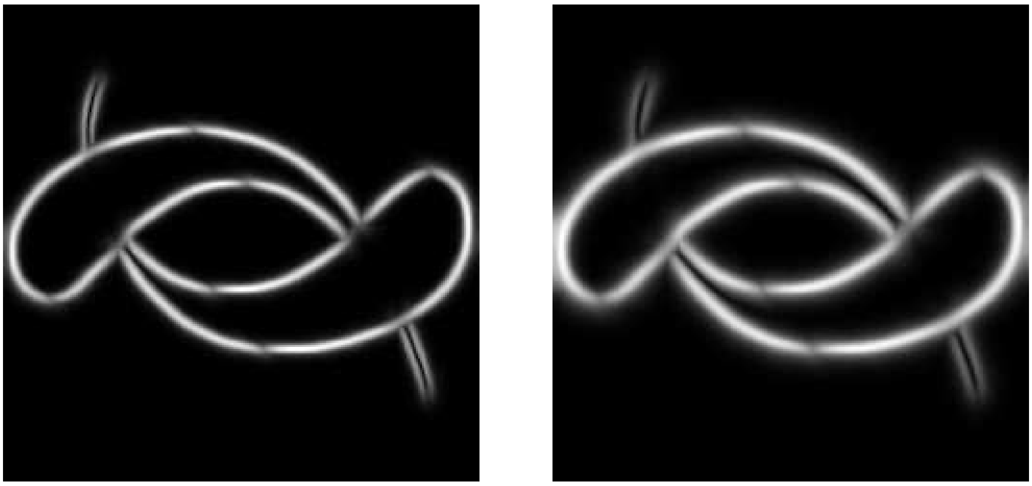
The last example we are going to consider is Kanizsa's fishes. The original, and the original representation in $\mathcal{RT}$, can be seen in Figure 7.24. The image on the left is perceived as two fishes, where the head of each fish occludes the others tail. There are no clear depth ordering present though, since the two fishes are represented with the same colour. On the right we see Kanizsa's fishes represented as the initial function $u_0\delta_{\Sigma_0}$. As above, only the boundaries where the gradients are non-zero can be seen.



(a) Original image. (b) Original image in $\mathcal{RT}$.

Figure 7.24: Kanizsa's fishes.

Figure 7.25 shows the results when applying the CS model for 20 and 50 iterations. We see that the boundaries of the heads diffuse through the tails, and it looks as if the fishes are spiralling upwards. In a slice plot of the three-dimensional function u, we would see that, below the heads, the tails also diffuses. This means that both objects could have been extracted in its entirety.



(a) 20 iterations. (b) 50 iterations.

Figure 7.25: Kanizsa's fishes completed with the CS model.

Chapter 8

Conclusion and future work

Three different perceptual completion models has been described and implemented. They are all based on the sub-Riemannian geometry on $\mathcal{RT}$ introduced in [1], which models the first layer of the visual cortex. Images are "lifted" to $\mathcal{RT}$, and completed. The first model completes both occluded and occluding objects in an image by a diffusion process, while the last two models are designed to complete a missing hole by directly calculate curves which spans this hole.

The model of Citti and Sarti has the advantage that only an image with no extra information is assumed. The different objects in an image "lives" at different heights in the three-dimensional space $\mathcal{RT}$, and we can extract the object of interest after the completion process.

The model of Hladky and Pauls [2] completes a circular hole by spanning the lifted boundary with rules. This will produce a minimal surface, and from the model above these solutions are completions of occluded visual data. When projecting the rules to $\mathbb{R}^2$, holes in images can be reconstructed.

The last model, which were designed to match equal Dirichlet boundary values on the boundary of an open hole, uses geodesics instead of rules. As above, these span a minimal surface covering the hole, and projected geodesics are completed contour-lines in the original image.

The two latter models are considerably less computationally intensive than the diffusion based model. However, these models can only complete images with a well-defined hole. On the other hand, when assuming a known underlying continuous intensity function, the solutions with these models are "exact". By exact we here mean that the solution curves spanning the minimal surface are constructed directly, and for the HP model we also have explicit expressions for these.

Chosen example images were completed and discussed. The last two images in Section 7.4 are great examples of the perceptual completion phenomena, and a verification that the model of Citti and Sarti works. The "direct" methods did also work well, especially on images which suited them. Both models preserved concavity, which is a feature that diffusion based models often have trouble with. We saw that for an image with circular contours, the model of Hladky and Pauls completed the hole perfectly. However, on other images there were some distortions and mismatch of boundary values on each end-point of the rules. This was expected, as no information about the intensity of the image at the boundaries are used when constructing these rules. The geodesic model had some problems with one of the examples. The calculated geodesics seemed to be shorter than what we want. This is not necessarily "wrong" since the result is a solution to *some* occluded visual data. Beacause we only use information on the boundary of the hole, the result will not always match the properties of the image in the surroundings.

Future work can include the following. The model of Citti and Sarti can be implemented in a more sophisticated way. The overall strategy regarding the implementation here was simplicity. For instance, the simple forward Euler scheme can be replaced by a more sophisticated time integrator. Further, the Lie group structure of the problem can be exploited more. In [26], a grid is generated based on the structure of the Heisenberg group, and a similar extension to the roto-translational group could be beneficial.

The model of Hladky and Pauls can be adjusted to consider arbitrary holes. It would also be interesting to adjust it to the discrete case by interpolating the discrete lifted points on the boundary of the hole, as mentioned in Section 5.6. In natural circumstances, we do not have the knowledge of an underlying intensity function $I(x, y)$.

