ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea Magistrale in Matematica

Diffusion maps in the subriemannian motion group and perceptual grouping

Tesi di Laurea in Analisi Matematica

Relatore:
Chiar.ma Prof.ssa
Giovanna Citti

Correlatore:
Chiar.mo Prof.
Alessandro Sarti

Presentata da:
Chiara Nardoni

III Sessione
Anno Accademico 2012-2013

ACKNOWLEDGMENTS

Questa tesi è frutto di un lavoro tenace che nasce come proseguimento ideale dell'attivita professionalizzante svolta sotto la tutela della prof.ssa Giovanna Citti. Pertanto si tratta del frutto di una assidua interazione, grazie alla quale sono emersi man mano nuovi stimoli e apporti fondamentali ai fini della conclusione di questo lavoro. Desidero ringraziare il dott. Giacomo Cocci per gli esperti suggerimenti riguardo l'implementazione numerica dei risultati.

Desidero inoltre ringraziare la dott.ssa Marta Favali per la proficua collaborazione e l'apporto decisivo per conclusione delle prove sperimentali.

Un ringraziamento particolare va al prof. Alessandro Sarti, Correlatore della tesi, per l'importante sostegno e preziosi suggerimenti nella redazione del lavoro.

Il pensiero più importante va alla prof.ssa Giovanna Citti, Relatrice della tesi, per avermi incoraggiata e sostenuta con entusiasmo, per il tempo prezioso dedicato al mio lavoro, per l'approccio originale, per le conversazioni mai scontate, per avermi aperto le porte a una visione inedita della matematica.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

This thesis is motivated by the problem of constitution of perceptual units in visual perception. The issue deals with the mechanism allowing the information distributed in the visual areas to get bound together into coherent object representations, and can be faced with dimension reduction technique, as for example the diffusion maps, recently introduced by R.R. Coifman and S. Lafon in [29],[12][11]. These phenomena take place in the first layer of the visual cortex, which has been modelled as contact structure with a subriemannian metric. The main scope of this thesis will be to exploit instruments of potential theory and spectral analysis in Lie groups to formalize dimensionality reduction techniques in subriemannian context. The problem will be reduced to the convergence of a graph Laplacian to the sublaplacian differential operator, whose eigenfunctions will be interpreted as the objects present in the image.

When the brain processes the visual stimulus, it extracts meaningful features from the image, such as spatial position, color, brightness, orientation, movement, stereo. The acquisition of the visual stimulus is an integrative process, taking into account the multi-level structure of the visual cortex. Geometrical models of this structure have been provided in [33], [34], [38], [28], [42], [15],[19],[20]. This thesis focuses on the geometrical model of V1 proposed by Citti-Sarti in [8]. The 2D retinal images are lifted to the higher dimensional manifold of the cortex encoding both the physical variables and the engrafted variables such as orientation and velocity preferred. Any level line of the image is lifted to a new curve in the cortical space, whose geometric properties will be encoded and described by the definition of a subriemannian metric.

A subriemannian metric in $\mathbb{R}^n$ is defined by the the choice of $m \leqslant n$ vector fields $X_1, \cdots, X_m$, called horizontal vector fields, at

every point. The selected subspace is called horizontal tangent space. The horizontal vector fields satisfy the Hörmander condition of hypoellipticity when a bracket generating condition holds. Under these assumptions, the differential calculus in subriemannian setting can be introduced by replacing the derivatives with the vector fields. Following the classical presentations of Nagel, Stein, Wainger [31] and [4], the horizontal tangent space is endowed with an horizontal metric which allows the definition of the length of curves and distance between points . The subriemannian differential operators are defined by the horizontal vector fields. The gradient of a function f is defined as $\nabla f = (X_1, \cdots, X_m)$, laplacians and heat operators are accordingly defined. Due to the strong degeneracy of the metric these operators are totally degenerate at every point. However the existence of fundamental solution is well known, since the works of Hormander [21] and Rothshild and Stein [35]. The results regarding subriemannian geometry are collected in Chapter (3).
The cortico-cortical connectivity is described in terms of families of integral curves of horizontal vector fields. The outcomes are compatible with both the neurophysiological findings [17],[6] and the principles of proximity, good-continuation and co-circularity prescribed by Gestalt pshycological theory [26].

Since the perceptual saliency is described in terms of stochastic diffusion processes, a brief overview of stochastic calculus is presented in chapter (4), following the classical presentation in [32]. The fundamental definitions of Brownian motion and Ito stochastic integrals are stated. The transition probability of a stochastic process is recognized as the fundamental solution of a parabolic second order differential operator of Fokker Planck type or Heat type in the considered dynamic. The stochastic models of good continuation of contours were proposed by Mumford [30]. Stochastic cortical connectivity models are due to [1],[42]. These models are able describe the anisotropic diffusion dynamics of the cortical brain. The present work concentrates on diffusion by the heat operator on the rototranslation Lie group.

Chapters (5) and (6) contain the original contribute of the work. The previous result provides just an existence result for the fundamental solution of the heat equation. In chapter (5) a fine approximation of the fundamental solutions in the neighborhood of the pole is accomplished . The estimate is performed by an adaptation of the parametrices method described in [24],[37], which will be adapted to the non nilpotent setting of the motion group. The heat fundamental solution on the motion group is locally approximated by the heat fundamental solution on the Heisenberg group: the horizontal vector fields are approximated by frozen vector fields defined on a stratified Lie algebra. The approximation is validated by homogeneity arguments. On the other hand the fundamental solution of the Heisenberg heat operator is explicitly known, providing a careful approximation of the $SE(2)$ heat kernel.

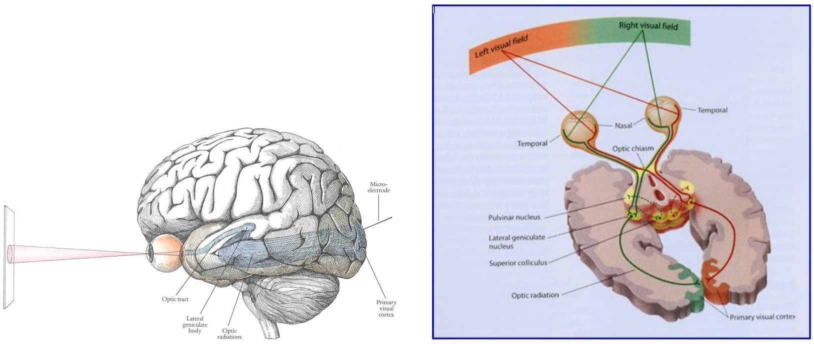
The fundamental solution local estimate is used in chapter (6) to study the subriemannian version of the diffusion maps algorithm, and to run it to constitution of perceptual units. The diffusion maps method uses the spectral decomposition of a properly normalized diffusion operator to embed the data and to recognize underlying low-dimensional geometric structures inside the data. In [29] a graph Laplacian is defined in terms of an average on a submanifold of $\mathbb{R}^n$, weighted with the heat kernel in the euclidean group. Then it is proved the convergence of this operator to the Laplace-Beltrami operator for Gaussian kernels on submanifolds of $\mathbb{R}^n$. In chapter (6) the analogous proof is performed in the subriemannian context. The main difficulties are related to the strong degeneracy of the operator and to the non nilpotency of the space. In addiction we restrict to curves, since the definition of Laplace-Beltrami operator on a surface is not totally known in the subriemannian setting. The approximate heat kernel admits a spectral decomposition. Moreover the approximate eigenfunctions converge to the eigenfunctions of the heat operator on the motion group. Then the first few eigenfunctions are used to perform clustering. In the end numerical results are presented and allow to recover constitution of perceptual units.

2

THE PRIMARY VISUAL CORTEX AS A LIE GROUP

The primary visual cortex is the best studied area in the brain. In recent years formal models of the visual brain have been provided by different authors: Koenderink [28], Hoffman [19], Zucker [42], Mumford [30], Petitot [33][34], Citti Sarti [8] and Sarti, Citti, Petitot [38]. Here a brief excursus on the functional architecture of the visual cortex, taking into account its multi-level structure, is given. A detailed description of the Citti Sarti and Sarti, Citti, Petitot models is shown.

2.1 THE FUNCTIONAL ARCHITECTURE OF THE VISUAL CORTEX



(a) (b)

Figure 1: The retinogeniculate visual pathways

2.1.1 *Visual cortex areas*

The primary visual cortex is divided into six functionally distinct layers, labeled from 1 to 6. The visual cortex shows a modular structure which organizes many families of cells, according to their functionality. Each family is sensible to a specific feature of the image: position, orientation, ocular dominance, scale, color, curvature, move-

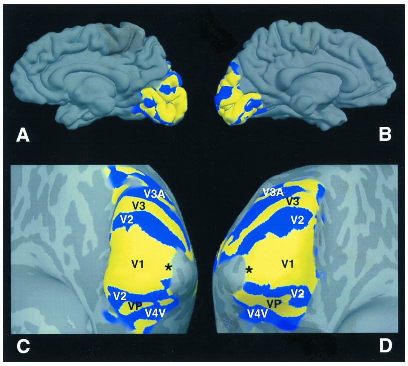


Figure 2: Visual cortex layers

ment, stereo. Hubel and Wiesel (see [23]) classified three types of V1 neurons (simple, complex and hypercomplex cells), depending on their response to visual stimuli. The recorded response decrease as the orientation of the visual stimulus changes from the preferred orientation. Moreover, the strength of perception increase of movement is considered. We will focus on cells skilled in orientation and velocity selectivity.

2.1.2 *Receptive fields and receptive profiles*

A visual neuron is connected to the retina through the neural connections of the retino-geniculo-cortico pathways. The projection from retina to the cortex is acted through the talamic way.
The receptive field (RF) of a visual neuron is classically defined as the domain of the retina to wich the neuron is connected and whose stimulation elicitate a spike response. Each RF is well-defined over a small regions of the space. A RF is decomposed into ON (positive contrast) and OFF (negative contrast) zones, referring to the type of response to light and dark stimultions.
The receptive profile (RP) of a visual neuron is a real valued function $\phi(x, y)$ which associates to each couple of points (x, y) of the retina domain the response (positive or negative) to stimulus. There exist different models for RF, changing with the type and the spatial organization of cells. Classically the RFs are modeled using derivatives of Gaussians. Since each family of cells is sensible to a specific feature,

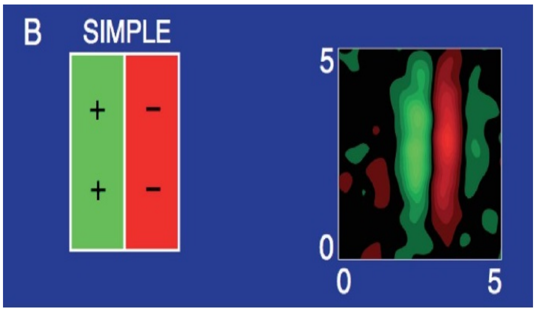


Figure 3: The RP of a simple cell sensible to orientation

the receptive profile model also depends on a vector e enconding the eingrafted variables. The the receptive profile is denoted by:

$$\phi = \phi_e(x, y) \tag{1}$$

where e denotes the dependence on orientation and velocity.
The neural activity of the receptive profile (1) in response to a visual stimulus is modeled as:

$$O_e = \int I(x, y)\phi_e(x, y)\, dx\, dy \tag{2}$$

where $I(x, y)$ denotes the optical signal at retinal point (x, y).

2.1.3 *The retinoptic map*

The retinoptic structure of the visual cortex refers to the mappings existing from the retina to the cortical layers. A retinoptic map is a map:

$$\rho : D \rightarrow M$$

which is an isomorphism and preserves the retina topology.
A mathematical model well fitting the empirical data is a logarithm conformal map (see fig. (4)).
From now on we will identify the two planes.

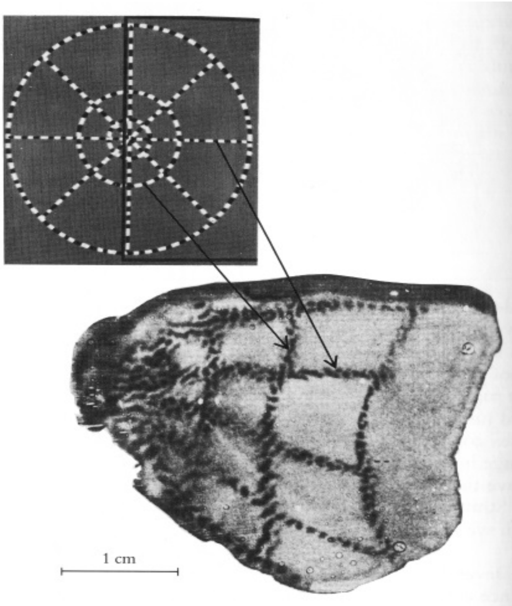


Figure 4: Retinoptic map

2.1.4 *The hypercolumnar structure of* V1

The columnar functional organization of the cortex, originally discovered by Vernon Mountcastle, was framed by the Nobel Prize D. Hubel and T. Wiesel [22],[23].
The cortex is organized in a modular structure consisting of repeating sets of functional columns. This functional structure has been accurately described in the primary visual cortex. The V1 simple cells are organized in columnar and hypercolumnar structure corresponding to features such as orientation, ocular dominance and color. Over every point (x, y) of the retinal plane there is an entire set of cells, each one sensible to a particular orientation or, when movement is considered, to a particular velocity. The set of simple cells will be parametrized in order to take into account the the invariance by translation. Then (1) will denote the whole set of receptive profiles over the point (x, y) and each column will be parametrized by varying the parameter vector e. Then:

$$\phi_e(x - x_0, y - y_0) \tag{3}$$

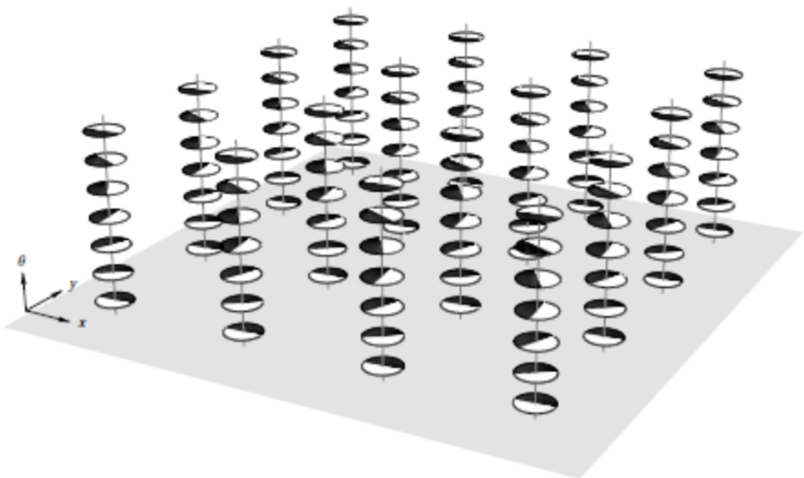


Figure 5: Visualization of the hypercolumnar structure of V1 with the engrafted variable θ

will model the columnar structure over each other retinal point (x_0, y_0). The action (2) turns into the convolution:

$$O_e(x_0, y_0) = \int I(x, y)\phi_e(x - x_0, y - y_0)\, dx\, dy = (I * \phi_e)(x_0, y_0) \quad (4)$$

In presence of the visual stimulus centered at point (x_0, y_0) with edge orientation and velocity e, all the hypercolumn over point (x_0, y_0) is activated and the simple cells sensible to the values e show maximal response. Even if the mechanism producing strong orientation and velocity selectivity is controversial, it is an experimental evidence that intracortical circuitry realizes the suppression of all the non-maximal directions. Then each retinal point (x, y) is lifted to a point (x, y, e^*) with e^* such that:

$$O_{e^*}(x_0, y_0) = \max_e O_e(x_0, y_0) \quad (5)$$

2.2 LIFTED CURVES AND CONSTRAINED DYNAMICS

Since the RP of a visual neuron depends on both spatial variables and engrafted variables, a reasonable assumption is to obtain the family

ϕ_e of RPs simple cells sensible to the image features encoded by vector e as a rigid transformation of a mother profile ϕ_0. The set of RPs ϕ_θ with a preferred orientation θ is modeled as a rigid transformation of a mother profile ϕ_0. The mother RP defined over the retinical point (x, y) is given by:

$$\phi_0(x, y) = \partial_y \exp(-(x^2 + y^2)) \tag{6}$$

Therefore the family ϕ_θ of RPs over point (x, y) is achieved by applying a counterclockwise planar rotation R_θ to ϕ_0:

$$\phi_\theta(x, y) = \phi_0 \circ R_\theta(x, y) = \phi_0(x \cos\theta + y \sin\theta, -x \sin\theta + y \cos\theta). \tag{7}$$

The RP of preferred orientation θ over each point (x', y') of retinal domain is simply obtained by applying a translation $T_{x',y'}$ to ϕ_θ:

$$\phi_\theta(x', y') = \phi_\theta \circ T_{x',y'}(x, y) = \phi_0(x - x', y - y'). \tag{8}$$

Applying a rotation of an angle θ (6) becomes:

$$\phi_\theta(x, y) = X_3 \exp(-(x^2 + y^2))$$

where X_3 is the vector field obtained by rotation of ∂_y:

$$X_3 = -\sin\theta\partial_x + \cos\theta\partial_y. \tag{9}$$

The filtering output (2) is a function $O(x, y, \theta)$ obtained by convolution with the image I:

$$O(x, y, \theta) = \int \phi_\theta(x - x', y - y')I(x', y')\, dx'\, dy' = \tag{10}$$

$$= -X_3 \exp(-(x^2 + y^2)) * I = -X_3 I_s(x, y),$$

where I_s is a smoothing of the image I:

$$I_s = I * \exp(-(x^2 + y^2)).$$

In order to integrate the filtering process with the non-maximal suppression mechanism, the direction of maximum response is considered. Applying (5) we impose the maximality condition:

$$O(x, y, \theta^*) = \max_\theta O(x, y, \theta).$$

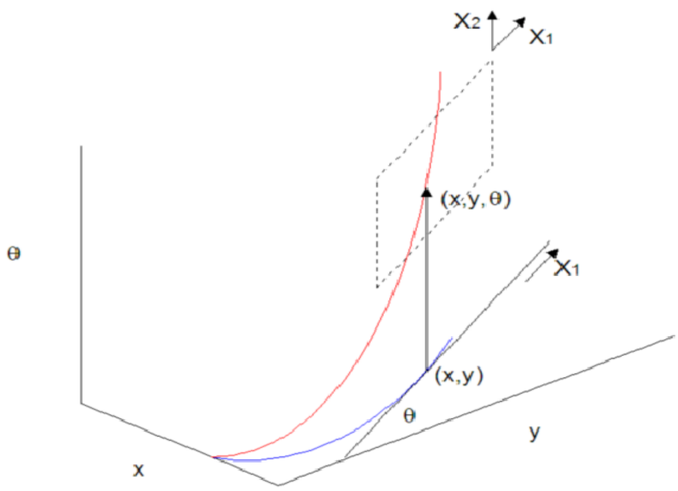


Figure 6: Visualization of a lifted curve

Consequently the preferred orientation θ^* is selected requiring:

$$\partial_\theta O(x, y, \theta^*) = 0$$

Calling:

$$X_1 = \cos\theta\partial_x + \sin\theta\partial_y$$

the following geometrical condition holds:

$$0 = \partial_\theta O(x, y, \theta^*) = \partial_\theta X_3(\theta^*)I = -X_1(\theta^*)I = -\langle X_1(\theta^*), \nabla I)\rangle \quad (11)$$

The relation (11) imposes that the vector $X_1(\theta^*)$ is orthogonal to the gradient of I at every point, which means that it has the direction of the level lines of I. Each point (x, y) is lifted to the 3D cortical point $(x, y, \theta^*) \in R^2 \times S^1$. Then any level line of I is lifted to a curve in $R^2 \times S^1$.

By construction the tangent vector to the lifted curve can be expressed as a linear combination of the vector fields:

$$X_2 = \partial_\theta$$

$$X_1 = \cos\theta\partial_x + \sin\theta\partial_y$$

on the lifted space $\mathbb{R}^2 \times S^1$. The lifting process select at every point a two-dimensional distribution of tangent planes, generated by X_1 and X_2 in the three dimensional space $R^2 \times S^1$.

In the next section this mechanism will be mathematically formalized in terms of a subriemannian contact structure defined on the space $\mathbb{R}^2 \times S^1$.

2.2.1 V1 *as a contact structure*

The vector fields X_1, X_2 select a 2D-subspace of the 3D tangent space at every point, prescribing a constrained dynamic on the tangent space. The condition can be equivalently formulated in terms of a contact 1-form defined on the cotangent space at every point. From the vector field X_3, we endow the cotangent bundle $T^*\mathcal{RT}$ with the contact 1-form:

$$\omega_{x,y,\theta} = -\sin\theta\, dx + \cos\theta\, dy.$$

The geometrical condition (11) is reformulated by imposing that the lifted curves lie in the kernel of ω, i.e. the subspace:

$$ker\omega_{x,y,\theta} = \{(v_1, v_2, v_3) \in T_{(x,y,\theta)}\mathcal{RT} : -\sin\theta v_1 + \cos\theta v_2 = 0\}$$

Note that $ker\omega$ is the subspace orthogonal to X_3, i.e. the plane spanned by vector fields X_1 and X_2 at every point. In chapter (3) the constrained dynamic arising from the lifting model will be formalized by the definition of a subriemannian contact structure on the lifted space.

2.2.2 V1 *as the rototranslation group* $\mathcal{RT}$

The result of lifting process is a costrained dynamic on the lifted space $\mathbb{R}^2 \times S^1$. Since each cell is obtained by a rotation and a translation from a fixed one, we will identify the space of cells with the structure of the Lie group of translations and rotations.

The rototranslation Lie group $\mathcal{RT}$ is the space:

$$\mathbb{R}^2 \times S^1$$

equipped with the composition law:

$$(x, y, \theta) +_{\mathcal{RT}} (x', y', \theta') = \left(((x, y) + R_\theta(x', y'))^T, \theta + \theta'\right)$$

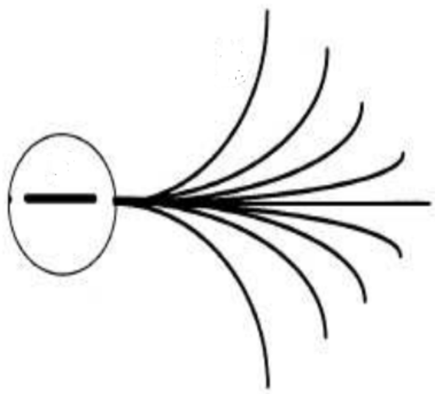


Figure 7: Association Field of Field, Hayes and Hesse

where R_θ is a counterclockwise planar rotation of an angle θ.

In addition we have recognized that the all curves of the space are integral curves of two vector fields in a 3D structure. We will see that this induces a Sub-riemannian structure.

2.2.3 *Cortical connectivity and Association fields*

Horizontal cortico-cortical connections link cells of similar orientation and velocity in distinct hypercolumns. These connections are long range (up to 6-8mm) and strickly anisotropic. The cortical connectivity seems at the basis of many visual phenomena as for example perceptual completion. Indeed the phenomenological conterpart of cortical connectivity are the so called association fields modelled by Field, Hayes and Hesse (see fig. (7)). A stimulus at the retinical point $(0, 0)$ with horizontal tangent can be joined by a subjective boundary only to a stimulus which is tangent to the association lines depicted in the figures. The association field defines a notion which agrees with the good-continuation, proximity and cocircularity principles prescribed by Gestalt psychological theory.

2.3 A SPATIO-TEMPORAL MODEL

An extension of the Citti-Sarti model [8] is proposed in [1] to accomplish the processing of spatio-temporal visual stimuli.
In order to take account of both local orientation preference and velocity-selectivity, we need to consider the temporal behaviour of the simple cells. The Lie group properties allow to describe the modular structure of different families of cells, detecting more and more engrafted variables, simply increasing the dimension of the underling physical space. The receptive profiles are now defined in the joint domain of space and time. De Angelis et al. in [14] provide a description of such RPs, distinguishing between space-time separable (if the spatial spatial and temporal response characteristics can be dissociated) and inseparable profiles. Slightly modifying the model of Barbieri et al. in [1] we will choose a family of filters, which is able to fit very well the experimental data of both separable and inseparable RPs. Under this view we choose Gaussian RPs:

$$\phi_{\theta,\nu}(x,y,t) = (X_3 - \nu\partial_t)e^{-x^2-y^2-t^2},$$

These filters act on a spatio temporal stimulus $f = f(x,y,t)$ by convolution:

$$O_{\theta,\nu}(x,y,t) = \phi_{\theta,\nu} * f(x,y,t),$$

In perfect analogy with the lower dimensional case, a non maximal suppression mechanism associates to the action of these filters the contact structure generated by the 1-form with the same coefficients as the vector $X_3 - \nu\partial_t$:

$$\omega = -\sin(\theta)dx + \cos(\theta)dy - \nu ds$$

The vanishing condition $\omega = 0$ identifies ν as the velocity in the direction of the vector X_3. Since this is orthogonal to the boundaries, ν describes the perceived velocity. In perfect analogy with the lower dimensional case, the connectivity among cells in this setting is described in terms of admissible directions of the tangent space. The contact structure provides a constraint on the dynamic, selecting a sub-algebra of admissible directions, namely the directions belonging to the kernel of ω. The horizontal tangent space is spanned by the frame $\{X_1, X_2, X_4, X_5\}$ given by:

$$\begin{aligned}
X_1 &= \cos\theta\partial_x + \sin\theta\partial_y \\
X_2 &= \partial_\theta \\
X_4 &= \partial_v \\
X_5 &= -v\sin\theta\partial_x + v\cos\theta\partial_y + \partial_s
\end{aligned}$$

The horizontal tangent space will be responsible of the neural connectivity among cells. The lifted space

$$\mathbb{R}^2 \times \mathbb{R}^+ \times S^1 \times R^+ = \{(x, y, t, \theta, v)\}$$

endowed with the contact structure prescribed by ω will define a 5D subriemannian manifold which will be denoted by $\mathcal{M}$.

3 THE SUBRIEMANNIAN GEOMETRY OF V1

We have seen in the previous section that the internal geometry of the visual cortex $V1$ is strongly anisotropic. The space is endowed with a contact structure which models the constraint on admissible connectivity among cells. Even if the tangent space to $\mathbb{R}^2 \times \mathbb{S}^1$ is a vector space of dimension 3 at every point, the $V1$ neurogeometrical model selects at every point of the $3D$ cortical structure a two dimensional plane, subset of the tangent space at every point. The induced metric structure on the tangent plane, which is an example of sub-riemannian metric.

In this chapter we will depict the subriemannian geometry of the horizontal distribution of vector fields. An overview on homogoeneous Carnot groups is presented. Differential operators and calculus in subriemannian setting are described (we refer to [4] for a detailed presentation).

3.1 HÖRMANDER VECTOR FIELDS

Definition 3.1.1. *Let $\mathcal{N}$ be a manifold of dimension n. A smooth distribution $\mathcal{H}$ of costant rank $m \leqslant n$ is a subbundle of dimension m of the tangent bundle. That is, at every point, the distribution is a subspace of dimension m of the tangent space.*

Remark 3.1.1. *The vector fields*

$$X_1 = \cos\theta\,\partial_x + \sin\theta\,\partial_y \tag{12}$$

$$X_2 = \partial_\theta \tag{13}$$

constitute an horizontal frame for $\mathcal{RT}$.

Then the vector space

$$\mathcal{H} = \mathrm{span}\{X_1, X_2\}$$

is a distribution on $\mathcal{RT}$.

Remark 3.1.2. *The orthogonal vector fields:*

$$X_1 = \cos\theta\partial_x + \sin\theta\partial_y \quad (14)$$

$$X_2 = \partial_\theta \quad (15)$$

$$X_4 = \partial_\nu \quad (16)$$

$$X_5 = -\nu\sin\theta\partial_x + \nu\cos\theta\partial_y + \partial_s \quad (17)$$

$$(18)$$

constitute an horizontal frame for $\mathcal{M}$.

Obviously $\mathcal{RT}$ and $\mathcal{M}$, being Lie groups, carry the structure of a manifold.
An explicit computation shows that the non zero Lie bracket of the vector fields X_1 and X_2 is the vector field X_3. We remark that the commutator is linearly indipendent of the horizontal distribution at every point.
Due to the non commutative relation, the horizontal vectors together with their commutators span the whole tangent space at every point. That is, the Lie algebra structure on the horizontal distribution recovers the missing dimension. This is the so called Hörmander condition.

Definition 3.1.2. *A distribution is bracket generating, or verifying the Hörmander condition, or completely non integrable, if any local frame together with a finite number of iterated Lie bracket spans the whole tangent space at every point. Equivalently the vector fields* $X_1, ..., X_k$ *satisfies the Hörmander condition iff:*

$$\mathit{span}\{[X_1, [X_2, ..., [X_{k-1}, X_k](q)\} = T_q\mathcal{N} \quad \forall q \in \mathcal{N}$$

Vectors fields satysfing the Hormander condition are also called Hörmander vector fields.

Remark 3.1.3. *The compute of all non-zero commutation relations between the vector fields (14) yields:*

$$[X_1, X_2] = -X_3$$
$$[X_2, X_3] = -X_1$$
$$[X_4, X_5] = X_3$$
$$[X_2, X_5] = -\nu X_1$$

Since X_3 is linearly independent from the horizontal frame both in $\mathcal{RT}$ and $\mathcal{M}$, the horizontal vector fields together with their commutators span the tangent space at every point. The horizontal distribution is naturally endowed with a metric structure.

Definition 3.1.3. *Let $\mathcal{H}$ be a bracket generating distribution of dimension m on a manifold $\mathcal{N}$. An horizontal scalar product is a scalar product on $\mathcal{H}$ which makes the vector fields $X_1, ..., X_m$ an orthogonal basis.*
An horizontal norm on $\mathcal{H}$ is a norm induced by an horizontal scalar product on $\mathcal{H}$.

We note that the eucliden metric on the $\mathcal{RT}$ horizontal tangent plane makes the vector fields X_1 and X_2 orthogonal. Analogously, the eucliden metric on the $\mathcal{M}$ horizontal tangent plane makes the vector fields X_1,X_2,X_4,X_5 orthogonal.

Definition 3.1.4. *A subremannian manifold is a triple $(\mathcal{N}, \mathcal{H}, g)$ where $\mathcal{N}$ is a manifold, $\mathcal{H}$ is a smooth bracket generating distribution endowed with a positive definite non degenerate metric g.*
Such a metric is called a subriemannian metric or a Carnot-Charathodory metric.

The Lie groups $\mathcal{RT}$ and $\mathcal{M}$ carries a natural structure of subriemannian manifold.

Definition 3.1.5. *Let $\mathcal{N}$ be a subriemannian manifold. An horizontal curve is a curve γ having tangent vector at each point in the horizontal distribution, i.e. an absolutely continuous curve $\gamma : [0, 1] \to \mathcal{N}$ such that $\dot{\gamma}(t) \in \mathcal{H}_{\gamma(t)}$.*

The subriemannian metric on the space enable us to define the length of a horizontal curve.

Definition 3.1.6. *Let $\gamma : [0, 1] \to \mathcal{N}$ be a horizontal curve of a subriemannian manifold $\mathcal{N}$. The length l of γ is defined as:*

$$l(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}\, dt.$$

3.2 CONNECTIVITY PROPERTY

The Hörmander condition is a meaningful geometric condition, as the following result states

Theorem 3.2.1 (Chow 1939)**.** *Any two points in a connected subriemannian manifold can be joined by a horizontal integral curve.*

When Chow theorem applies, the horizontal vector fiels together with all the commutators are able to reconstruct all the missing directions.
The CC metric and the consequences of the Chow theorem enable us to define the distance between every couple of points. The distance on the space is computed in terms of horizontal curves, in analogy with the well known Riemannian case.

Definition 3.2.1. *Let $\mathcal{N}$ be a subriemannian manifold. Let p, q be two points on $\mathcal{N}$. The C-C distance between p and q is the minimal lenght of horizontal integral curves joining the two points, i.e.*

$$d_{CC}(p, q) = \inf_{\gamma} \quad l(\gamma)$$

where the infimum is taken over all horizontal integral curves such that $\gamma(0) = p$ and $\gamma(1) = q$.

3.3 HOMOGENEOUS METRIC ON CARNOT GROUPS

Definition 3.3.1. *A Carnot group (or stratified group) $\mathbb{H}$ is a connected and simply connected Lie group whose Lie algebra $\mathfrak{h}$ admits a stratification, i.e. a direct sum decomposition:*

$$\mathfrak{g} = V_1 \oplus ... \oplus V_r$$

such that $[V_1, V_{i-1}] = V_i$ *if* $2 \leqslant i \leqslant r$ *and* $[V_i, V_r] = \{0\}$.

The stratification implies that the Lie algebra $\mathfrak{h}$ is nilpotent of step r. Every Carnot group is (isomorphic to) a homogeneous group.
Since the step r and the dimensions n_i of the subspaces V_i are indipendent from the stratification, the following definition is well-posed.

Definition 3.3.2. *The homogeneous dimension of* $\mathbb{H}$ *is the integer:*

$$Q = \sum_{i=1}^{r} i n_i$$

The homogeneous dimension Q plays the same role of Euclidean dimension in the definition of the metric structure.
Since the Lie algebra is nilpotent, the exponential map is a diffeomorphism.

The Heisenberg group $\mathbb{H}^n$, the most-studied among Carnot groups, is the space $\mathbb{R}^{2n+1}$, equipped with the composition law:

$$(x, y, z) \diamondsuit (x', y', z') = (x + x', y + y', z + z' + 2\langle y, x'\rangle - 2\langle x, y'\rangle).$$

The group $\mathbb{H}^n$ equipped with the parabolic dilatations $\delta_\lambda(z, t) = (\lambda z, \lambda^2 t), \lambda > 0$, carries the structure of homogeneous group.
A stratified basis of left invariant vector fields for the Lie algebra $\mathfrak{g}_n$ of $\mathbb{H}^n$ is given by:

$$X_j = \partial_{x_j} + 2y_j \partial_t \qquad j = 1, ..., n \tag{19}$$

$$Y_j = \partial_{y_j} - 2x_j \partial_t$$

$$T = \partial_t.$$

Let $\mathbb{G}$ be a Carnot group and, let $\mathfrak{g}$ be its stratified algebra with dilatations $\{\sigma_\lambda\}_{\lambda>0}$.

Definition 3.3.3. *A function* f *on* $\mathbb{G}$ *is homogeneous of degree* r *with respect to the dilatations* $\{\sigma_\lambda\}_{\lambda>0}$ *iff*

$$f \circ \sigma_\lambda = \lambda^r f$$

for each $\lambda > 0$.

Definition 3.3.4. *An homogeneous norm* f *on* $\mathbb{G}$ *is a continuous function* $\nu : \mathbb{G} \to [0, +\infty)$ *such that:*

(i) $\nu \in C^\infty(\mathbb{G} - 0)$;

(ii) ν *is homogeneous of degree* 1;

(iii) $\nu(x) = 0$ *iff* $x = 0$.

Moreover, ν *is symmetric if* $\nu(x) = \nu(x^{-1})$ *for all* $x \in \mathbb{G}$.

Let $d_{\mathbb{H}}$ be the control distance defined on the Heisenberg group $\mathbb{H}$. Then

$$d_0(x) := d_{\mathbb{H}}(x, 0)$$

is a symmetric homogeneous norm on $\mathbb{H}$.

Remark 3.3.1. *All the homogeneous norm on* $\mathbb{G}$ *are equivalent to the homogeneous norm:*

$$|x|_{\mathbb{G}}^{2r!} = \sum_{j=1}^{r} |x^{(j)}|^{\frac{2r!}{j}}. \tag{20}$$

Consequently the distance induced by this norm is equivalent to the CC one.

3.4 SUBRIEMANNIAN METRIC ON $\mathcal{RT}$

Even though the group $\mathcal{RT}$ is not a Carnot group, we can mimic the homogeneous structure by the choice of a suitable metric depending on the commutation properties of vector fields. The associated metric balls are squeezed in the directions of the commutators, reflecting the anisotropic nature of the distance (see [31]).

Definition 3.4.1. *Let* X *be a vector field on* $\mathbb{R}\mathcal{T}$. *Then:*

(i) X *is a vector field of degree* s *if* X *is a commutator of length* s;

(ii) The formal degree deg(X*) of* X *is the minumum integer* s *such that* X *has degree* s, *i.e.:*

$$\deg(X) = \min\{s : X = [X_{i_1}, [X_{i_2}, ..., [X_{i_{s-1}}, X_{i_s}]]]\}$$

where $i_1, ...i_s \in 1, 2, 3$.

Then the vector fields belonging to the horizontal distribution show degree 1, while the commutator X_3 is of degree 2.
Setting:

$$V_i = \{X \in \mathfrak{g} : \mathrm{deg}(X) = i\},$$

then the stratification is replaced by the direct sum decomposition:

$$\mathfrak{g} = V_1 \oplus V_2.$$

where the first layer V_1 is the horizontal distribution. In analogy with the stratified case, the homogeneous dimension is the integer:

$$Q_{\mathcal{RT}} = \sum_{i=1,2} i \dim(V_i)$$

Note that $Q_{\mathcal{RT}} = Q_{\mathbb{H}} = 4$. The horizontal norm can be extended to an homogeneous norm to the whole tangent space. In analogy with the stratified case, the norm is given by:

$$|x|^Q_{\mathcal{RT}} = \sum_{j=1}^{3} |x^{(j)}|^{\frac{Q}{\deg(X_j)}} \tag{21}$$

Note that the homogeneous dimension Q is greater than th euclidean dimension $n = 3$. In order to single out the dimension n, the calculations will be performed in terms of the exponential mapping.
Let $\xi_0 \in \mathcal{RT}$ be a fixed point and $e = (e_1, e_2, e_3)$ be the exponential coordinates of a point ξ in the neighborhood of ξ_0, such that:

$$\xi = \exp\Big(\sum_{j=1}^{3} e_j X_j\Big)(\xi_0).$$

The norm (21) induces the following distance between $\xi_0, \xi \in \mathcal{RT}$:

$$d_{\mathcal{RT}}(\xi_0, \xi) = |e|_{\mathcal{RT}} \tag{22}$$

and this distance is locally equivalent to the CC one. The metric ball centered in ξ_0 of radius $r > 0$ is denoted by:

$$B_{\mathcal{RT}} = \{\xi \in \mathcal{RT} : d_{\mathcal{RT}}(\xi_0, \xi) < r\} \tag{23}$$

Remark 3.4.1. *Since the Lie groups $\mathcal{RT}$ and $\mathbb{H}$ show the same homogeneous dimension Q, the metric balls are locally comparable. Precisely, the metric balls $B_{\mathcal{RT}}$ and $B_{\mathbb{H}}$ of radius r show the same size:*

$$|B_{\mathcal{RT}}| = |B_{\mathbb{H}}| = r^Q$$

3.5 RIEMANNIAN APPROXIMATION OF THE METRIC

A Riemannian metric on a manifold is a section of the cotangent bundle. That is, a symmetric definite positive bilinear form at every point of the tangent space is given.
A riemannian metric restricted to the horizontal distribution is a subriemannian metric. Instead a subriemannian metric is not always induced by a riemannian metric.
A direct computation shows that the subriemannian metric defined on $\mathcal{RT}$ fails to induce a riemannian metric.
Let $\|\|$ be the horizontal norm defined on the horizontal distribution which makes X_1 and X_2 an orthonormal frame. As we have observed, it is the standard euclidean norm.
We can extend it on all tangent space defining a norm of a vector as the euclidean norm if its projection on the horizontal tangent space.
Let $\xi_0 = (x_0, y_0, \theta_0)$ be a point in $\mathcal{RT}$. Let $v = (v_1, v_2, v_3)$ be a vector in $T_{\xi_0}\mathcal{RT}$ represented in the standard basis ∂_x,∂_y,∂_θ.

$$|v|_g^2 = \| \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \|^2 = \|\big(v_1\cos\theta + v_2\sin\theta, v_3\big)\|^2 =$$

$$= \big(v_1\cos\theta + v_2\sin\theta\big)^2 + v_3^2 = v_1^2\cos^2\theta + v_2^2\sin^2\theta + 2v_1v_2\cos\theta\sin\theta + v_3^2$$

The formal inverse g^{ij} of the metric is given by the singular matrix:

$$g^{ij} = \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta & 0 \\ \sin\theta\cos\theta & \sin^2\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In several problems an useful task is to look for a riemannian approximation of the metric. Here the riemannian metric is constructed adding a viscosity term in the direction X_3.
The approximate riemannian norm of the tangent vector v is given by:

$$|v|_{g_\varepsilon}^2 = \big(v_1\cos\theta + v_2\sin\theta\big)^2 + v_3^2 + \varepsilon^2\big(v_2\cos\theta - v_1\sin\theta\big)^2$$

The inverse g_ε^{ij} of the riemannian metrics defined by the invertible matrix:

$$g_\varepsilon^{ij} = \begin{pmatrix} \cos^2\theta + \varepsilon^2 \sin\theta & (1-\varepsilon^2)\sin\theta\cos\theta & 0 \\ (1-\varepsilon^2)\sin\theta\cos\theta & \sin^2\theta + \varepsilon^2\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

More generally, if $\mathcal{H} = \{X_1, ..., X_m\}$ is the horizontal distribution of a n-dimensional subriemannian manifold, then by Hörmander condition there exist commutators $X_{m+1}, ..., X_n$ such that $X_1, ...X_n$ span the tangent space at every point. We can define the approximate basis:

$$X_j^\varepsilon = X_j \quad j = 1, ..., m \tag{24}$$

$$X_j^\varepsilon = \varepsilon X_j \quad j > m \tag{25}$$

where $\varepsilon > 0$ is a parameter.
A riemannian metric g_ε on the whole tangent space is defined by requiring that the left invariant basis $\{X_j^\varepsilon\}_{i=1,...,n}$ is orthonormal.
Clearly the subriemannian inner product on $\mathcal{H}$ can be recovered by restricting the inner product $\langle\rangle_{g_\varepsilon}$ to the horizontal directions.
The approximation scheme is a classical tool in the context of stratified groups, especially in the case of Heisenberg groups.
In [7] it has been shown that the sequence of metric spaces $(\mathbb{R}^{2n+1}, d_\varepsilon)$ converges to $(\mathbb{H}^n, d)$ in the Gromov-Hausdorff sense as $\varepsilon \to 0$.