The geodesic model can also be adjusted to the discrete case. From the results, it is also clear that including information outside the hole would be beneficial. When the pairing of points with equal boundary values is non-trivial, information from outside the hole can possibly indicate the right combinations. Further, with the geodesic equations (2.18) we can include more boundary points with equal colours outside the hole. Numerical approximations of the solution to this boundary value problem would be horizontal paths, but not necessarily geodesics. Applying this procedure to the example in Figure 7.7, where the calculated geodesics were “too short”, could possibly produce better results. As with the model of Citti and Sarti, the Lie group structure can be exploited. When finding numerical solutions of the geodesic equations, a Lie group integrator can be used.

In Section 3.4, we saw that approximations to the horizontal mean curvature motion problem could be found using a stochastic approach. Since a minimal surface is a surface with mean curvature zero, this stochastic approach can be used to find a converging horizontal minimal surface in $\mathcal{RT}$. Numerical solutions of the “horizontal” stochastic differential equation (3.28) could be found using, for instance, the Euler-Maruyama method. Then a good way to approximate the infimum in Theorem 3.1 over all stochastic controls must be found.

Bibliography

[1] G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. *J Math Imaging Vis*, 24:307–326, 2006.

[2] R. K. Hladky and S. D. Pauls. Minimal surfaces in the roto-translational group with applications to a neuro-biological image completion model. Submitted, 2008.

[3] P. Perona and J. Malik. Scale-space and edge detection using anisotropic detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[4] D. Mumford, M. Nitzberg, and T. Shiota. *Filtering, segmentation and depth*, volume 662 of *Lecture notes in Computer Science*. Springer-Verlag, Berlin, 1993.

[5] S. Masnou and J.-M. Morel. Level lines based disocclusion. *IEEE International Conference on Image Processing*, 3:259–263, 1998.

[6] S. Esedoglu and R. March. Segmentation with depth but without detecting junctions. *J. Math. Imaging Vision*, 18(1):7–15, 2003.

[7] G. Bellettini and R. March. An image segmentation variational model with free discontinuities and contour curvature. *Math. Models Methods Appl. Sci.*, 14(1):1–45, 2004.

[8] J. Petitot and Y. Tondut. Toward a neurogeometry: cortical fibrations, contact structures and modal subjective contours. *Math. Inform. Sci. Humaines*, 145:5–101, 1999.

[9] R. R. Seeley, T. D. Stephens, and P. Tate. *Essentials of Anatomy & Physiology.* McGraw-Hill, 6th edition, 2006.

[10] M. Carandini, D. J. Heeger, and J. A. Movshon. Linearity and gain control in v1 simple cells. *Cerebral Cortex*, 13, 1999. Models of cortical circuits.

[11] P. M. Viddal. A perceptual completion model. Pre-master at the Master's degree programme within Industrial Mathematics at NTNU, Trondheim, 2008.

[12] R. Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics and Applications*, volume 91. Mathematical Surveys and Monographs, American Mathematical Society, 2006.

[13] H. Geiges. Contact geometry. *Handbook of differential geometry*, 2:315–382, 2006. Elsevier/North-Holland, Amsterdam.

[14] J. Jost. *Riemannian Geometry and Geometric Analysis.* Springer-Verlag Berlin Heidelberg, 5th edition, 2008.

[15] B. Franchi, R. Serapioni, and F. S. Cassano. On the structure of finite perimeter sets in step 2 of carnot groups. *J. geom. Anal.*, 13(3):421–466, 2003.

[16] G. Dziuk and C. M. Elliot. An eulerian level set method for partial differential equations on evolving surfaces. *Computing and Visualization in Science*, pages 1–12, 2007.

[17] L. C. Evans and J. Spruck. Motion of level sets by mean curvature i. *J. Differ. Geom.*, 33:635–681, 1991.

[18] Y.-G. Chen, Y. Giga, and S. Goto. Uniqueness and existence of viscosity solution of generalized mean curvature flow equations. *J. Differential Geom.*, 33:749–786, 1991.

[19] J. Bence ans B. Merriman and S. Osher. Diffusion generated motion by mean curvature. *Computational Crystal Growers Workshop, J. E. Taylor, ed., Sel. Lectures Math., AMS, Providence*, pages 73–83, 1992.

[20] L. C. Evans. Convergence of an algorithm for mean curvature motion. *Indiana Univ. Math. J.*, 42(2):533.557, 1993.

[21] R. Buckdahn, P. Cardaliaguet, and M. Quincampoix. A representation formula for the mean curvature motion. *SIAM J. Math. Anal.*, 33(4):827–846, 2001.

[22] W. H. Fleming and H. M. Soner. *Controlled Markov Processes and Viscosity Solutions.* Stochastic Modelling and Applied Probability 25. Springer Science + Business Media, Inc, 2nd edition, 2006.

[23] N. Dirr, F. Dragoni, and M. V. Renesse. Evolution by mean curvature flow in subriemannian geometries: A stochastic approach. Submitted, 2008.

[24] B. Øksendal. *Stochastic Differential Equations, An Introduction with Applications.* Springer-Verlag Berlin Heidelberg, 1985.

[25] D. J. Field, A. Hayes, and R. F. Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Res.*, 33(2):173–193, 1993.

[26] Y. Achdou and N. Tchou. A finite difference scheme on a non commutative group. *Numer. Math.*, 89:401–424, 2001.