3.6 SUB-RIEMANNIAN DIFFERENTIAL OPERATORS

The sub-riemannian differential operators are defined by the vectors fields of the horizontal distribution, in analogy to the well known riemannian case.

Definition 3.6.1. *Let $\mathcal{M}$ be a subriemannian manifold. Let $\mathcal{H} = \{X_1, ..., X_m\}$ be the horizontal distribution. Let f be a function. The horizontal gradient of f is given by:*

$$\nabla_{\mathcal{H}} f = \big(X_1, ..., X_m\big)$$

Definition 3.6.2. *A function f is said to be of class $C^1_{\mathcal{H}}$ if its horizontal gradient is continuous with respect to the control distance.*

By inductive argument, a function f will belong to the class $C^k_{\mathcal{H}}$ if its horizontal gradient is of class $C^{k-1}_{\mathcal{H}}$, for $k > 1$.
Note that a function can be of class $C^1_{\mathcal{H}}$ even though it is not differentiable in the riemannian sense. However one can prove that if the horizontal distribution is bracket generating, then a function of class $C^1_{\mathcal{H}}$ is differentiable.

Definition 3.6.3. *The riemannian gradient of the function* f *referred to the approximate basis (24) is defined as:*

$$\nabla_\varepsilon f = (X_1^\varepsilon, ..., X_n^\varepsilon)$$

Formally:

$$\nabla_\varepsilon f \to \nabla_{\mathcal{H}} f$$

as $\varepsilon \to 0$.

3.7 SUBLAPLACIANS AND HEAT OPERATORS

Definition 3.7.1. *Let* $\phi = (\phi_1, ..., \phi_n)$ *be a* $C^1_{\mathcal{H}}$ *section of the tangent space of a subriemannian manifold* $\mathcal{N}$. *Let* $\mathcal{H} = (X_1, ..., X_m)$ *its horizontal distribution. The divergence of* ϕ *is defined as:*

$$div_{\mathcal{H}}(\phi) = \sum_{i=1}^{m} X_i^* \phi_i$$

where X_j^* *is the formal adjoint of the operator* X_j.

Definition 3.7.2. *The sublaplacian operator is defined as:*

$$\Delta_{\mathcal{H}} = div_{\mathcal{H}}(\nabla_{\mathcal{H}})$$

Definition 3.7.3. *The subriemannian heat operator is defined as:*

$$L_{\mathcal{H}} = \Delta_{\mathcal{H}} - \partial t$$

If the horizontal vector fields are self-adjoint, then the sublaplacian operator turns into the sum of square:

$$\Delta_{\mathcal{H}} = \sum_{i=1}^{m} X_i^2.$$

In analogy with definition (3.6.3) the riemannian approximation of the metric allows to define the riemannian counterpart of sublaplacians and second order operators.

Definition 3.7.4. *The riemannian laplacian referred to the basis (24) is defined as:*

$$\Delta^{\varepsilon}_{\mathcal{H}} = \sum_{i=1}^{m} X_i^2 + \varepsilon^2 \sum_{i=m+1}^{n} X_i^2.$$

The riemannian heat operator referred to the basis (24) is defined as:

$$L^{\varepsilon}_{\mathcal{H}} = \Delta^{\varepsilon}_{\mathcal{H}} - \partial t$$

Note that the riemannian approximate laplacian turns into the sub-laplacian as $\varepsilon \to 0$.

3.8 INTEGRAL CURVES AND ASSOCIATION FIELDS

We conclude this section with a model of Citti Sarti which proposes integral curves of the vector fields X_1, X_2 as model of association fields. The similarity between the phenomenological association field and the integral curves validates the model of subriemannian metric for the description of the structure of the visual cortex.
In [8] Citti-Sarti propose to model the local association field as a family of integral curves of the horizontal distribution $\mathcal{H} = \{X_1, X_2\}$. Such a family of curves $\gamma : [0,1] \to \mathcal{RT}$ is obtained as solution of the following ODE:

$$\dot{\gamma} = X_1(\gamma) + kX_2(\gamma)$$

with a starting condition $\gamma(0) = \xi_0 \in \mathcal{RT}$.
By varying the parameter $k \in \mathbb{R}$, we obtain the family of curves depicted in figure (8).
More explicitely, $\gamma(t) = (x(t), y(t), \theta(t))$ satisfies:

$$\begin{cases} \dot{x}(t) = \cos\theta(t) \\ \dot{y}(t) = \sin\theta(t) \\ \dot{\theta}(t) = k \end{cases} \tag{26}$$

The system describes a deterministic evolution process with the orientation changing in accordance with the parameter k. The coefficient k turns to be the curvature of the curve obtained by projecting γ on the spatial plane.

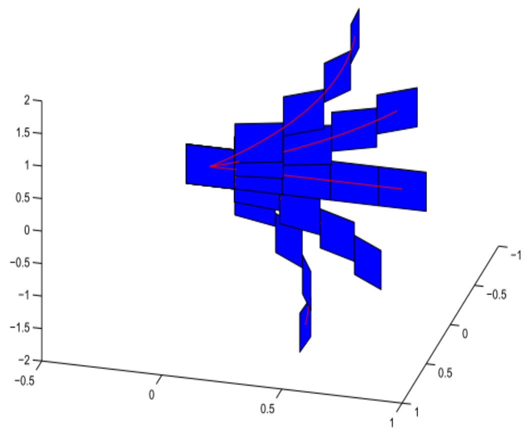


Figure 8: Integral curves of (26)

The connectivity property ensured by Chow theorem is carried on the integral curves by the exponential mapping. Starting from a fixed point $\xi_0 \in \mathcal{RT}$ every point in $\mathcal{RT}$ is riched by a curve

$$\gamma(t) = \exp\left(t(X_1 + kX_2)\right)(\xi_0).$$

Due to the non-commutativity between vector fields X_1,X_2, the non-vanishing difference between the two curves:

$$\gamma_1(t) = \exp\left(tX_1\right)\exp\left(tX_2\right)(\xi_0)$$

$$\gamma_2(t) = \exp\left(tX_2\right)\exp\left(tX_1\right)(\xi_0)$$

is estimated by the integral curve of the higher-order commutator X_3:

$$\gamma_3(t) = \exp\left(t^2X_3 + o(t^2)\right)(\xi_0).$$

The integral curves of the structure are depicted in figure (**??**). The 2D projections of the 3D integral curves fit the shape of association fields and prove that the subriemannian model is efficient and capable of predicting the direction of diffusion of the visual signal.

In analogy with the low-dimensional case, the horizontal connections between points of $\mathcal{M}$ are modeled as a family of integral curves satisfying:

$$\dot{\gamma} = a_1X_1(\gamma) + a_2X_2(\gamma) + a_4X_4(\gamma) + a_5X_5(\gamma).$$

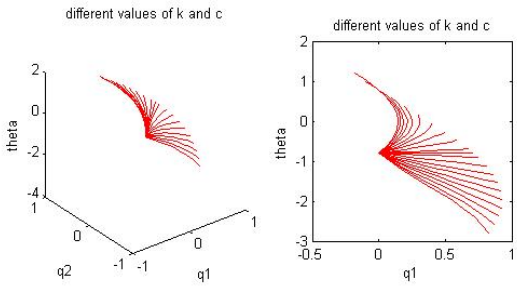


Figure 9: Visualization of the integral curves of (27) w.r.t. the variables (x, y, θ) and projection on the (x, θ) plane

for suitable coefficients a_i.
By varying the parameters a_i we can describe different spatio-temporal dynamics. In [1] two special cases of interest are described. The contour motion detected at a fixed time $(a_5) = 0$ is described by the system:

$$\dot{\gamma} = X_1(\gamma) + kX_2(\gamma) + cX_4(\gamma) \tag{27}$$

where k is a parameter of curvature and c is the rate of change of local velocity along the curve. By varying the parameters, we obtain the family of curves depicted in fig. (9).
The evolution in time can be described by the equation:

$$\dot{\gamma} = X_5(\gamma) + \omega X_2(\gamma) + aX_4(\gamma) \tag{28}$$

where the parameters ω and a represent respectively the angular velocity and the tangential acceleration. The fan of curves is depicted in fig. (10).

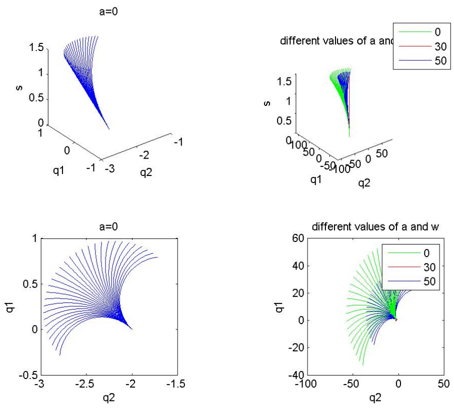


Figure 10: Visualization of the integral curves of (28) w.r.t. the variables (x, y, s) and projection on the spatial plane

4

STOCHASTIC APPROACH TO CORTICAL CONNECTIVITY

A stochastic approach to study general properties of boundaries of images was proposed by Mumford (see [30]) and applied to cortical connections in [40],[1]. The deterministic ODE system defined in chapter (3) is replaced by a system of stochastic differential equations (SDE) which takes into account the indeterminacy of the cortical connections.
In this chapter the main properties of stochastic differential equations are recalled and the transition probability of the process is recognized as the fundamental solution of the associated heat operator (see [32]). This property holds in a very general setting. When applied to the group $\mathcal{RT}$, it provides a kernel, the fundamental solution, which is interpreted as a model of cortical connectivity.

4.1 STHOCASTIC DIFFERENTIAL EQUATIONS

In order to afford a probabilistic description of the association fields, an essential background on the theory of SDE, based on the Ito calculus, is reported. The starting point of the SDE theory is to find a theoretical foundation of time-dependent differential equation:

$$\frac{dY}{dt} = b(t, Y_t) + \sigma(t, Y_t)W_t \tag{29}$$

where Y_t is a stochastic process on a probability space $(\Omega, \mathcal{F}, P)$, b,σ are given functions and W_t is a stochastic process representing noise. When $\sigma(t, Y_t) = 0$ the (29) turns to an ordinary first order differential equation. Indeed a stochastic evolution system can be view as a deterministic dynamic affected by some noise.

4.1.1 *White noise process and Brownian motion*

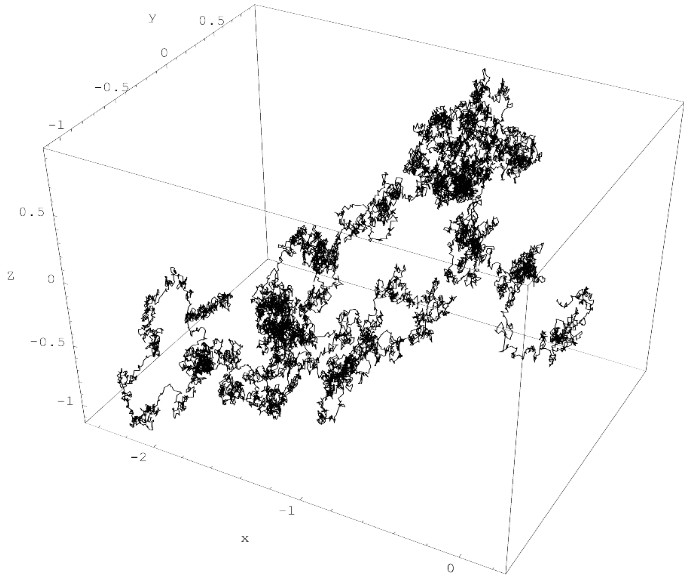


Figure 11: Brownian motion

The noise term of a SDE is classically modeled as a generalized stochastic process, called white noise process, which is constructed as the generalized derivative of a stochastic process B_t called Brownian motion or Wiener process.

Definition 4.1.1. *A Brownian motion with starting point* $y \in \mathbb{R}$ *is a real-valued stochastic process* $\{B_t\}_{t \geqslant 0}$ *satisfying the following properties:*

(i) $B_0 = y$ *almost surely;*

(ii) B_t *has independent increments, i.e. the increments* $\{B(t_i) - B(t_{i-1})\}_{i=1,\dots,k}$ *are independent random variables for all time sequences* $0 \leqslant t_1 \leqslant t_2 \leqslant \dots \leqslant t_k$*;*

(iii) the increments $B_{t+h} - B_t$ *are normally distribuited with zero mean and variance* h *for all* $t \geqslant 0, h > 0$*;*

(iv) the function $t \to B_t$ *is continuous almost surely.*

Definition 4.1.2. *A n-dimensional Brownian motion is a vector valued stochastic process:*

$$B_t = (B_t^{(1)}, \dots, B_t^{(n)})$$

where $\{B_t^{(i)}\}_{t\geqslant 0}$ *is a Brownian motion for all* $i \in \{1, ..., n\}$.

For a fixed $\omega \in \Omega$, the function defined by $t \to B_t(\omega)$ is called a path of B. The Brownian paths, which are continuous from (iv), are almost surely not differentiable at any point. Indeed, almost all Brownian paths are of unbounded variation on any time interval. As we will see, this peculiarity reflects on the definition of stochastic integral driven by Brownian motion.

4.1.2 *Ito integrals and SDE*

In order to define mathematically the differential of a stochastic process according with equation (29), one needs a notion of stochastic integral. Formally, the integral form of equation (29) is given by:

$$Y_t = Y_0 + \int_0^t b(s, Y_s)\, ds + \int_0^t \sigma(s, Y_s)\, dB_s \tag{30}$$

where the last integral is a stochastic Ito integral driven by the standard Brownian motion (starting at the origin) B_t.

A complete presentation of the Ito integral construction can be found in [32]. Here we will point out some remarkable difference with the Riemann integral. In analogy with the Riemann integral, the definition is stated for a class of elementary functions and then extended by approximation procedure. The elementary approximation of a given function f that is used to define its integral is:

$$\sum_j f(t_j^*, \omega)\chi_{(t_j, t_{j+1})}(t)$$

where f is evaluated at fixed points $t_j^* \in [t_j, t_{j+1}]$. Unlike to the Riemann case, the presence of Brownian motion implies that the choice of the point t_j^* give rise to different definitions of stochastic integral. Here we discuss the Ito choice $t_j^* = t_j$.

Following the Ito construction, we restrict to functions $\omega \to f(t_j, \omega)$ that are mesurable with respect to the natural filtration induced by B_t up to time t_j, i.e. which depend only on the past history of the Brownian process. In order to do so, we will denote with $\mathcal{F}_{t_j}$ the σ-algebra generated by random variables $\{B_s\}_{s\leqslant t_j}$, that is the σ-algebra containing the preimages of Borel sets under the random variables $\{B_s\}_{s\leqslant t_j}$. The Ito integral is defined for the class $N = N(T)$ of functions $f : [0,\infty) \times \Omega \to \mathbb{R}$ such that:

- $(t,\omega) \to f(t,\omega)$ is $\mathcal{B} \times \mathcal{F}$ measurable;
- the function $\omega \to f(t,\omega)$ is $\mathcal{F}_t$-measurable for each $t \geqslant 0$;
- $E[\int_0^T f(t,\omega)^2 dt] < \infty$.

where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^+$.
If $\phi \in N$ is an elementary function:

$$\phi(t,\omega) = \sum_j e_j(\omega)\chi_{(t_j,t_{j+1})}(t)$$

the Ito integral is defined as:

$$\int_0^T \phi(t,\omega)dB_t(\omega) = \sum_{j\geqslant 0} e_j(\omega)\big(B_{t_{j+1}} - B_{t_j}\big)(\omega). \tag{31}$$

If $f \in N$, there exists a sequence of elementary functions $\phi_n \in N$ such that:

$$E\Big[\int_0^T |f - \phi_n|^2 dt\Big] \to 0. \tag{32}$$

The existence of such a sequence is achieved using the Ito isometry for a function $\phi(t,\omega)$ elementary and bounded:

$$E[\Big(\int_0^T \phi(t,\omega)\, dB_t(\omega)\Big)^2] = E[\int_0^T \phi(t,\omega)^2\, dt]$$

It is easy to show (32) when $f \in N$ is continuous w.r.t. the first variable, for each ω. Indeed in this case choosing ϕ_n as the elementary approximation of f, for each ω the integral $\int_0^T |f - \phi_n|\, dt \to 0$ and (32) can be deduced by dominated convergence. For a refined argument that applies to general $f \in N$ see [32].

Now we are ready to define the stochastic integral for a function $f \in N$ as limit operator.

Definition 4.1.3 (Ito Integral)**.** *Let* $f \in N$ *and* $\phi_n \in N$ *be a sequence satisfying (32). The Ito integral of the function* f *is defined as:*

$$\int_0^T f(t,\omega)dB_t(\omega) = \lim_{n\to\infty} \int_0^T \phi_n(t,\omega)dB_t(\omega)$$

where the limit is intended in $L^2(\Omega, P)$.

The Ito isometry for elementary functions implies indeed that the limit exists as an element of $L^2(\Omega, P)$ and, moreover, that is independent of the sequence ϕ_n. The isometry also holds in the limit giving:

$$E[(\int_0^T f(t,\omega)\, dB_t)^2] = E[\int_0^T f^2(t\omega)\, dt].$$

Now we are ready to answer to the original question of this section. The Ito integral is used to define stochastic differential calculus.

Definition 4.1.4. *The stochastic process* Y_t *is a solution of the SDE:*

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)\, dB_t$$

if the following integral identity holds:

$$Y_t = Y_0 + \int_0^t b(s, Y_s)\, ds + \int_0^t \sigma(s, Y_s)\, dB_s.$$

Up to now no requirement about the coefficients of a SDE has been demanded. The next theorem shows, under reasonable conditions imposed on the coefficients, the well posedness of the initial value problem:

$$\begin{cases} dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dB_t \quad t \in [0, T] \\ Y_0 = y_0 \quad y_0 \in \mathbb{R}^n \end{cases} \tag{33}$$

where σ is an $n \times m$ matrix, B_t is an m-dimensional Brownian motion and we have used the shorthand notation y_0 to mean a random variable which takes value y_0 almost surely.

Theorem 4.1.1. *Let* $T > 0$ *be a fixed time. Let*

$$b : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$$

$$\sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$$

be measurable functions such that for some constants C, D *the following growth conditions hold:*

(i) $|b(t, y)| + |\sigma(t, x)| \leqslant C(1 + |y|)$*;*

(ii) $|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leqslant D|x - y|$

Then the stochastic initial value problem 33 admits a unique solution Y_t *continous in time with components belonging to* N. *where* $|\sigma| = \sum_{i,j} |\sigma_{ij}^2|, x, y \in \mathbb{R}^n, t \in [0, T]$.

Remark 4.1.1. *If the coefficients of the SDE in (4.1.1) do not depend explicitely on time, i.e. the coefficients are functions:*

$$b : \mathbb{R}^n \to \mathbb{R}^n,$$

$$\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times m},$$

then the conditions (i) *and* (ii) *of theorem (4.1.1) reduce to the requirement:*

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leqslant D|x - y|.$$

for each $x, y \in \mathbb{R}^n$.

4.1.3 *The generator of an Ito diffusion*

We will start this section defining the stochastic process which will be used to model time-evolution connectivity equations.

Definition 4.1.5. *Let* $b : \mathbb{R}^n \to \mathbb{R}^n$,$\sigma : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ *be functions satisfying the condition of remark (4.1.1). Let* B_t *be an* m*-dimensional Brownian motion. A (time-homogeneous) Ito diffusion is a stochastic process* $\{Y_t\}_{t \geqslant 0}$ *satisfying a stochastic differential equation of the form:*

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t \quad t \geqslant 0$$

with initial condition $Y_0 = y \in \mathbb{R}^n$.

The functions b and σ are called respectively the drift coefficient and the diffusion coefficient of the process.
Now we will state a milestone of stochastic calculus, which will be used to effort the main theorem of this section.

Theorem 4.1.2 (Ito formula)**.** *Let* Y_t *be a* n*-dimensional Ito diffusion:*

$$dY_t = b(Y_t)dt + \sigma(Y_t)dB_t.$$

Let $g \in C^2\left([0,\infty) \times \mathbb{R}^n, \mathbb{R}^p\right)$. *Then the stochastic process*

$$Z(t,\omega) = g(t, Y_t)$$

satisfies the stochastic chain rule:

$$dZ_I = \frac{\partial g_I}{\partial t} dt + \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}(t, Y_t) dY_{i,t} + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial y_i \partial y_j}(t, Y_t) dY_{i,t} dY_{j,t} \tag{34}$$

for each $I = 1, ..., p$.

Proof. See [32]. □

The surprising term in (34) is the second order derivative in the right side hand. This additional term is due to the non-differentiability of brownian paths.

For our purpose we need to associate a second order partial differential operator A to an Ito diffusion Y_t. The classical procedure is to define a differential operator, called the infinitesimal generator of the process, which encodes informations regarding the evolution process.

Definition 4.1.6. *The infinitesimal generator* A *of a (time homogeneous) Ito diffusion* $\{Y_t\}_{t \geqslant 0}$ *is defined by:*

$$Af(y) = \lim_{t \to 0} \frac{E^y(f(Y_t)) - f(y)}{t}$$

The following theorem shows how to build the infinitesimal generator of a given stochastic Ito process.

Theorem 4.1.3. *Let* $\{Y_t\}_{t \geqslant 0}$ *be an Ito diffusion with drift coefficient* b *and diffusion coefficient* σ. *Lef* $f \in C^2$ *a bounded function with first and second derivative bounded. Then the generator* A *of the Ito diffusion satisfies the following second order differential equation:*

$$Af(y) = \sum_i b_i(x) \frac{\partial f}{\partial_{y_i}} + \frac{1}{2} \sum_{i,j} (\sigma\sigma^T)_{i,j}(y) \frac{\partial^2 f}{\partial_{y_i} \partial_{y_j}}$$

Proof. Setting $Z_t = f(Y_t)$ the Ito formula yields:

$$
\begin{aligned}
dZ_t &= \sum_i \frac{\partial f}{\partial y_i}(Y_t)dY_{i,t} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial y_i \partial y_j}(Y_t)dY_{i,t}dY_{j,t} \\
&= \sum_i b_i \frac{\partial f}{\partial y_i}dt + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial y_i \partial y_j}(\sigma dB_t)_i(\sigma dB_t)_j \\
&+ \sum_i^j \frac{\partial f}{\partial y_i}(\sigma dB_t)_i
\end{aligned}
$$

From the properties of Brownian motion we have:

$$
(\sigma dB_t)_i(\sigma dB_t)_j = \big(\sum_k \sigma_{ik} dB_k\big)\big(\sum_l \sigma_{jl} dB_l\big) = \big(\sigma\sigma^T\big)_{ij}\, dt.
$$

Then the stochastic integral turns into:

$$
f(Y_t) = f(Y_0) + \int_0^t \big(\sum_i b_i \frac{\partial f}{\partial y_i} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial y_i \partial y_j}\big)\, ds + \sum_{i,j}\int_0^t \sigma_{ij}\frac{\partial f}{\partial y_i}\, dB_j.
$$

Hence:

$$
\begin{aligned}
E^y[f(Y_t)] = f(y) + E^y\Big[\int_0^t \big(\sum_i b_i \frac{\partial f}{\partial y_i} + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial y_i \partial y_j}\big)\, ds\Big] + \\
+ \sum_{i,j} E^y\Big[\int_0^t \sigma_{ij}\frac{\partial f}{\partial y_i}\, dB_j\Big].
\end{aligned}
$$

To get the proof we need to show that the last term of the right hand side vanishes. We recall that a family $\{f_j\}_{j\in\mathbb{N}}$ of measurable functions is called uniformly integrable w.r.t. the probability measure P iff:

$$
\lim_{l\to\infty} \sup_{j\in\mathbb{N}} \int_{\{|f_j|>l\}} |f_j| dP\} = 0.
$$

Since the first derivative of f is bounded, for all $k \in \mathbb{N}$:

$$
E^y[\int_0^{t\wedge k} \sigma_{ij}\frac{\partial f}{\partial y_i}(Y_s)\, dB_s] = E^y[\int_0^k \chi_{\{s<t\}}\sigma_{ij}\frac{\partial f}{\partial y_i}(Y_s)\, dB_s] = 0.
$$

Moreover:

$$E^y[(\int_0^{t\wedge k} \sigma_{ij} \frac{\partial f}{\partial y_i}(Y_s)\, dB_s)^2] = E^x[\int_0^{t\wedge k} (\sigma_{ij} \frac{\partial f}{\partial y_i}(Y_s))^2\, ds]$$
$$\leqslant M^2 E^y[t] < \infty$$

for a suitable constant M.
It follows that the family $\{\int_0^{t\wedge k} \sigma_{ij} \frac{\partial f}{\partial y_i}(Y_s)\, dB_s\}_{k\in\mathbb{N}}$ is uniformly integrable w.r.t. the probability measure P. As a consequence we obtain:

$$0 = \lim_{k\to\infty} E^y[\int_0^{t\wedge k} \sigma_{ij} \frac{\partial f}{\partial y_i}(Y_s)\, dB_s] =$$
$$= E^y[\lim_{k\to\infty} \int_0^{t\wedge k} \sigma_{ij} \frac{\partial f}{\partial y_i}(Y_s)\, dB_s] = E^y[\int_0^t \nu_{ij} \frac{\partial f}{\partial y_i}\, dB_j]$$

Then the expectation $E^y[f]$ turns into:

$$E^y[f(Y_t)] = f(y) + E^y\Big[\int_0^t \Big(\sum_i u_i \frac{\partial f}{\partial y_i}(Y_s) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial y_i \partial y_j}(Y_s)\Big)\, ds\Big]$$

Combining with the definition of infinitesimal generator we get the statement. □

4.2 DIFFUSION OPERATORS

Integration of a system of SDE produces random paths whose ensemble defines time dependent probability distributions. In order to study the dynamics of the system, it is convenient to consider the time evolution of these probability distribution.

Theorem 4.2.1 (Kolmogorov backward equation)**.** *Let Y_t an n-dimensional Ito diffusion with generator A. Let $f \in C_0^2$. Define:*

$$u(t, y) = E^y[f(Y_t)]$$

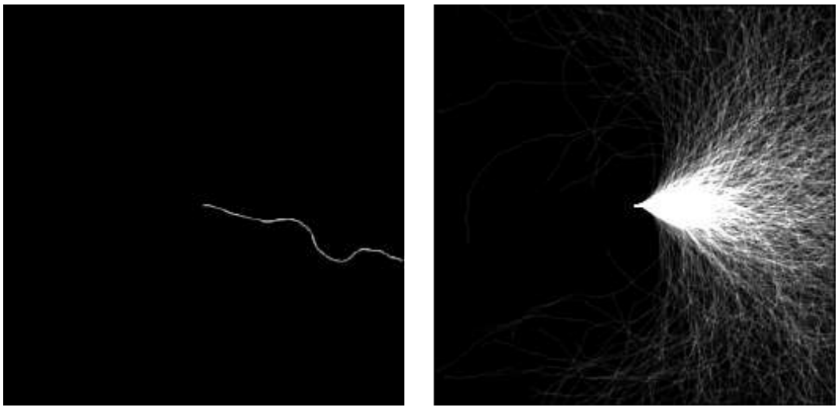
Then u satisfies the following second order differential equation:

$$\partial_t u = Au \tag{35}$$

Proof. See [32]. □

The equation (35) is called the Kolmogorov backward equation of the process. The forward equation is obtained by replacing the operator A with its adjoint A^* (see [32]).

4.3 STOCHASTIC MODELS OF CONNECTIVITY



(a) Random walk (b) 1000 random walks

Figure 12: Mumford stochastic process (from [40])

Now we will go on the construction of the probabilistic association fields. In the previous section a general method to relate a diffusion process to a second order partial differential operator has been given. The fundamental assumption we will make is that the prior probability distribution can be modeled by a random walk. The stochastic counterpart of the deterministic process (26) is the Mumford process:

$$\begin{cases} \dot{x}(t) = \cos\theta(t) \\ \dot{y}(t) = \sin\theta(t) \\ \dot{\theta}(t) = dB(t) \end{cases} \tag{36}$$

with starting point $\xi_0 = (0, 0, \theta_0) \in \mathcal{RT}$ at time $t_0 = 0$.

The process can be written in terms of the horizontal vector fields:

$$d\gamma(t) = X_1(\gamma)dt + X_2(\gamma)dB(t)$$

where $\gamma(t) = (x(t), y(t), \theta(t))$ with $\gamma(0) = \xi_0$.
These equations describe the motion of a particle moving with constant velocity in a direction randomly changing in accordance with the stochastic process $B(t)$. Here the cortical connectivity is modeled as a variant of (36). Our purpose is to estimate the density of points $(x(T), y(T), \theta(T))$ reached at a fixed time $T > 0$ by the connectivity process starting when $t_0 = 0$ at point $(0, 0, \theta_0)$. A single random walk with start at position $(0, 0, \theta_0)$ at time t_0 and end at position $(x(T), y(T), \theta(T))$ gives the probability of transition from $(0, 0, \theta_0)$ to $(x(T), y(T), \theta_(T))$. Under the previous assumptions, the density function is evaluated by evolving in time the process from the initial state up to time T.

A variant of the stochastic connectivity process (36) is considered by evolving both the vector fields X_1 and X_2 in accordance with two independent Brownian motions. Then the dynamic is driven by a family of horizontal curves $\gamma(t) = (x(t), y(t), \theta(t))$ satisfying:

$$\dot{\gamma}(t) = X_1(\gamma(t))dB_1(t) + X_2(\gamma(t))dB_2(t)$$

or equivalently:

$$\begin{cases} \dot{x}(t) = \cos\theta(t)dB_1(t) \\ \dot{y}(t) = \sin\theta(t)dB_1(t) \\ \dot{\theta}(t) = dB_2(t) \end{cases} \tag{37}$$

This technique can be directly implemented, and the fundamental solution is obtained solving numerically the system of stochastic differential equations (37) and applying standard Markov Chain Monte Carlo methods. The results are depicted in figure (13). By theorem (4.2.1), the probability density evolves in time in accordance with the heat equation. Precisely, the density $p(\xi, t|\xi_0, 0)$ of points reached at time t by the sample paths of the process (37) is obtained as the fundamental solution of the operator:

$$\partial_t - L \tag{38}$$

where L is the Laplacian $L = X_1^2 + X_2^2$. The problem of local estimate of the fondamental solution will be discussed in chapter (5).

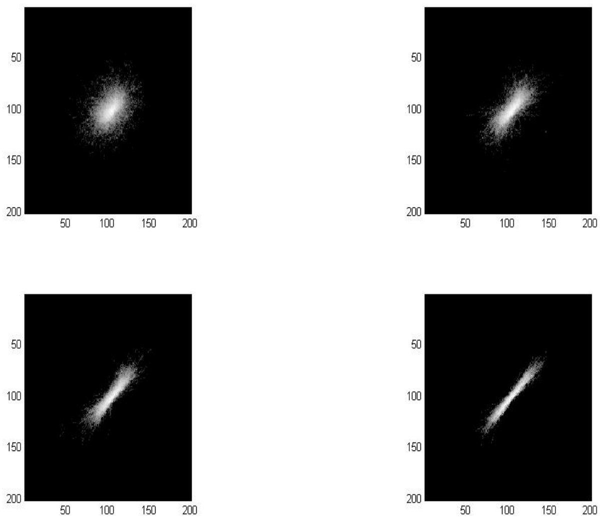


Figure 13: Heat kernels on $\mathcal{RT}$ with initial point $\xi_0 = (0, 0, \frac{\pi}{4})$ and decreasing values of variance.

5

DIFFUSION KERNELS ESTIMATES

As we have seen in chapter (4), the fundamental solution of a parabolic second order differential operator. In the special case of the subriemannian visual cortex, we have considered a diffusion driven by the subriemannian heat operator. In this chapter we will point out a picture of the fundamental solution of the subriemannian heat operator on the group $\mathcal{RT}$. A local estimate of the fundamental solution will be achieved by lifting vector fields in the Heisenberg group $\mathbb{H}$. The explicit expression of the fundamental solution of the heat operator on $\mathbb{H}$ will yield a parametrix of the fundamental solution of the heat operator on $\mathcal{RT}$.

5.1 FUNDAMENTAL SOLUTION OF SECOND ORDER HYPOELLIPTIC DIFFERENTIAL OPERATORS

In this section we will point discuss the existence of the fundamental solution of the subelliptic heat operator:

$$L = \Delta_{\mathcal{RT}} - \partial_t = X_1^2 + X_2^2 - \partial_t \tag{39}$$

where X_1, X_2 are the horizontal self adjoint vector fields defined in remark (3.1.1).
The operator L belongs to a wide class of second order hypoelliptic operators, namely the Hörmander type operators. As we will see, these operators admit a fundamental solution, even if they are affected by strong degeneracy.
A presentation of the essential background concerning fundamental solutions of hypoelliptic operators is reported below.

Let $\mathbb{G}$ be a Lie group. We will use the standard Schwartz notation $\mathcal{D}(\mathbb{G}), \mathcal{D}'(\mathbb{G})$ for the space $C_0^\infty(\mathbb{G})$ of test functions and its dual space of distributions. The pairing of $\tau \in \mathcal{D}'$ with $u \in \mathcal{D}$ will be indicated $\langle\tau|u\rangle$.

We recall that in this setting the convolution of $\tau \in \mathcal{D}'(\mathbb{G})$ and $f \in \mathcal{D}(\mathbb{G})$ is given by:

$$\tau * f(x) = \langle\tau|l_x\check{f}\rangle = \int_{\mathbb{G}} \tau(y)f(y^{-1} \odot x)dy$$

where l_x is the lef translation on $\mathbb{G}$, $\check{f}(x) = f(x^{-1})$, $\odot$ is the group law, dy is the Haar measure on $\mathbb{G}$. Taking into account the time dependence, the spatio-temporal convolution of $\tau \in \mathcal{D}'(\mathbb{G} \times [0,T])$ and $f \in \mathcal{D}(\mathbb{G} \times [0,T])$ is given by:

$$\tau * f(x,t) = \langle\tau|l_x\check{f}\rangle = \int_{\mathbb{G}} \tau(y,s)f(y^{-1} \odot x, t-s)\, dy\, ds$$

where ds is the Lebesgue measure on the real line.

Definition 5.1.1. *A distribution $E \in \mathcal{D}'(\mathbb{G} \times [0,T])$ is called a fundamental solution of 39 iff*

$$L(E) = \delta$$

wheere δ is the dirac mass defined by $\langle\delta, u\rangle = u(0)$ for $u \in \mathcal{D}(\mathbb{G} \times [0,T])$.

Remark 5.1.1. *If E is a fundamental solution for the operator L, then $L * g$ is a solution u of the differential equation $Lu = f$.*

In accordance with the findings of previous section, we will look for solutions which can be expressed in terms of Green kernels.

Let Ω_{x_1},$\Omega_{x_2} \subset \mathbb{G}$ open sets. A function $k \in C(\Omega_{x_1} \times \Omega_{x_2})$ defines an integral operator from $C_0(\Omega_{x_2})$ to $C(\Omega_{x_1})$ given by

$$(\mathcal{K}\phi)(x_1) = \int k(x_1,x_2)\phi(x_2)dx_2 \qquad \phi \in C_0(\Omega_{x_2}), x_1 \in \Omega_{x_1}.$$

The definition can be extended to arbitrary distributions $k \in \mathcal{D}(\Omega_{x_1} \times \Omega_{x_2})$, as the following theorem states.

Theorem 5.1.1 (Schwartz Kernel Theorem)**.** *A distribution $k \in \mathcal{D}'(\Omega_{x_1} \times \Omega_{x_2})$ defines a linear map $\mathcal{K}$ from $\mathcal{D}(\Omega_{x_2})$ to $\mathcal{D}'(\Omega_{x_1})$, according to:*

$$\langle\mathcal{K}\phi, \psi\rangle = k(\psi \otimes \phi) \quad \psi \in \mathcal{D}(\Omega_{x_1}), \phi \in \mathcal{D}(\Omega_{x_2}), \tag{40}$$

where $\psi \otimes \phi$ is the tensor product of ψ and ϕ. Such a linear map is continuous in the sense that $\mathcal{K}\phi_j \to 0$ in $\mathcal{D}'(\Omega_{x_1})$ when $\phi_j \to 0$ in $\mathcal{D}(\Omega_{x_2})$. Conversely, to every such linear map $\mathcal{K}$ there is one and only one distribution k such that (40) is valid. k is called the kernel of $\mathcal{K}$.

From now on we we refer to fundamental solutions which are fundamental kernels of (39), i.e. distributions $K \in \mathcal{D}(\mathbb{G} \times \mathbb{G} \times [0,T])$ such that:

$$D_{x_1} K(x_1, x_2, t, \tau) = \delta({x_2}^{-1} \odot x_1, t - \tau).$$

where D_x means that D acts on the variable x.

The hypoellipticity answer the problem of existence of a fundamental solution of a given differential operator, as the theorem (5.1.2) states.

Definition 5.1.2. *If the vector fields $X_1, \cdots, X_m$ are of Hörmander type and selfadjoint, the associated subriemannian Laplacian and the heat operator (also called subelliptic Laplacian and Heat) are called Hörmander type operators.*

Definition 5.1.3. *A differential operator L is called hypoelliptic if for every distribution $u \in \mathcal{D}'(\Omega)$, u is C^∞ in every open set where Lu is C^∞.*

Theorem 5.1.2. *A second order differential operator L is hypoelliptic if and only if there exist a fundamental solution of L which is C^∞ outside the origin.*

The following theorem states the existence of a fundamental solution for the operator (39).

Theorem 5.1.3. *(Hörmander 1967) An Hörmander operator is hypoelliptic.*

The celebrated Hörmander theorem states that we can control the the characteristic direction of a differential operators via conditions involving commutators of vector fields. We recall that the vector fields X_1, X_2, X_3 do not depend explicitly on time, being the generators of the tangent space to $\mathcal{RT}$ at every point. When Hörmander theorem applies the existence of a fundamental solution is guarantee, even though its explicit expression might be unknown.
Neverthless, also in this case, we can provide uniform estimates of the fundamental solution and all derivatives in terms of the metric balls measures.
The main difficult here is that $\mathcal{RT}$ is not equipped with a homogeneous structure. In the following, we will provide an approximate of

the sub-riemannian heat kernel via parametrice method described in [37].
The fundamental solution of L can be locally estimated with a suitable adaptation of the freezing method.
If L is an elliptic operator, the frozen approximating operator is simply obtained evaluating the coefficients of L at a fixed point. Here the approximate is achieved by a first order Taylor expansion of the coefficients of L. The Taylor developement is computed in the directions prescribed by the vector fields $X_{1,}, X_2, X_3$.
Firstly we will define a frame of first order operators with polynomial coefficients which locally approximate the vector fields X_1,X_2,X_3 in the neighborhood of a given point $\xi_0 = (x_0, y_0, \theta_0) \in \mathcal{RT}$.
The resulting Lie algebra will admit a 2-step stratification. As a consequence, the approximation will be performed in a Carnot group of dimension three, namely a group isomorphic to the Heisenberg roup. It will be shown that, up to a change of variables, the approximate operator is the heat operator attached to the Kohn laplacian on the Heisenberg group, which will mimic the subelliptic heat operator on the $\mathcal{RT}$ group. The fundamental solution of the heat operator on the Heisenberg group is explicitly known and much about its properties has been already studied.
The explicit fundamental solution Γ_{ξ_0} of the approximate operator will be the the candidate local approximation of the fundamental solution Γ of the operator L.

5.1.1 *Heat kernels on Carnot groups*

Let

$$L_{\mathbb{G}} = \Delta_{\mathbb{G}} - \partial t \tag{41}$$

be the heat operator defined on the Carnot group $\mathbb{G}$. The heat kernel $\Gamma_{\mathbb{G}}$ defined on $\mathbb{G} \times \mathbb{G} \times [0, T]$, like its sublaplacian, satisfies some homogeneity properties, which can be stated in terms of the homogeneous dimension Q.

Definition 5.1.4. *A differential operator* D *is homogeneous of degree* r *w.r.t. the dilatations* $\{\sigma_\lambda\}$ *iff*

$$D(u \circ \sigma_\lambda) = \lambda^r (Du) \circ \sigma_\lambda$$

for all $u \in \mathcal{D}$ *and* $\lambda > 0$.

Definition 5.1.5. *A distribution* $\tau \in \mathcal{D}'(\mathbb{G})$ *is a kernel of type* α *iff*

(i) $\tau \in C^\infty((\mathbb{G} - 0))$;

(ii) τ *is homogeneous of degree* $\alpha - Q$.

Remark 5.1.2. *If D is homogeneous of degree* r *and* K *is a kernel of type* α, *then* DK *is a kernel of type* $\alpha - \lambda$.

Definition 5.1.6. *(Parabolic dilatations)* $\mathbb{G} \times [0, T]$ *is equipped with the family* $\{\sigma^L_\lambda\}_{\lambda>0}$ *of parabolic dilatations given by:*

$$\sigma^L_\lambda(\xi, t) = (\sigma_\lambda(x), \lambda^2 t)$$

where $\{\sigma_\lambda\}_{\lambda>0}$ *is a family of homogeneous dilatations on* $\mathbb{G}$.

Theorem 5.1.4. *There exists a smooth function* Γ *on* $\mathbb{G} \times [0, T] - 0$ *such that the fundamental solution of (41) is given by:*

$$\Gamma_{\mathbb{G}}(\xi, t, \eta, s) := \Gamma_{\mathbb{G}}(\eta^{-1} \circ \xi, t - s).$$

The kernel $\Gamma_{\mathbb{G}}$ *satisfies:*

(i) $\Gamma_{\mathbb{G}}$ *is homogoeneous of degree* $-Q$ *w.r.t. the parabolic dilatations* $\{\sigma^L_\lambda\}_{\lambda>0}$, *i.e:*

$$\Gamma_{\mathbb{G}}(\sigma^L_\lambda(\xi, t)) = \lambda^{-Q} \Gamma_{\mathbb{G}}(\xi, t);$$

(ii) $\Gamma_{\mathbb{G}}(\xi, t, \eta, s) = \Gamma_{\mathbb{G}}(\eta, -s, \xi, t) = \Gamma_{\mathbb{G}}(\eta^{-1} \circ \xi, t - s, 0, 0)$;

(iii) *There exists a positive constant* C *such that:*

$$\Gamma_{\mathbb{G}}(\xi, t) \leqslant C\left(d_{\mathbb{G}}(\xi, 0) + |t|^{\frac{1}{2}}\right)^{-Q},$$

where $d_{\mathbb{G}}$ *is a control distance on* $\mathbb{G}$ *and* Q *is the homogeneous dimension of* G;

(iv) (Reproduction property)

$$\Gamma_{\mathbb{G}}(\xi, t+s) = \int_{\mathbb{G}} \Gamma_{\mathbb{G}}(\eta^{-1} \circ \xi, t)\Gamma(\eta, s)\, d\eta$$

for every $\xi \in \mathbb{G}, t > 0, s > 0.$

Proof. See [5]. □

Then Γ_G is a kernel of type 0 w.r.t. parabolic dilatations (5.1.6). From the homogeneity behavioural of the fundamental solution, uniform gaussian estimates in terms of the control distance d_G are inferred.

Theorem 5.1.5 (Uniform Gaussian estimates)**.** *There exist positive constants* $C_1, C_{p,q}$ *such that the fundamental solution* Γ_G *of the heat operator* $L_{\mathbb{G}}$ *and its derivatives satisfies:*

(i) $|\Gamma_G(x,t)| \leqslant C_1 t^{-\frac{Q}{2}} \exp\left(\frac{-d^2_{\mathbb{H}}(\xi,0)}{C_1 t}\right)$;

(ii) $|X_{i_1}...X_{i_p}(\partial t)^q \Gamma(\xi,t)_G| \leqslant C_{p,q} t^{-\frac{Q+p+2q}{2}} \exp\left(\frac{-d^2_{\mathbb{H}}(\xi,0)}{C_1 t}\right)$;

for every $p, q \in \mathbb{N}$, $i_1, ..., i_p \in \{1,2,3\}$.

Proof. See [5]. □

As the above theorem shows, we can control the homogeneity of the fundamental solution and its derivatives in terms of metric balls defined by vector fields.
If $B_G(\xi, r)$ is the metric ball induced by the control distance d_G, with center in $\xi \in \mathbb{G}$ and radius $r > 0$, we define the exponential term:

$$E(\xi, t) = \frac{\exp\left(-\frac{C\, d^2_G(\xi,0)}{t}\right)}{|B_G(0, \sqrt{t})|}.$$

where C is a positive constant.

Remark 5.1.3. *There exist a positive constant* M *such that the following inequality holds:*

$$d_G(\xi,0)E(\xi,t) = \sqrt{t}\frac{d_G(\xi,0)}{\sqrt{t}} \frac{\exp\left(-\frac{C\, d^2_G(\xi,0)}{t}\right)}{|B_G(0,\sqrt{t})|} \tag{42}$$

$$\leqslant M\sqrt{t}\frac{\exp\left(-\frac{C\, d^2_G(\xi,0)}{2t}\right)}{|B_G(0,\sqrt{t})|} \tag{43}$$

$$\leqslant M\sqrt{t}E(\xi, 2t) \tag{44}$$

As a consequence, the gaussian estimates of theorem (5.1.5) can be refurmulated as:

(i) $|\Gamma_G(\xi, 0, t)| \leqslant C_0 E(\xi, t)$

(ii) $|X_i \Gamma_G(\xi, \eta, t)| \leqslant \frac{C_1}{\sqrt{t}} E(\xi, t)$

for suitable positive constants C_0, C_1.

5.2 APPROXIMATE VECTOR FIELDS

We will look towards the homogeneity behavioural of vector fields in $\mathcal{RT}$ in terms of the homogeneity properties of the stratified group of same dimension, namely the Heisenberg group $\mathbb{H}$.
Let $X_1, X_2, X_3 \in \mathcal{RT}$ the vector fields defined in section (2.2). Let $\xi_0 = (x_0, y_0, \theta_0) \in \mathcal{RT}$ be a fixed point.

Definition 5.2.1. *The vector fields frozen at ξ_0 are defined as:*

$$X_{1,\xi_0} = \cos\theta_0 - (\theta - \theta_0 \sin\theta_0 \partial_x + \sin\theta_0 + (\theta - \theta_0) \cos\theta_0) \partial_y$$
$$X_{2,\xi_0} = X_2 = \partial_\theta$$
$$X_{3,\xi_0} = -\sin\theta_0 \partial_x + \cos\theta_0 \partial_y$$

The vector fields $\{X_{i\xi_0}\}_{i=1,2,3}$ are Hörmander vector fields defined in terms of the first order Taylor development of the coefficients of the vector fields $\{X_i\}_{i=1,2,3}$.

Remark 5.2.1. *The only non-zero commutation relation between the frozen vector fields is:*

$$[X_{1,\xi_0}, X_{2,\xi_0}] = X_{3,\xi_0}$$

Then the Lie algebra spanned by $X_{1,xi_0}, X_{2,\xi_0}, X_{3,\xi_0}$ is nilpotent of step two. Up to isomorphism, it is the Heisenberg algebra defined in (19).

5.2.1 *Exponential coordinates turning the approximate vector fields into the Heisenberg algebra*

Let $\xi = (x, y, \theta)$ a point in the neighborhood of ξ_0. Let $e = (e_1, e_2, e_3)$ the exponential coordinates of ξ with respect to the basis $X_{1\xi_0}, X_{2\xi_0}, X_{3\xi_0}$, such that:

$$\xi = \exp\left(e_1 X_{1\xi_0} + e_2 X_{2\xi_0} + e_3 X_{3\xi_0}\right)(\xi_0)$$

By definition of exponential mapping, $\xi = (x, y, \theta) = \gamma(1)$, where $\gamma(t) = (\gamma_1(t), \gamma_2(t), \gamma_3(t))$ solves the Cauchy problem:

$$\begin{cases} \dot{\gamma} = e_1 X_{1\xi_0} + e_2 X_{2\xi_0} + e_3 X_{3\xi_0} \\ \gamma(0) = \xi_0 \end{cases}$$

A trivial computation yields:

$$\gamma_1(1) = x = x_0 + \cos\theta_0 e_1 - \frac{1}{2}\sin\theta_0 e_2 e_1 - \sin\theta_0 e_3 \tag{45}$$

$$\gamma_2(1) = y = y_0 + \sin\theta_0 e_1 + \frac{1}{2}\cos\theta_0 e_1 e_2 + \cos\theta_0 e_3 \tag{46}$$

$$\gamma_3(1) = \theta = \theta_0 + e_2 \tag{47}$$

Plugging-in $e_2 = \theta - \theta_0$, one obtain the linear map:

$$\begin{pmatrix} y - y_0 \\ x - x_0 \end{pmatrix} = \begin{pmatrix} \sin\theta_0 + \frac{1}{2}(\theta - \theta_0)\cos\theta_0 & \cos\theta \\ \cos\theta_0 - \frac{1}{2}(\theta - \theta_0)\sin\theta_0 & -\sin\theta_0 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

The inverse is:

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \sin\theta_0 & \cos\theta_0 \\ \cos\theta_0 - \frac{1}{2}(\theta - \theta_0)\sin\theta_0 & -\sin\theta_0 - \frac{1}{2}(\theta - \theta_0)\cos\theta_0 \end{pmatrix} \begin{pmatrix} y - y_0 \\ x - x_0 \end{pmatrix}.$$

The regular change of variables

$$\begin{cases} e_1 = \sin\theta_0 (y - y_0) + \cos\theta_0 (x - x_0) \\ e_2 = \theta - \theta_0 \\ e_3 = \cos\theta_0 (y - y_0) - \sin\theta_0 (x - x_0) - \frac{1}{2} e_1 e_2 \end{cases} \tag{48}$$

provides a diffeomorphism:

$$\Phi_{\xi_0}(\xi) = \Phi_{x_0, y_0, \theta_0}(x, y, \theta) = (e_1, e_2, e_3) \tag{49}$$

turning the coordinates (x, y, θ) into the coordinates (e_1, e_2, e_3).
For a function $u : \mathcal{RT} \to \mathbb{R}$ we set

$$u_{\mathbb{H}} = u \circ \Phi_{\xi_0}^{-1}.$$

Using the diffeomorphism, we can express the distance between two points in terms of the control distance attached to the Heisenberg group.

Definition 5.2.2. *The distance attached to the frozen metric in the neighborhood of $\xi_0 \in \mathcal{RT}$ is given by:*

$$d_{\xi_0}(\xi, \eta) = d_{\mathbb{H}}(\Phi_{\xi_0}(\xi), \Phi_{\xi_0}(\eta))$$

The metric ball induced by the distance d_{ξ_0}, with center in $\xi \in \mathcal{RT}$ and radius $r > 0$, will be denoted by $B_{\xi_0}(\xi, r)$.

5.2.2 *Approximate heat kernel*

Definition 5.2.3. *The approximate subelliptic heat operator is defined as:*

$$L_{\xi_0} = X_{1,\xi_0}^2 + X_{2,\xi_0}^2 - \partial_t.$$

This operator admits an explicit fundamental solution Γ_{ξ_0} which, up to the change of variables 48, is the well-known fundamental solution of the Heisenberg heat operator.
If

$$L_{\mathbb{H}} = \Delta_{\mathbb{H}} - \partial t = X_{\mathbb{H}}^2 + Y_{\mathbb{H}}^2 - \partial_t$$

is the canonical Heisenberg heat operator in the standard basis (19), the related fundamental solution $\Gamma_{\mathbb{H}}$ is given by:

$$\Gamma_{\mathbb{H}}((x, y, \theta), t) = \frac{1}{(2\pi t)^4} \int \cos\left(\frac{y\tau}{t}\right) \exp\left(\frac{1}{2}\tau \coth 2\pi \frac{x^2 + \theta^2}{t}\right) \frac{2\tau^4}{\sinh(2\tau)^4}, d\tau \tag{50}$$

The fundamental solution Γ_{ξ_0} satisfies the gaussian estimates prescribed by theorem (5.1.5), in terms of the distance d_{ξ_0} defined in (5.2.2):

(i) $|\Gamma_{\xi_0}(\xi, t)| \leqslant C_0 E(\xi, t)$;

(ii) $|X_{i\xi_0}\Gamma_{\xi_0}| \leqslant \frac{C_1}{\sqrt{t}}E(\xi, t)$.
$i = 1, 2, 3$

where $E(\xi, t)$ is the exponential term:

$$E(\xi, t) = \frac{\exp\left(-\frac{C d^2_{\xi_0}(\xi, 0)}{t}\right)}{|B_{\xi_0}(0, \sqrt{t})|}.$$

In order to make a comparison between the frames $\{X_i\}_{i=1,2,3}$ and $\{X_{i,\xi_0}\}_{i=1,2,3}$, we will now introduce a definition related to kernels on $\mathcal{RT}$.

Definition 5.2.4. *A distribution* $K \in \mathcal{D}'(\mathcal{RT})$ *is a kernel of class* F_λ *iff*

(i) $k \in C_0^\infty(\mathcal{RT} - \{0\})$*;*

(ii) $|X_{j_1\xi_0}...X_{j_s\xi_0}k(\xi, t)| \leqslant C_s t^{\frac{\lambda - s}{2} - 1}E(\xi, t)$.

where C_s *is a constant and* $j_1, ..., j_s \in \{1, 2, 3\}$.

For istance, a fundamental solution Γ belongs to the class F_2, while its gradient belongs to the class F_1.

5.3 PARAMETRICES METHOD

The parametrices method is a classical iterative method which will provide estimates of the fundamental solution Γ by homogeneity arguments. The following cumputations are inspired by [37],[24]. The paper [37] describes a procedure to lift a frame of vector fields into a (model) Carnot group. As a result, the fundamental solution of the sum of squares defined by these vector fields is written as a parametrix of the fundamental solution of the lifted operator. In [24], the method is adapted to the parabolic case. The following results are very close to the contents of lemma 3 and theorem 2 in [37].
We will start with some inequalities achieved by homogeneity arguments.
From now on we adopt the easier notation $k = O(f)$ meaning that there exists a constant $C > 0$ such that $|k| \leqslant C|f|$.

The $X_{i\xi_0}$'s appriximate the X_i's in the sense of the following lemma.

Lemma 5.3.1. *There exists constants* C_1, C_2 *such that the following estimates hold:*

(i) $|(X_i - X_{i\xi_0})\Gamma_{\xi_0}(\xi, t)| \leqslant C_0 E(\xi, t)$;

(ii) $|X_i \Gamma_{\xi_0}(\xi, t)| \leqslant \frac{C_1}{\sqrt{t}} E(\xi, t)$.

for $i = 1, 2$.

Proof. When $i = 2$ the statement is trivial. Let $i = 1$.

$$
\begin{aligned}
&(X_1 - X_{1\xi_0})\Gamma_{\xi_0}(\xi, t) = \\
&= \big(\cos\theta - \cos\theta_0 + (\theta - \theta_0)\sin\theta_0\big)\partial_x \Gamma_{\xi_0}(\xi, t) + \\
&+ \big(\sin\theta - \sin\theta_0 - (\theta - \theta_0)\cos\theta_0\big)\partial_y \Gamma_{\xi_0}(\xi, t) = \\
&= \big(\cos\theta - T_{1,\theta_0}\cos\theta\big)\big(\cos\theta X_{1\xi_0} - \big(\sin\theta_0 + (\theta - \theta_0)\cos\theta_0\big)X_{3\xi_0}\big)\Gamma_{\xi_0}(\xi, t) + \\
&+ \big(\sin\theta - T_{1,\theta_0}\sin\theta\big)\big(\sin\theta X_{1\xi_0} - \big(\cos\theta_0 + (\theta - \theta_0)\sin\theta_0\big)X_{3\xi_0}\big)\Gamma_{\xi_0}(\xi, t) = \\
&= O(\Gamma_{\xi_0}(\xi, \eta))
\end{aligned}
$$

where $T_{1,\theta_0} f(\theta) = f(\theta_0) + f'(\theta_0)(\theta - \theta_0)$ is the first order truncated Taylor series of f at θ_0.
As the above computation shows, the vector field $X_1 - X_{1\xi_0}$ applied on Γ_{ξ_0} shows the same homogeneity of Γ_{ξ_0}. As a consequence, it acts as a derivative of order zero. From the comparison with theorem (5.1.5) we conclude (ii). □

The operator L_0 approximates the operator L in the sense of the following lemma.

Lemma 5.3.2. *There exists constants* C_1, C_2 *such that the following estimates hold:*

(i) $|X^2_{i\xi_0}\Gamma_{\xi_0}(\xi, t)| \leqslant \frac{C_0}{t} E(\xi, t)$;

(ii) $|(L - L_{\xi_0})\Gamma_{\xi_0}(\xi, t)| \leqslant \frac{C_1}{\sqrt{t}} E(\xi, t)$.

for $i = 1, 2$.

Proof.

$$
\begin{aligned}
&(X_i^2 - X_{i\xi_0}^2)\Gamma_{\xi_0}(\xi,t) = \\
&= \big(X_i^2 - X_i X_{i\xi_0} + X_i X_{i\xi_0} - X_{i\xi_0}^2\big)\Gamma_{\xi_0}(\xi,t) = \\
&= X\big(X - X_{\xi_0}\big)\Gamma_{\xi_0}(\xi,t) + \big(X - X_{\xi_0}\big)X_{\xi_0}\Gamma_{\xi_0}(\xi,t) = \\
&= \big(X - X_{\xi_0} + X_{\xi_0}\big)\big(X - X_{\xi_0}\big))\Gamma_{\xi_0} + \big(X - X_{\xi_0}\big)X_{\xi_0}\Gamma_{\xi_0} = \\
&= \big(X - X_{\xi_0}\big)^2\Gamma_{\xi_0} + X_{\xi_0}\big(X - X_{\xi_0}\big)\Gamma_{\xi_0}(\xi,t) + \big(X_{\xi_0}\big)X_{\xi_0}\Gamma_{\xi_0}(\xi,t) = \\
&= O\big(\frac{1}{\sqrt{t}}E(\xi,t)\big)
\end{aligned}
$$

□

We conclude that $L - L_{\xi_0}$ acts on Γ_{ξ_0} as a differential operator of order 1.

Theorem 5.3.1. *If* L *is the subelliptic heat operator (39) and* Γ_{ξ_0} *is the fundamental solution for Heisenberg heat operator (5.2.3), then there exists a kernel* H_1 *such that:*

(i) $L\Gamma_{\xi_0}(\xi,t) = \delta(\xi,t) + H_1(\xi,t)$*;*

(ii) $|H_1(\xi,t)| \leqslant C\frac{E(\xi,t)}{\sqrt{t}}$*;*

where C *is a positive constant.*

Proof. As a direct consequence of lemma (5.3.2) we have:

$$L\Gamma_{\xi_0} = L_{\xi_0}\Gamma_{\xi_0} + \big(L - L_{\xi_0}\big)\Gamma_{\xi_0} = \delta + H_1$$

where $H_1 = \big(L - L_{\xi_0}\big)\Gamma_{\xi_0} = O\big(\frac{E(\xi,t)}{\sqrt{t}}\big)$. □

Γ_{ξ_0} is an approximate of the fundamental solution Γ in the sense of theorem (5.3.1). Indeed we have:

$$L\Gamma = \delta$$

$$L\Gamma_{\xi_0} = \delta + H_1$$

As a consequence, the kernel H_1 is a remainder term.
The parametrices method yields an iterative procedure to improve a better approximation. In order to expose the procedure, we begin with a theorem related to the behavioural of kernels under convolutions.

Theorem 5.3.2. *Let* H_1, H_s *be kernels such that:*

(i) $|H_\beta(\xi, t)| \leqslant C_0\big(t^\beta E(\xi, t)\big)$;

(ii) $|H_s(\xi, t)| \leqslant C_1\big(t^s E(\xi, t)\big)$;

for positive constants C_0, C_1.
Then there exists a positive constant M *such that:*

$$|H_\beta * H_s(\xi, t)| \leqslant M\big(t^{\beta+s+1} E(\xi, t)\big).$$

Proof. The reproduction property of theorem (5.1.4) yields:

$$\begin{aligned}
&H_\beta * H_s(\xi, t) = \\
&= \int \Gamma_{\xi_0}(\xi, \eta, t, \tau)\Gamma_{\xi_0}(\eta, \xi, \tau, \sigma)(t-\tau)^\beta(\tau-\sigma)^s \, d\eta \, d\tau \\
&= \Gamma_{\xi_0}(\xi, \eta, t, \sigma)\int (t-\tau)^\beta(\tau-\sigma)^s \, d\tau
\end{aligned}$$

since $\tau \in [0, T]$ by $t - \tau > 0, \tau - \sigma > 0$.
If we set $(\tau - \sigma) = (t - \sigma)r$, then

$$\begin{aligned}
\tau - \sigma &= (1 - r)(t - \sigma) \\
d\tau &= (t - \sigma) \, dr
\end{aligned}$$

Therefore a change of variables yields:

$$\begin{aligned}
H_\beta * H_s(\xi, t) &= \Gamma_{\xi_0}(\xi, \eta, t, \sigma)(t-\sigma)^{\beta+s+1}\int_0^1 r^s(1-r)^\beta \, dr \\
&\leqslant M\Gamma_{\xi_0}(\xi, \eta, t, \sigma)(t-\sigma)^{\beta+s+1}
\end{aligned}$$

□

The following theorem ends the presentation of the approximation scheme. We will show that a sequence of approximate of the fundamental solution can be constructed by convolution. As the following theorem shows, the step incresing can achieve any order of the remainder.

Theorem 5.3.3. *Assume the hypothesis of the corollary 5.3.1 and let* H_s *be a kernel such that:*

$$|H_s(\xi, t)| \leqslant C_1\big(t^{\frac{s}{2}-1} E(\xi, t)\big). \tag{51}$$

where C_1 *is a positive constant. Then the kernel* Γ_{s+1} *given by*

$$\Gamma_{s+1} = \Gamma_s - \Gamma_{\xi_0} * H_s$$

satisfies

$$L\Gamma_{s+1} = \delta + H_{s+1}$$

with $H_{s+1} \in F_{s+1}$.

Proof. Let H_1 be the kernel defined in corollary 5.3.1. By induction, if there exists H_s satisfying (51) then the following identities hold:

$$\begin{aligned} L\Gamma_{s+1} &= L(\Gamma_s - \Gamma_{\xi_0} * H_s) \\ &= L\Gamma_s - L\Gamma_{\xi_0} * H_s \\ &= L\Gamma_s - (L - L_{\xi_0} + L_{\xi_0})\Gamma_{\xi_0} * H_s \\ &= \delta + H_s - (L - L_{\xi_0})\Gamma_{\xi_0} * H_s - H_s \\ &= \delta - (L - L_{\xi_0})\Gamma_{\xi_0} * H_s \end{aligned}$$

If we set:

$$H_{s+1}(\xi, t) = -(L - L_{\xi_0})\Gamma_{\xi_0} * H_s(\xi, t)$$

then we have:

$$(L - L_{\xi_0})\Gamma_{\xi_0}(\xi, t) = O\big(t^{-\frac{1}{2}} E(\xi, t)\big);$$

$$H_s(\xi, t) = O\big(t^{\frac{s}{2}-1} E(\xi, t)\big).$$

By theorem (5.3.2) we conclude that there exists a positive constant M such that:

$$H_{s+1}(\xi, t) \leqslant M t^{\frac{s-1}{2}} E(\xi, t)$$

Then $H_{s+1} \in F_{s+1}$. □

5.3.1 *A diffeomorphism turning* L_{ξ_0} *into the Heisenberg heat operator* $L_{\mathbb{H}}$

By the applying of parametrices method, we have shown that the fundamental solution Γ of the heat operator L is locally approximated by the fundamental solution Γ_{ξ_0} of the operator L_{ξ_0}. Following [9], we will show that the fundamental solution Γ_{ξ_0} can be written in terms of the fundamental solution $\Gamma_{\mathbb{H}}$ on the Heisenberg group, whose expression is explicitly known. As the following theorem shows, the change of variables (49) provided by exponential coordinates, turns the the approximate operator L_{ξ_0} into the heat operator $L_{\mathbb{H}}$.

Theorem 5.3.4. *The diffeomorphism* $\Phi_{\xi_0}(\xi) = e$ *changes the heat operator* L_{ξ_0} *into the Heisenberg heat operator* $Ł_{\mathbb{H}}$.
Precisely the following identities hold:

(ii) $X_{i,\xi_0}u(\xi) = X_{i,\mathbb{H}}u_{\mathbb{H}}(\Phi_{\xi_0}(\xi))$ *i=1,2*

(iii) $L_{\xi_0}u(\xi) = Ł_{\mathbb{H}}u_{\mathbb{H}}(\xi)$

Corollary 5.3.1. *For every given point* $\xi_0 \in \mathcal{RT}$ *the fundamental solution* Γ_{ξ_0} *of the operator* L_{ξ_0} *is explicitly known and is given by:*

$$\Gamma_{\xi_0}(\xi, t) = \Gamma_{\mathbb{H}}(\Phi_{\xi_0}(\xi), t) \tag{52}$$

where $\Gamma_{\mathbb{H}}$ *is the fundamental solution reported in (50).*

6 DIMENSIONALITY REDUCTION

In this chapter we will go on the problem of dimensionality reduction in the specific context of visual brain. When the brain develops the visual stimulus, it performs grouping and categorization in perceptual units. Natural images are complex data sets which we depict as high-dimensional matrices of pixels, encoding features like color, position, brightness. Such a complex structure is decomposed into simpler figures corresponding to specific features of the natural image.
We will present framework of diffusion map methods ([11],[12],[29]). The main argument is the adaptation of the technique to the subriemannian structure of the visual brain. The convergence of a discrete averaging operator to a diffusion operator on subriemannian curves is proved. Then, this discrete operator will be used to perfom spectral clustering driven by the visual brain. The chapter will end with the numerical simulations.

6.1 THE PROBLEM OF DIMENSIONALITY REDUCTION

The problem of finding a simpler description of complex data sets is often tied to the problem of finding a low-dimensional embedding of the data. In many fields like information theory, machine learning, statistical analysis, computer vision, the understanding of the data go through the discovering of low-dimensional units which carry a great deal of information about the data.
Let us consider an input composed of l points $\xi_1, ..., \xi_l$ where each $\xi_j = (\xi_j^{(1)}, ..., \xi_j^{(n)})$ $(j = 1, ..., l)$ belongs to a n-dimensional manifold $\mathcal{N}$, whose global geometry might be quite complicate.
The reduction of the data deals with two aspects:

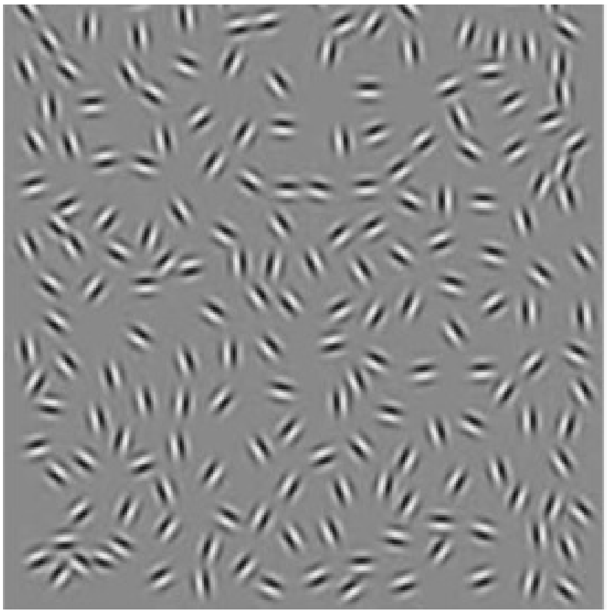


Figure 14: Visual stimulus composed of randomly oriented segments with the emergency of perceptual units

- to represent the data $\Xi = \{\xi_1, ..., \xi_l\} \in \mathcal{N}$ by $Y = \{y^{(1)}, ..., y^{(l)}\} \in \mathcal{Z}$ where $\mathcal{Z}$ is a m-dimensional manifold such that $m \ll n$ (dimensionality reduction);
- to cluster the data into a small number r of groups based on the similarity between points, in order to discover low-dimensional meaningful geometric structures inside the data (clustering and grouping).

The two problems are strickly related with the problem of changing the representation of the data into an easier description which involves a low-dimensional number of free parameters.
Several techniques have been proposed to answer the problem of dimensionality reduction. We will focus on spectral methods, based on the spectral decomposition of a given kernel representing some notion of affinity or similarity between points. The grouping is driven by the affinities between points prescribed by the kernel. Then, the choice of the kernel is crucial: it defines the local geometry of the data.

6.2 GRAPH LAPLACIAN

The clustering problem can be riformulated in terms of graph theory. Let $\Xi = \{\xi_1, ..., \xi_l\} \subset \mathcal{N}$ the given data points. Let $k : \mathcal{N} \times \mathcal{N} \to \mathbb{R}$ be the kernel encoding the affinities between points.
The kernel satisfies the following assumption:

- k is positively preserving: $k(\xi_i, \xi_j) \geqslant 0$;
- k is symmetric: $k(\xi_i, \xi_j) = k(\xi_j, \xi_i)$.

for all $i, j = 1, ..., l$.
The informations carried from the initial datum are collected in a similarity graph.
A graph on the space $\mathcal{N}$ is a pair (V, E) where $V \subset \mathcal{N}$ is a discrete set of points, called nodes, and $E \subset \mathcal{N} \times \mathcal{N}$ is a set of couples of points of V, called edges.
The graph (E, V) is weighted if each edge between two vertices v_i and v_j carries a non-negative entries k_{ij}, called the weight of the edge.
The matrix $K = k_{ij}$ is called the adjacency matrix of the graph.
The degree d_i of a vertex $v_i \in V$ is defined as:

$$d_{ij} = \sum_j k_{ij}$$

The degree matrix D of the graph is defined as:

$$D = diag\{d_{11}, ..., d_{nn}\}$$

We build a similarity graph with the data points as nodes and the kernel entries as the weights of edges connecting points. The assumption of symmetry of the kernel implies that the graph is undirected. For data assembled in a similarity graph, the clustering problem is restated in terms of partitioning the similarity graph. A partition of a graph (V, E) is a collections of nonempty subsets $A_1, ..., A_r$ of V, such that $A_i \cap A_j = \emptyset$ if $i \neq j$ and $A_1 \cup ... \cup A_r = V$. The clustering problem on the graph can be formulated as follows: to find a partition of the graph such that the edges between differents groups have a very low weight, while the edges within a group have high weight. That is,

points in different clusters are dissimilar from each other and points within the same cluster are similar.
The starting point of spectral clustering on the graph is the construction of a matrix representation called graph Laplacian matrix. In literature there exist different variants of graph Laplacians. The main distinction lies in normalized and unnormalized versions. In a large variety of problems one needs to consider suitable normalized versions of graph Laplacians. These matrices show spectral properties similar to the spectral properties of the unnormalized version and are able to approximate anisotropic dynamics. The most pupular choices are the symmetric graph Laplacian and the random walk graph Laplacian. We refer to [10],[41] for a detailed presentation of the topic. Let (V, E) be a weighted graph with adjacency matrix K symmetric and positive definite and degree matrix D.

Definition 6.2.1. *The "random walk" normalized graph Laplacian is defined as:*

$$L_{rw} = D^{-1}L = \mathbb{1} - D^{-1}K$$

If f is a bounded function defined on the data points $\xi_1, ..., \xi_l$, setting $f_k = (f_1, ..., f_l) = f(\xi_1, ..., \xi_l) \in \mathbb{R}^n$, the quadratic form associated to the linear operator L_{rw} acts on f_l is:

$$f_l^T L_{rw} f_l = \frac{1}{2} \sum_{i,j=1}^{l} k(\xi_i, \xi_j)(f_i - f_j)^2 \tag{53}$$

The "random walk" normalized graph Laplacian models a random walk on the graph, i.e. a stochastic Markov process which randomly jumps from node to node. The random walk on the graph is a discrete approximation of a continuous brownian motion. The transition matrix P of the random walk is just the Laplacian L_{rw} renormalized to represent a probability matrix:

$$P = \mathbb{1} - L_{rw} = D^{-1}K. \tag{54}$$

The element $(P)_{ij}$ represents the transition probability of jumping in one step from node ξ_i to node ξ_j. The evolution of the process is driven by the iterates $\{P^s\}_{s \in \mathbb{N}}$ of the transition matrix. The element $(P^s)_{ij}$ represents the transition probability of jumping in s steps from node ξ_i to node ξ_j. In this context the step increasing admits a double interpretation. On one hand, the step is the discrete time step at which the random process jumps from one state to another state. On the other hand, the step can be viewed as a threshold of resolution on the data.

The graph Laplacian is strickly related to the continuous Laplace operator. Setting $k(\xi_i, \xi_j) = \frac{1}{d(\xi_i,\xi_j)^2}$, where d is the euclidean distance defined on the data points, the graph Laplacian (53) turns into:

$$f_l^T L_{rw} f_l = \frac{1}{2} \sum_{i,j=1}^{l} \Big(\frac{f_i - f_j}{d(\xi_i, \xi_j)}\Big)^2$$

Then the operator L_{rw} looks like the discrete analogous of the Laplace operator Δ. Indeed the quadratic form associated to Laplace operator acting on functions defined on a metric space endowed with a inner product $\langle\rangle$ is defined as:

$$\langle f, \Delta f\rangle = \int |\nabla f|^2.$$

The formalization of this intuition has only been provided in special cases. In [29],[3] it has been shown that the graph Laplacian constructed on a similarity graph of randomly sampled data points converges to some continuous Laplace operators. In [29] it is proved that a graph Laplacian properly rescaled converges to the Laplace-Beltrami operator on compact submanifolds of a riemannian manifold. In order to adapt the proof to the subriemannian context of visual brain, in the following section we will show that the graph Laplacian converges to the sublaplacian differential operators computed along horizontal curves.

The spectral clustering we are going to perform is based on the approximation of a given diffusion operator, namely the heat subriemannian operator on the group $\mathcal{RT}$. This operator will be achieved as a continuous limit of a suitable discrete diffusion operator. In order to construct the diffusion operator, we will focus on the generator of the diffusion, in the sense of definition (4.1.6).

The graph Laplacian acts on functions as a discrete averaging operator. If f is a function defined on the data points, we consider the averaging operator defined by L:

$$Lf = \sum_{j=1}^{l} f(\xi_j)k(\xi_i, \xi_j)$$

The choice of the kernel is crucial. In a large variety of problems the euclidean distance between points is not representative of the geometry of the data. For instance, in figure (14) we perceive as grouped elements with comparable alignement, polarity, or frequency. An euclidean neighborhood of a fixed point x of the image will include both points which are similar to x and points dissimilar from x. As a consequence, the choice of the euclidean distance will make the clustering meaningless, while it is necessary to introduce cortical inspired distances encoding the related eingrafted variables. The recognition of perceptual units takes into account the Gestalt principles of good continuation, proximity and co-circularity. The notion of similarity we will use is encoded in the connectivity kernel modeling the association field mechanism. The partition in perceptual units is a integrate process locally driven by the diffusion connectivity kernel. In the neighborhood of a fixed point x, the similarity between points is established by the connectivity kernel starting at x. The clustering problem applied to the visual brain deals with the connection between the peculiar geometry of the perceptual connectivity model and the spectral clustering features techniques. We will look for a dimensionality reduction algorithm skilled in diffusion processes, namely the diffusion maps algorithm.

6.3 DIFFUSION MAPS

Diffusion maps is a machine learning algorithm introduced by R. R. Coifman and S. Lafon (see [29],[12],[11]).
Given the data set defined on a smooth manifold $\mathcal{N}$, the diffusion map algorithm embeds the data into a low-dimensional euclidean space. The embedding is realized computing the first few eigenvectors of the graph Laplacian matrix. The eigenvectors are used as coordinate set. The mapping from the data points space to these eigenfunctions

is denoted as the diffusion map. The affinities between points are measured by a suitable distance called the diffusion distance. In order to introduce the approach, consider a set Ξ of points on a n-dimensional manifold $\mathcal{N}$. The local geometry is inferred by a symmetric definite positive matrix $k \in \mathbb{R}^{n^2}$. Starting from the data Ξ and the matrix k, the method constructs a normalized averaging kernel k^* and a diffusion operator $\mathcal{A}$ such that:

$$\mathcal{A}f(\xi) = \int_Y k^*(\xi,\eta)f(\eta)\, d\sigma(\eta) \tag{55}$$

where $d\sigma$ is a probability measure on Ξ and f is a real-valued bounded function defined on Ξ. The integral operator $\mathcal{A}$, properly renormalized, admits a spectral decomposition. Then the kernel k^* can be written as:

$$k^*(\xi,\eta) = \sum_{i\geqslant 0} \mu^2 \phi_i(\xi)\phi_i(\eta) \tag{56}$$

where $\{\phi\}_{i\geqslant 0}$ is an orthonormal basis of $L^2(\Xi)$. The eigenfunctions $\{\phi_i\}_{i\geqslant 0}$ define a family of diffusion maps $\{\Phi\}_m, m \in \mathcal{N}$, as follows:

$$\Phi_m(\xi) = \begin{pmatrix} \mu_0^{(m)}\phi_0(\xi) \\ \mu_1^{(m)}\phi_1(\xi) \\ \vdots \end{pmatrix} \tag{57}$$

The diffusion maps induce a family of diffusion distances $\{d\}_m$ defined as:

$$d_m^2(\xi,\eta) = k^{*(m)}(\xi,\xi) + k^{*(m)}(\eta,\eta) - 2k^{*(m)}(\xi,\eta) \tag{58}$$

where $k^{*(m)}$ is the kernel of the operator $A^{(m)}$. The kernels k^* and $k^{*(m)}$ can be interpreted as the transition probabilities of a diffusion process define on Ξ. Then the diffusion distance measures the rate of connectivity in accordance with the paths of the diffusion process. As the following identity shows:

$$d^{2(m)}(\xi,\eta) = \| \Phi_m(\xi) - \Phi_m(\eta) \|^2$$

the diffusion map Φ_m realizes an embedding of the data into an Euclidean space. The diffusion distances, underlying the geometry of manifold Ξ, can be easily computed as the euclidean distance on the embedding space. The relevant feature is that the diffusion distance d_m can be accurately approximate by the first few eigenvectors. Up to a permutation, we can assume $\mu_0 \geqslant \mu_1 \geqslant \mu_2 \geqslant \ldots$. The method selects an integer m_0 and a small number of eigenfunctions $\phi_0, \ldots, \phi_{j_0}$ corresponding to eigenvalues $\{\mu_i\}_{i=0,\ldots,j_0}$ which are numerically significant in terms of a fixed resolution threshold. Precisely, it can been shown (see [29]) that there exists an integer m_0 and a constant $C > 0$ such that for all $m \geqslant m_0$ the diffusion map embeds the data into $\mathbb{R}^{j_0}$ and:

$$d_m^2(\xi, \eta) = C \parallel \xi_{j_0} - \eta_{j_0} \parallel \left(1 + O(e^{Cm})\right)$$

where $\xi_{j_0} = (\mu_0^{(m)} \phi_0, \ldots, \mu_{j_0}^{(m)} \phi_{j_0})$.

6.4 CONVERGENCE PROOF

Now we will go on the proof that a suitable graph Laplacian discrete operator converges to the infinitesimal generator of the heat subriemannian diffusion. Similar computations have been pointed out by S. Lafon [29] in the case of diffusion processes on submanifolds of $\mathbb{R}^n$. The proof needs to be adapted to the subriemannian geometry of the space $\mathcal{RT}$.

The proof consists of two steps. Firstly, we will show the convergence of a properly normalized averaging operator to the sublaplacian operator on compact submanifolds. The, the outcome will be used to approximate the heat kernel.

We will start estabilishing an asymptotic expansion as $\varepsilon \to 0$ for the following averaging operator:

$$\mathcal{A}_\varepsilon f(\xi) = \frac{1}{\varepsilon} \int_\Lambda k_\varepsilon(\xi, \eta) f(\eta)\, d\eta \tag{59}$$

where k_ε is a kernel satisfying the hypothesis (6.2). We will consider the special case of diffusion along compact submanifolds such as curves ad surfaces. This restriction simplifies the computation and is sufficient to simulate a large variety of problems of interest. We will expose the proof in two different settings. Firstly, we will consider an asymptotic expansion along 2-dimensional hypersurfaces on $\mathbb{R}^3$ endowed with the euclidean standard metric. Firstly, we will consider an asymptotic expansion along 2-dimensional hypersurfaces on $\mathbb{R}^3$ endowed with the euclidean standard metric. The proof in this setting was originally due to Lafon. Here we present a new, simplified proof. Then, we will consider the expansion along an horizontal curves on the subriemannian space $\mathcal{RT}$ In this setting we use as a kernel of similarity the fundamental solution of the heat kernel, studied from a stochatic and deterministi point of view in the previous chapters.

6.4.1 *Asymptotics for averaging kernels in Euclidean setting*

Let $\mathbb{G} = \mathbb{R}^3$ with the euclidean group structure. Let θ be a real function.The smooth immersion:

$$i : \mathbb{R}^2 \to \mathbb{R}^3$$
$$(x_1, x_2) \to \theta(x_1, x_2)$$

defines a 2-dimensional hypersurface Λ. Let $x = (x_1, x_2)$ be a point on Λ. Up to a change of variables we can suppose that x is the origin. Let $B_{\sqrt{\varepsilon}}(0)$ an euclidean ball centered in x. Since we are interested in modeling the heat diffusion, we make the following assupumption on the kernel:

$$k_\varepsilon(\xi, \eta) = h\left(\frac{|\xi - \eta|^2}{\varepsilon}\right)$$

where h is any differentiable function vanishing at infinity, $\xi = (x_1, x_2, \theta(x_1, x_2)), \eta \in \mathbb{R}^3$.

The operator (59) turns into:

$$\varepsilon \mathcal{A}_\varepsilon f(0) = \int_\Lambda h\left(\frac{|\xi|^2}{\epsilon}\right) f(\xi)\, d\xi = \tag{60}$$

$$= \int_{B_{\sqrt{\varepsilon}}} h\left(\frac{|(x_1, x_2, \theta(x_1, x_2)|^2}{\varepsilon}\right) f(x_1, x_2, \theta(x_1, x_2)) \sqrt{1 + |\nabla\theta|^2}\, dx_1\, dx_2$$

Since $\nabla\theta(0) = 0$ by the choice of coordinates, a Taylor expansion near the origin gives:

$$\partial_{x_i}\theta(x) = \partial_{x_i}\theta(0) + \nabla\partial_{x_i}\theta(0)x + O(x^2) = \sum_{j=1}^{2} x_j \partial_{x_j x_i}{}^2\theta(0) + O(x^2)$$

$$|\nabla\theta(0)|^2 = \sum_{i=1}^{2}(\sum_{j=1}^{2} x_j\theta_{ij}(0))^2 + O(x^2)$$

where $\theta_{ij} = \frac{\partial^2\theta}{\partial x_i \partial x_j}$.

The expansion of the squared term is given by:

$$\sqrt{1 + |\nabla\theta(0)|^2} = 1 + \frac{1}{2}\sum_{i=1}^{2} x_i{}^2(\sum_{j=1}^{2}\theta_{ij}(0))^2 + O(x^2) \quad (61)$$

We Taylor expand the kernel h at $\frac{|x|^2}{\epsilon}$ with respect to increment $\frac{|\theta(x)|^2}{\epsilon}$:

$$h(\frac{|(x,\theta(x))|^2}{\epsilon}) = h(\frac{|x|^2}{\epsilon}) + \frac{1}{\epsilon}|\theta(x)|^2 h'(\frac{|x|^2}{\epsilon}) + O(\epsilon^{\frac{3}{2}})$$
$$= h(\frac{|x|^2}{\epsilon}) + \frac{1}{2\epsilon}(\sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j \theta_{ij})^2 h'(\frac{|x|^2}{\epsilon}) + O(\epsilon^{\frac{3}{2}})$$

We Taylor expand the function f near the origin:

$$f(x,\theta(x)) = f(0) + \sum_{i=1}^{2} x_i f_i(0) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j f_{ij}^2(0) + O(\epsilon^{\frac{3}{2}})$$

Therefore the integral operator turns into:

$$\epsilon\mathcal{A}_\epsilon f(x) = \int_{B_{\sqrt{\epsilon}}} \left[h(\frac{|x|^2}{\epsilon}) + \frac{1}{2\epsilon}(\sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j \theta_{ij})^2 h'(\frac{|x|^2}{\epsilon}) + O(\epsilon^{\frac{3}{2}})\right]$$
$$\left[f(0) + \sum_{i=1}^{2} x_i f_i(0) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2} x_i x_j f_{ij}^2(0) + O(\epsilon^{\frac{3}{2}})\right]$$
$$\left[1 + \frac{1}{2}\sum_{i=1}^{2} x_i{}^2(\sum_{j=1}^{2}\theta_{ij}(0))^2 + O(x^2)\right] dx$$

The simmetry of the kernel gives:

$$\int_{B_{\sqrt{\varepsilon}}} x_i h(\frac{|x|^2}{\varepsilon}) = 0$$
$$\int_{B_{\sqrt{\varepsilon}}} x_i x_j h(\frac{|x|^2}{\varepsilon}) = 0 \quad i \neq j$$

Consequently the integral operator turns into:

$$\varepsilon\mathcal{A}_\varepsilon f(0) = f(0)\int_{B_{\sqrt{\varepsilon}}} h(\frac{|x|^2}{\varepsilon})\,dx + \int_{B_{\sqrt{\varepsilon}}} h(\frac{|x|^2}{\varepsilon})f(0)\frac{1}{2}\sum_{i=1}^{2} {x_i}^2(\sum_{j=1}^{2}\theta_{ij}(0))^2\,dx +$$
$$\int_{B_{\sqrt{\epsilon}}} h(\frac{|x|^2}{\epsilon})\frac{1}{2}\sum_{i=1}^{2}{x_i}^2 f_{ii}^2(0) + \int_{B_{\sqrt{\epsilon}}}\frac{1}{2\epsilon}(\sum_{i=1}^{2}\sum_{j=1}^{2}x_i x_j\theta_{ij})^2\,h'(\frac{|x|^2}{\epsilon})f(0)\,dx + O(\epsilon^{\frac{3}{2}})$$

$$\varepsilon\mathcal{A}_\varepsilon f(0) = f(0)\int_{B_{\sqrt{\varepsilon}}} h(\frac{|x|^2}{\varepsilon})\,dx + f(0)\frac{1}{2}\sum_{i=1}^{2}(\sum_{j=1}^{2}\theta_{ij}(0))^2\int_{B_{\sqrt{\varepsilon}}}{x_i}^2 h(\frac{|x|^2}{\varepsilon})\,dx +$$
$$-\frac{1}{2}\Delta f(0)\int_{B_{\sqrt{\varepsilon}}}{x_1}^2 h(\frac{|x|^2}{\varepsilon})\,dx + \frac{1}{2\varepsilon}f(0)\sum_{i=1}^{2}\sum_{j=1}^{2}\theta_{ij}^4\int_{B_{\sqrt{\varepsilon}}}{x_i}^2{x_j}^2 h'(\frac{|x|^2}{\varepsilon})\,dx + O(\varepsilon^{\frac{3}{2}})$$

Now we set:

$$m_0 = \int_{B_{\sqrt{\varepsilon}}} h(|x|^2)\,dx$$
$$m_2 = \int_{B_{\sqrt{\varepsilon}}} {x_1}^2 h(|x|^2)\,dx$$
$$m_{ij} = \int_{B_{\sqrt{\varepsilon}}} {x_i}^2{x_j}^2 h'(|x|^2)\,dx$$

The integral expansion simplifies:

$$\mathcal{A}_\varepsilon f(0) = m_0 f(0) + \varepsilon\frac{1}{2}f(0)\sum_{i=1}^{2}(\sum_{j=1}^{2}\theta_{ij}(0))^2 m_2 - \frac{\varepsilon}{2}\Delta f(0)m_2$$
$$+\frac{\varepsilon}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\theta_{ij}^2 m_{ij} + O(\varepsilon^{\frac{3}{2}})$$

Integration by parts yields:

$$m_{ij} = -\frac{1}{2}m_2 \quad i = j$$
$$m_{ij} = -\frac{3}{2}m_2 \quad i \neq j$$

The integral operator turns into:

$$\mathcal{A}_\varepsilon f(0) = m_0 f(0) - \frac{\varepsilon}{2}\Delta f(0) m_2 + \frac{\varepsilon}{2} f(0)\Big[\sum_{i=1}^{2} \theta_{ii}^4 m_{ii} + \sum_{i=1}^{2}\sum_{j\neq i} \theta_{ij}^2 m_{ij}\Big] + O(\varepsilon^{\frac{3}{2}}) =$$

$$= m_0 f(0) - \frac{\varepsilon}{2}\Delta f(0) m_2 + \frac{\varepsilon}{2} f(0)\Bigg[\sum_{i=1}^{2} \theta_{ii}^4 \Big[-\frac{1}{2}m_2\Big] + \sum_{i=1}^{2}\sum_{j\neq i} \theta_{ij}^2\Big[-\frac{3}{2}m_2\Big]\Bigg] + O($$

$$= m_0 f(0) - \frac{\varepsilon}{2}\Delta f(0) m_2 + \frac{\varepsilon}{2} f(0) m_2\Bigg[-\frac{1}{2}\sum_{i=1}^{2} \theta_{ii}^4 2 - \frac{3}{2}\sum_{i=1}^{2}\sum_{j\neq i} \theta_{ij}^2\Bigg] + O(\varepsilon^{\frac{3}{2}})$$

Now we set:

$$E(x) = -\frac{1}{2}\sum_{i=1}^{2} \theta_{ii}^2(x) - \frac{3}{2}\sum_{i=1}^{2}\sum_{j\neq i} \theta_{ij}^2(x)$$

The averaging integral turns into:

$$\mathcal{A}_\varepsilon f(0) = m_0 f(0) + \frac{\varepsilon}{2} m_2\Bigg[-\Delta f(0) + E(0) f(0)\Bigg] + O(\varepsilon^{\frac{3}{2}})$$

Eventually we effort the following statement.

Proposition 6.4.1. *The averaging integral operator (59) admits the following asymptotic expansion:*

$$\mathcal{A}_\epsilon f(z) = m_0 f(z) + \frac{\epsilon}{2} m_2\Bigg[-\Delta f(z) + E(z) f(z)\Bigg] + O(\varepsilon^{\frac{3}{2}}) \qquad (62)$$

for each $z \in \mathbb{R}^3, \varepsilon > 0$.

6.4.2 *Asymptotics for the weighted graph Laplacian*

In order to model diffusion processes, we need to consider a properly normalized version of the averaging operator of the previous section. Precisely, we will look for a normalized graph Laplacian.
We set:

$$v_\varepsilon{}^2(z) = \int_\Lambda k_\varepsilon(z,\xi) p(\xi)\, d\xi.$$

The normalized averaging operator is defined as:

$$\mathcal{A}_\varepsilon f(z) = \frac{1}{v_\epsilon{}^2(z)}\int_\Lambda k_\varepsilon(z,\xi) f(\xi) p(\xi)\, d\xi \qquad (63)$$

From (62) we have:

$$\int_{\Lambda} k_\varepsilon(z,\xi)f(\xi)p(\xi)\,d\xi = \epsilon\Big[m_0 f(z)p(z) + \frac{\varepsilon}{2}m_2\Big[-\Delta f(z)p(z) + E(z)f(z)p(z)\Big]\Big] + O(\varepsilon^{\frac{3}{2}})$$

Now plugging-in $f = 1$ yields:

$$\int_{\Lambda} k_\varepsilon(z,\xi)p(\xi)\,d\xi = \varepsilon\Big[m_0 p(z) + \frac{\varepsilon}{2}m_2\Big[-\Delta p(z) + E(z)p(z)\Big]\Big] + O(\varepsilon^{\frac{3}{2}})$$

Taking the ratio yields:

$$\mathcal{A}_\varepsilon f(z) = \frac{\epsilon\Big[m_0 f(z)p(z) + \frac{\varepsilon}{2}m_2\Big[-\Delta f(z)p(z) + E(z)f(z)p(z)\Big]\Big] + O(\varepsilon^{\frac{3}{2}})}{\varepsilon\Big[m_0 p(z) + \frac{\varepsilon}{2}m_2\Big[-\Delta p(z) + E(z)p(z)\Big]\Big] + O(\varepsilon^{\frac{3}{2}})}$$

A Taylor expansion yields:

$$\Big[m_0 p(z) + \frac{\varepsilon}{2}m_2\Big[-\Delta p(z) + E(z)p(z)\Big] + O(\varepsilon^{\frac{3}{2}})\Big]^{-1} =$$

$$\Big(m_0 p(z)\Big)^{-1}\Big(1 + \frac{\varepsilon m_2}{2m_0 p(z)}\Big[-\Delta p(z) + E(z)p(z)\Big] + O(\varepsilon^{\frac{3}{2}})\Big)^{-1} =$$

$$\Big(m_0 p(z)\Big)^{-1}\Big(1 - \frac{\varepsilon m_2}{2m_0 p(z)}\Big[-\Delta p(z) + E(z)p(z)\Big] + O(\varepsilon^{\frac{3}{2}})\Big) =$$

The integral operator turns into:

$$\begin{aligned}
\mathcal{A}_\varepsilon f(z) = {} & \frac{1}{m_0 p(z)}\Big(m_0 f(z)p(z) + \frac{\varepsilon}{2}m_2\Big(-\Delta f(z)p(z) + E(z)f(z)p(z)\Big)\Big) \\
& \Big(1 - \frac{\varepsilon m_2}{2m_0 p(z)}(-\Delta p(z) + E(z)p(z))\Big) + O(\varepsilon^{\frac{3}{2}}) =
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_\varepsilon f(z) = & \Big(f(z) + \frac{\varepsilon}{2}\frac{m_2}{m_0 p(z)}(-\Delta(f(z)p(z)) + E(z)f(z)p(z))\Big) \\
& \Big(1 - \frac{\varepsilon m_2}{2m_0 p(z)}(-\Delta p(z) + E(z)p(z))\Big) + O(\varepsilon^{\frac{3}{2}}) = \\
= & f(z) + \frac{\varepsilon}{2}\frac{m_2}{m_0} - \frac{\varepsilon}{2}\frac{m_2}{m_0 p(z)}\Delta(f(z)p(z)) \\
& + \frac{\varepsilon}{2}\frac{m_2}{m_0 p(z)}f(z)\Delta p(z) - \frac{\varepsilon}{2}\frac{m_2}{m_0}f(z)E(z) + O(\varepsilon^{\frac{3}{2}}) = \\
= & f(z) + \frac{\varepsilon}{2}\frac{m_2}{m_0}\Big(\frac{\Delta p(z)}{p(z)}f(z) - \frac{\Delta(f(z)p(z))}{p(z)}\Big) + O(\varepsilon^{\frac{3}{2}})
\end{aligned}$$

Now plugging-in $p = 1$ yields the following statement.

Proposition 6.4.2. *The normalized averaging operator (63) admits the following asymptotic expansion:*

$$\mathcal{A}_\varepsilon f(z) = f(z) - \frac{\varepsilon}{2}\frac{m_2}{m_0}\Delta f(z) + O(\varepsilon^{\frac{3}{2}}) \tag{64}$$

for each $\varepsilon > 0$.

6.4.3 *Asymptotics for averaging kernels on* $\mathcal{RT}$

Now we will go on the main content of this chapter. Arguing as in [29], we will show an asymptotic expansion for the graph Laplacian operator along integral curves defined on the motion group $\mathcal{RT}$.
Let ξ_0 be a fixed point on $\mathcal{RT}$. Let Γ be the fundamental solution of the heat operator (39) and Γ_{ξ_0} its local approximate (see chapter 5). In accordance with the the definition of horizontal curves we have introduced, we will look for a diffusion along an integral curve $\gamma : [-\sqrt{\varepsilon}, \sqrt{\varepsilon}] \to \mathcal{RT}$ satisfying:

$$\dot{\gamma}(t) = X_1(\gamma(t)) + k(t)X_2(\gamma(t))$$

with $\gamma(t) = (x(t), y(t), \theta(t))$ and $\dot{\theta}(t) = k(t)$. We will start estabilishing an asymptotic expansion as $\varepsilon \to 0$ for the following averaging operator:

$$\mathcal{A}_\varepsilon f(\xi) = \frac{1}{\varepsilon} \int_\gamma \Gamma_{\xi_0}(\xi, \eta, \varepsilon) f(\eta) d\sigma(\eta) \tag{65}$$

By theorem (5.3.4), the operator (65) turns into:

$$\begin{aligned} \varepsilon \mathcal{A}_\varepsilon f(\xi_0) &= \int_\gamma \Gamma_{\mathbb{H}}\big(\Phi_{\xi_0}(\xi_0), \Phi_{\xi_0}(\eta)\big) f(\eta) d\sigma(\eta) \\ &= \int_\gamma \Gamma_{\mathbb{H}}\big(0, \Phi_{\xi_0}(\eta), \varepsilon\big) f(\eta) d\sigma(\eta) \\ &= \int_\gamma \Gamma_{\mathbb{H}}\big(0, \Phi_{\xi_0}(x(t), y(t), \theta(t)), \varepsilon\big) f(x(t), y(t), \theta(t)) \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{\theta}^2}\, dt \end{aligned}$$

where Φ is the diffeomorphism (49).
Taking into account the explicit expression of Γ_{ξ_0} (50) the kernel Γ_{ξ_0} can be written as:

$$\Gamma_{\mathbb{H}}\big(0, \Phi_{\xi_0}(x(t), y(t), \theta(t)), \varepsilon\big) = h\left(\frac{e_1^2(t) + e_2^2(t)}{\varepsilon}, \frac{|e_3(t)|}{\varepsilon}\right)$$

where h is a differentiable function vanishing at infinity and e_1, e_2, e_3 are the coordinates defined by the change of variables (48). Therefore the operator (65) turns into:

$$\varepsilon \mathcal{A}_\varepsilon f(\xi_0) = \int_\gamma h\left(\frac{e_1^2(t) + e_2^2(t)}{\varepsilon}, \frac{|e_3(t)|}{\varepsilon}\right) f(x(t), y(t), \theta(t)) \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{\theta}(t)^2}\, dt$$

Plugging-in the change of variables (48), a Taylor expansion around $t=0$ gives:

$$x(t)-x(0)=t\dot{x}(0)+\frac{t^2}{2}\ddot{x}(0)+O(t^3)=t\cos\theta_0-\frac{t^2}{2}\sin\theta_0+O(t^3)$$
$$y(t)-y(0)=t\dot{y}(0)+\frac{t^2}{2}\ddot{y}(0)+O(t^3)=t\sin\theta_0+\frac{t^2}{2}\cos\theta_0+O(t^3)$$

Then:

$$\begin{aligned}e_1(t)&=\sin\theta_0(y-y_0)+\cos\theta_0(x-x_0)=\\&=t\sin^2\theta_0+\frac{t^2}{2}\cos\theta_0\sin\theta_0+t\cos^2\theta_0-\frac{t^2}{2}\cos\theta_0\sin\theta_0=\\&=t+O(t^3)\end{aligned}$$

$$e_2(t)=\theta(t)-\theta_0=\frac{t^2}{2}\ddot{\theta}(0)+O(t^3)$$

$$\begin{aligned}e_3(t)&=\cos\theta_0(y-y_0)-\sin\theta_0(x-x_0)-\frac{1}{2}e_1e_2=\\&=t\cos\theta_0\sin\theta_0+\frac{t^2}{2}\cos^2\theta_0-t\cos\theta_0\sin\theta_0+\\&\frac{t^2}{2}\sin^2\theta_0-\frac{1}{2}\big(t+O(t^3)\big)\big(\frac{t^2}{2}\ddot{\theta}(0)+O(t^3)\big)=\\&=\frac{t^2}{2}\big(1-\dot{\theta}(0)\big)+O(t^3)\end{aligned}$$

We Taylor expand the kernel h at $\frac{e_1(t)^2+e_2(t)^2}{\varepsilon}$ with respect to increment $\frac{|e_3(t)|}{\varepsilon}$:

$$\begin{aligned}h\left(\frac{e_1^2+e_2^2}{\varepsilon},\frac{|e_3|}{\varepsilon}\right)&=h\left(\frac{e_1^2+e_2^2}{\varepsilon}\right)+\frac{|e_3|}{\varepsilon}h'\left(\frac{e_1^2+e_2^2}{\varepsilon}\right)+O(\varepsilon^{\frac{3}{2}})=\\&=h\left(\frac{t^2\big(1+\dot{\theta}^2(0)\big)}{\varepsilon},\frac{t^2|1-\dot{\theta}(0)|}{2\varepsilon}\right)\\&+\frac{t^2|1-\dot{\theta}(0)|}{2\varepsilon}h'\left(\frac{t^2\big(1+\dot{\theta}^2(0)\big)}{\varepsilon},\frac{t^2|1-\dot{\theta}(0)|}{2\varepsilon}\right)+O(\varepsilon^{\frac{3}{2}})\end{aligned}$$

A Taylor expansion of the function f gives:

$$\begin{aligned}f\big(\gamma(t)\big)=\big(f\circ\gamma\big)(0)&=\big(f\circ\gamma\big)(0)+t\big(f\circ\gamma\big)'(0)+\frac{t^2}{2}\big(f\circ\gamma\big)''(0)+O(t^3)=\\&=f(\xi_0)+t\big(X_1+\dot{\theta}(0)X_2\big)f(\xi_0)+\frac{t^2}{2}\big(X_1+\dot{\theta}(0)X_2\big)^2f(\xi_0)+O(t^3)\end{aligned}$$

We Taylor expand the derivative $\dot\theta(t)$:

$$\dot\theta(t) = \dot\theta(0) + t\ddot\theta(0) + \frac{t^2}{2}\dddot\theta(0) + O(t^3)$$
$$\dot\theta^2(t) = \dot\theta^2(0) + t^2\ddot\theta^2(0) + 2t\dot\theta(0)\ddot\theta(0) + 2t^2\dot\theta(0)\dddot\theta(0) + O(t^3)$$

A Taylor expansion of the squared term gives:

$$\sqrt{\dot x(t)^2 + \dot y(t)^2 + \dot\theta(t)^2} = \sqrt{1 + \dot\theta(t)^2} =$$
$$= 1 + \frac{1}{2}\dot\theta^2(0) + t\dot\theta(0)\ddot\theta(0) + \frac{t^2}{2}\ddot\theta^2(0) + t^2\dot\theta(0)\dddot\theta(0) + O(t^3)$$

The averaging kernel (65):

$$\varepsilon\mathcal{A}_\varepsilon = \int_\gamma h\left(\frac{e_1^2(t) + e_2^2(t)}{\varepsilon}, \frac{|e_3(t)|}{\varepsilon}\right) f(x(t), y(t), \theta(t))\sqrt{\dot x(t)^2 + \dot y(t)^2 + \dot\theta(t)^2}\, dt$$

turns into:

$$\varepsilon\int_{B_{\sqrt\varepsilon}} \left[h\left(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\right) + \frac{t^2}{2\varepsilon}(\dot\theta(0))h'\left(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\right)\right]$$
$$\left[f(\xi_0) + t\big(X_1 + \dot\theta(0)X_2\big)f(\xi_0) + \frac{t^2}{2}\big(X_1 + \dot\theta(0)X_2\big)^2 f(\xi_0) + O(t^3)\right]$$
$$\left[1 + \frac{1}{2}\dot\theta^2(0) + t\dot\theta(0)\ddot\theta(0) + \frac{t^2}{2}\ddot\theta^2(0) + t^2\dot\theta(0)\dddot\theta(0) + O(t^3)\right] dt$$
$$=$$

The simmetry of h gives:

$$\int_{B_{\sqrt\varepsilon}} th\left(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\right) dt = 0$$

The operator (65) simplifies:

$$\begin{aligned}
\varepsilon\mathcal{A}_\varepsilon f(\xi_0) &= \int_{B_{\sqrt{\varepsilon}}} h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\Big)\, dt \\
&+ \frac{1}{2}\big(X_1+\dot\theta(0)X_2\big)^2 f(\xi_0)\int_{B_{\sqrt{\varepsilon}}} t^2 h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\Big)\, dt \\
&+ \frac{1}{2} f(\xi_0)\dot\theta^2 \int_{B_{\sqrt{\varepsilon}}} h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\Big)\, dt \\
&+ f(\xi_0)\big(\frac{1}{2}\ddot\theta^2(0)+\dot\theta(0)\dddot\theta(0)\big)\int_{B_{\sqrt{\varepsilon}}} t^2 h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\Big)\, dt \\
&+ \frac{1}{2\varepsilon} f(\xi_0)\big(1-\dot\theta(0)\big)\big(1+\frac{\dot\theta^2(0)}{2}\big)\int_{B_{\sqrt{\varepsilon}}} t^2 h'\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2|1-\dot\theta(0)|}{2\varepsilon}\Big)\, dt
\end{aligned}$$

Now we set:

$$\begin{aligned}
m_0 &:= \int_{B_1} h\big(t^2(1+\dot\theta^2(0)), \frac{t^2}{2}|1-\dot\theta(0)|\big)\, dt \\
m_2 &:= \int_{B_1} t^2 h\big(t^2(1+\dot\theta^2(0)), \frac{t^2}{2}|1-\dot\theta(0)|\big)\, dt \\
m_3 &:= \int_{B_1} t^2 h'\big(t^2(1+\dot\theta^2(0)), \frac{t^2}{2}|1-\dot\theta(0)|\big)\, dt
\end{aligned}$$

A change of variables yields:

$$\begin{aligned}
&\int_{B_{\sqrt{\varepsilon}}} h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2(1-\dot\theta(0))}{\varepsilon}\Big)\, dt = \varepsilon m_0 \\
&\int_{B_{\sqrt{\varepsilon}}} t^2 h\Big(\frac{t^2(1+\dot\theta^2(0))}{\varepsilon}, \frac{t^2(1-\dot\theta(0))}{\varepsilon}\Big)\, dt = \varepsilon^2 m_2
\end{aligned}$$

The expansion turns into:

$$\begin{aligned}
\varepsilon\mathcal{A}_\varepsilon f(\xi_0) &= \varepsilon m_0 f(\xi_0) + \frac{1}{2}\varepsilon^2 m_2\big(X_1+\dot\theta(0)X_2\big)^2 f(\xi_0) + \frac{1}{2}\dot\theta^2(0)\varepsilon m_0 f(\xi_0) \\
&+ f(\xi_0)\big(\frac{1}{2}\ddot\theta^2(0)+\dot\theta(0)\dddot\theta(0)\big)\varepsilon^2 m_2 + \frac{1}{2\varepsilon} f(\xi_0)\big(1-\dot\theta(0)\big)\big(1+\frac{\dot\theta^2(0)}{2}\big)\varepsilon^2 m_3
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_\varepsilon f(\xi_0) &= m_0 f(\xi_0) + \frac{1}{2}\varepsilon m_2\big(X_1+\dot\theta(0)X_2\big)^2 f(\xi_0) + \frac{1}{2}\dot\theta^2(0) m_0 f(\xi_0) \\
&+ f(\xi_0)\big(\frac{1}{2}\ddot\theta^2(0)+\dot\theta(0)\dddot\theta(0)\big)\varepsilon m_2 + \frac{1}{2} f(\xi_0)\big(1-\dot\theta(0)\big)\big(1+\frac{\dot\theta^2(0)}{2}\big) m_3
\end{aligned}$$

Plugging-in $\dot\theta(0)=0$ the integral operator turns into:

$$\mathcal{A}_\varepsilon f(\xi_0) = m_0 f(\xi_0) + \frac{1}{2}\varepsilon m_2 \Delta_\gamma f(\xi_0) + \frac{1}{2}\varepsilon m_2 f(\xi_0)\ddot\theta^2(0) + \frac{1}{2} f(\xi_0) m_3$$

Now we set:

$$m_1 = m_0 + \frac{1}{2} m_3 \tag{66}$$

$$E(\xi_0) = \ddot{\theta}^2(0) \tag{67}$$

The computations yield the following statement.

Proposition 6.4.3. *For each $\varepsilon > 0$, the averaging integral operator (65) admits the following asymptotic expansion:*

$$A_\varepsilon f(\xi_0) = m_1 f(\xi_0) + \frac{\varepsilon}{2} m_2 \big(-\Delta_\gamma f(z) + E(\xi_0) f(\xi_0) \big) + O(\varepsilon^{\frac{3}{2}})$$

where Δ_γ is the Heisenberg sublaplacian computed along the integral curve γ, m_0,E are the terms defined in (66),(67).

6.4.4 *Asymptotics for the weighted graph Laplacian on $\mathcal{RT}$*

As in the euclidean case, we will consider a normalized version of the averaging kernel.
Setting:

$$\nu_\epsilon{}^2(\xi_0) = \int_\gamma k_\varepsilon(\xi_0, \xi)\, d\xi,$$

the normalized averaging operator is defined as:

$$\mathcal{A}_\epsilon f(\xi_0) = \frac{1}{\nu_\epsilon{}^2(\xi_0)} \int_\gamma k_\varepsilon(\xi_0, \xi) f(\xi)\, d\xi \tag{68}$$

Arguing as in section (6.4.4), we effort the following statement.

Proposition 6.4.4. *The normalized averaging operator (68) admits the following asympotics expansion:*

$$\mathcal{A}_\varepsilon f(\xi_0) = f(z) - \frac{\varepsilon}{2} \frac{m_2}{m_1} \Delta_\gamma f(\xi_0) + O(\varepsilon^{\frac{3}{2}}) \tag{69}$$

for each $\varepsilon > 0$.

6.5 HEAT KERNEL APPROXIMATION

The outcomes of previous section are collected in the following theorem.

Theorem 6.5.1. *Let A_ε be the normalized graph Laplacian integral operator (68).*

Let

$$\Delta_\varepsilon = \frac{I - \mathcal{A}_\varepsilon}{\varepsilon}$$

be the approximateLaplace operator defined by $\mathcal{A}_\varepsilon$. Then for all $f \in C_0^\infty(\mathcal{RT})$:

$$\lim_{\varepsilon \to 0} \Delta_\varepsilon f = \Delta_0 f$$

where $\Delta_0 = 2\frac{m_2}{m_1}\Delta_\gamma$.

Proof. The statement is a direct consequence of proposition (6.4.4). □

Following [29], we will go on the approximation of the heat operator along an horizontal integral curve, starting from the explicit expression of the associated infinitesimal generator.

Theorem 6.5.2. *Let $T > 0$ be a fixed time. For any $t \in [0, T]$:*

$$\lim_{\varepsilon \to 0} \mathcal{A}_\varepsilon^{\frac{t}{\varepsilon}} = e^{-t\Delta_0}$$

where $e^{-t\Delta_0}$ is the heat operator on $\mathcal{RT}$.

Proof. By proposition (6.4.4):

$$\mathcal{A}_\varepsilon = I - \varepsilon\Delta_0 + \varepsilon^{\frac{3}{2}} R_\varepsilon$$

where R_ε is a bounded operator.
Let $\frac{1}{\varepsilon} = 2^l$, with $l \in \mathbb{N}$. If l is sufficiently large, for each m such that $1 \leqslant m \leqslant l$ the following expansion hold:

$$\mathcal{A}_\varepsilon^{2^m} = (I - \varepsilon\Delta_0)^{2^m} + \varepsilon^{\frac{3}{2}} R_\varepsilon^{(m)}$$

with $\| R_\varepsilon^{(m)} \| \leqslant 2^{m+1} \| R_\varepsilon \|$.
Indeed by induction:

$$\mathcal{A}_\varepsilon^2 = \left(I - \varepsilon\Delta_0 + \varepsilon^{\frac{3}{2}} R_\varepsilon\right)^2 \tag{70}$$

$$= \left(I - \varepsilon\Delta_0\right)^2 + \varepsilon^{\frac{3}{2}}\left((I - \varepsilon\Delta_0)R_\varepsilon + R_\varepsilon(I - \varepsilon\Delta_0) + \varepsilon^{\frac{3}{2}} R_\varepsilon^2\right) \tag{71}$$

$$= \left(I - \varepsilon\Delta_0\right)^2 + \varepsilon^{\frac{3}{2}} R_\varepsilon^{(1)} \tag{72}$$

where

$$R_\varepsilon^{(1)} = \left((I-\varepsilon\Delta_0)R_\varepsilon + R_\varepsilon(I-\varepsilon\Delta_0) + \varepsilon^{\frac{3}{2}}R_\varepsilon^2\right)$$

satisfies the estimate:

$$\| R_\varepsilon^{(1)} \| \leqslant 2 \| R_\varepsilon \| + \varepsilon^{\frac{3}{2}} \| R_\varepsilon \|^2$$

since $\| I-\varepsilon\Delta_0 \| \leqslant 1$ if ε is sufficiently small.

Now suppose that:

$$\mathcal{A}_\varepsilon^{2^m} = (I-\varepsilon\Delta_0)^{2^m} + \varepsilon^{\frac{3}{2}} R_\varepsilon^{(m)}.$$

Then:

$$\begin{aligned}\mathcal{A}_\varepsilon^{2^{m+1}} &= (I-\varepsilon\Delta_0)^{2^{m+1}} \\ &+ \varepsilon^{\frac{3}{2}}\left((I-\varepsilon\Delta_0)^{2^m}R_\varepsilon^{(m)} + R_\varepsilon^{(m)}(I-\varepsilon\Delta_0)^{2^m} + \varepsilon^{\frac{3}{2}}R_\varepsilon^{(m)2}\right) \\ &= (I-\varepsilon\Delta_0)^{2^{m+1}} + \varepsilon^{\frac{3}{2}}R_\varepsilon^{(m+1)}\end{aligned}$$

with $R_\varepsilon^{(m+1)}$ satisfying for ε sufficiently small:

$$\| R_\varepsilon^{(m+1)} \| \leqslant 2 \| R_\varepsilon^{(m)} \| + \varepsilon^{\frac{3}{2}} \| R_\varepsilon^{(m)} \|^2 .$$

Let $r_m = 2^{-m} \| R_\varepsilon^{(m)} \|$. Then the following inequality holds:

$$r_{m+1} \leqslant r_m + 2^{m-1-\frac{3}{2}l} r_m^2$$

Now suppose that $u_m \leqslant 2r_0$ for all $m \leqslant m_0 \leqslant l$. Summing the previous inequality yields:

$$\begin{aligned}r_{m_0} &\leqslant r_0 + 2^{-1-\frac{3}{2}l} 4 r_0^2 \sum_{j=0}^{m_0-1} 2^j \\ &\leqslant r_0 + 2^{1-\frac{1}{2}l} 2^{m_0-l} r_0^2 \\ &\leqslant r_0 + 2^{1-\frac{1}{2}l} r_0^2 \\ &\leqslant 2r_0\end{aligned}$$

if l is sufficiently large. Then $r_m \leqslant 2r_0$ for all $m \leqslant i$. It follows:

$$\| R_\varepsilon^{(m+1)} \| \leqslant 2^{m+1} \| R_\varepsilon \|$$

with R_ε bounded as $\varepsilon \to 0$. Taking $m = l$ we conclude:

$$\begin{aligned}&\lim_{\varepsilon\to 0} \mathcal{A}_\varepsilon^{\frac{1}{\varepsilon}} = \\ &= \lim_{\varepsilon\to 0} (I-\varepsilon\Delta_0)^{\frac{1}{\varepsilon}} + 2\varepsilon^{\frac{1}{2}} R_\varepsilon = e^{-\Delta_0}\end{aligned}$$

□

Since we are interested in spectral grouping on $\mathcal{RT}$, we need to ensure that the eigenfunctions of the approximate heat operator converge to the eigenfunctions of the heat operator e^{Δ_0}.

Theorem 6.5.3 (Heat eigenfunctions approximation)**.** *The averaging operator $\mathcal{A}_\varepsilon$ is compact, then by spectral theorem admits the the following decomposition:*

$$\mathcal{A}_\varepsilon^{\frac{t}{\varepsilon}} = \sum_{j\geqslant 0} \mu_{\varepsilon,j}^{\frac{t}{\varepsilon}} P_{\varepsilon,j}$$

where $P_{\varepsilon,j}$ is the orthogonal projector on the eigenspace corresponding to $\mu_{\varepsilon,j}$. Moreover, if

$$e^{-t\Delta_0} = \sum_{j\geqslant 0} e^{-t\nu_j^2} P_j$$

is the spectral decomposition of the heat operator, then:

$$\lim_{\varepsilon\to 0} \mu_{\varepsilon,j}^{\frac{t}{\varepsilon}} = e^{-t\nu_j^2} \tag{73}$$

$$\lim_{\varepsilon\to 0} P_{\varepsilon,j} = P_j \tag{74}$$

Proof. See [29]. □

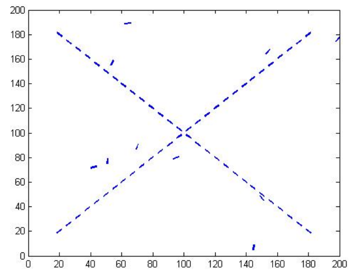
6.6 NUMERICAL EXPERIMENT

Numerical simulations will test the grouping performances of the heat kernels of connectivity. The discretization of the kernel k is performed via the stochastic method described in chapter (4). The operator $\mathcal{A}_\varepsilon$ is discretized on the set of points (x, y, θ) of the stimulus:

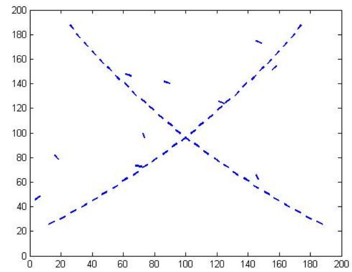
$$\mathcal{A}_\varepsilon(x_j, y_j, \theta_j) = \sum_i k((x_j, y_j, \theta_j), (x_i, y_i, \theta_i) f(x_i, y_i, \theta_i)$$

Since the operator acts on a finite dimensional space, we will idetify $\mathcal{A}_\varepsilon$ with the matrix K obtained restricting k on the data set. We will call this matrix K the affinity matrix. The matrix K will be normalized in accordance with formula (54), in order to keep its probabilistic meaning. The first eigenvalues of K cluster the stimulus in $r = 3$ perceptual groups.

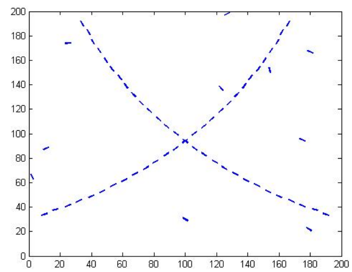
Consider a stimulus composed of oriented segments labeled with both spatial coordinates (x, y) and orientation θ. The input data set is of size $[n_x, n_y, n_\theta] = [200, 200, 64]$. The data set is perceptually organized in three groups, as shown in fig. (15): two perceptual units and a third group composed of noise segments.



(a)



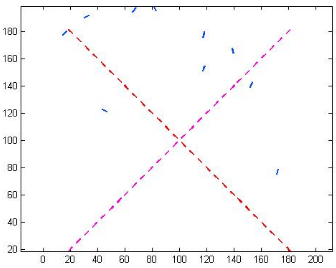
(b)



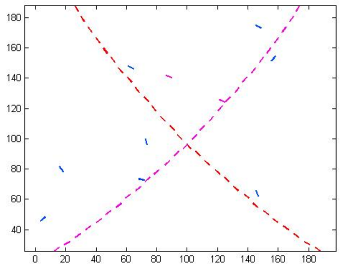
(c)

Figure 15: Original stimulus with $j = 10$ noise segments with increasing values of curvature

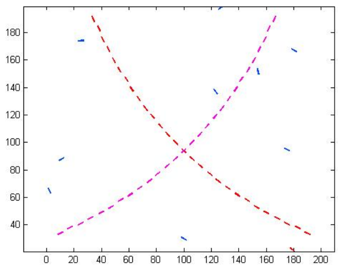
The output is depicted in fig. (16).



(a)



(b)



(c)

Figure 16: Spectral clustering performed with diffusion maps

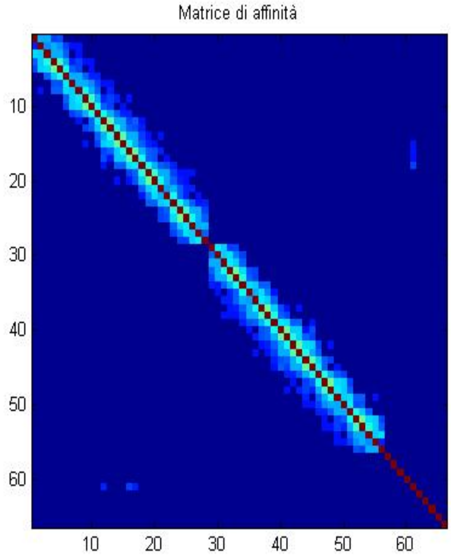


Figure 17: Visualization of the block diagonal affinity matrix K

BIBLIOGRAPHY

[1] D. Barbieri, G. Citti, G. Cocci, A. Sarti: A cortical-inspired geometry for contour perception and motion integration, 2013.

[2] D. Barbieri, G. Cocci, A. Sarti, Spatiotemporal receptive fields of cells in V1 are optimally shaped for stimulus velocity estimation. J. Opt. Soc. Am. A 29(1), 130-138, 2012.

[3] M. Belkin, P. Niyogi, Towards a Theoretical Foundation for Laplacian-Based Manifold Methods, The University of Chicago, Department of Computer Science.

[4] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni, Stratifed Lie groups and potential theory for their sub-Laplacians. Springer Monographs in Mathematics. Springer, Berlin, 2007.

[5] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni, Uniform Gaussian estimates of the fundamental solutions for heat operators on Carnot groups, Adv. Differential Equations 7, 1153-1192, 2002.

[6] W.H. Boskin, Y. Zhang, B. Scholfield, D. Fitzpatrick, Orientation selectivity and the arrangement of horizontal connections in the tree shrew striate cortex, 1998. J. Neurosci. 17(6), 2112-27, 1997.

[7] L. Capogna, D. Danielli ,S. Pauls, J. Tyson, An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem, Progress in Mathematics, vol. 259, Birkhauser Verlag, Basel, 2007.

[8] G. Citti, A. Sarti: A cortical based model of perceptual completion in the roto-translation space. J. Math. Imag. Vis. 24(3), 307-326,2006.

[9] G. Citti, Regularity of solutions of a quasilinear equation related to the Levi operator. Annali della Scuola Normale Superi-

ore di Pisa, Scienze Fisiche e Matematiche Serie IV. Vol. XXIII. Fasc 3, 1996.

[10] F. Chung, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics, No. 92, 1997.

[11] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner, S. Zucker, Geometric Diffusion as a tool for harmonic analysis and structure definition of data, part I: Diffusion Maps, Proceedings of the National Academy of Sciences, 102(21):7426-31, 2005.

[12] R.R. Coifman, S. Lafon, Diffusion maps, Applied and Computational Harmonic Analysis: Special issue on Diffusion Maps and Wavelets, Vol 21, July 2006, pp 5-30.

[13] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Am. A 2:1160-1169, 1985.

[14] G.C. DeAngelis, I. Ohzawa, R.D. Freeman, Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General Characteristics and Postnatal Development. J. Neurophysiol. 69(4), 1091-1117, 1993.

[15] R. Duits, M. van Almsick. The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2D euclidean motion group. Quart. Appl. Math., 66, 27-67, 2008.

[16] L. Evans, Convergence of an Algorithm for mean curvature motion, Indiana Univ. Math J.,vol 42, n.2, 553-557,1993.

[17] D.J. Field, A. Hayes, R.F. Hess, Contour integration by the human visual system: evidence for a local Association Field, Vision Research, 33, 173-193, 1993.

[18] G.B. Folland, Sub-elliptic estimates and function spaces on nilpotent groups, Arkiv for Math. 13 161-207, 1975.

[19] W.C. Hoffman, The visual cortex is a contact bundle, Applied Mathematics and Computation, 32, 137-167, 1989.

[20] W.C. Hoffman, The Lie algebra of visual perception, Journal of Mathematical Psychology 3, 65-98, 1966.

[21] L. Hormander. Hypoelliptic second order differential equations. Acta Math., 119, 147-171, 1967.

[22] D.H. Hubel, Eye, Brain, and Vision, Volume 22, Scientific American Library.

[23] D.H. Hubel and T.N. Wiesel. Ferrier Lecture: Functional Architecture of Macaque Monkey Visual Cortex. Proceedings of the Royal Society of London Series B 198 (1130) 1-59, 1997.

[24] D.S. Jerison, A. Sanchez-Calle, Estimates for the heat kernel for a sum of squares of vector felds, Indiana Univ. Math. J.35 (4),835â854, 1986.

[25] J.P. Jones, L.A. Palmer, An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex, Journal of Neurophysiology, Vol. 58, No. 6, 1987.

[26] G. Kanizsa, Organization in vision, Hardcover, 1979.

[27] G. Kanizsa, Grammatica del vedere, Il mulino, Bologna, 1980.

[28] J.J. Koenderink, A.J. van Doorn, Representation of local geometry in the visual system. Biol. Cybernet. 55,367-375, 1987.

[29] S. Lafon, Diffusion maps and geometric harmonics, Ph.D. dissertation, Yale University, May 2004.

[30] D. Mumford, Elastica and computer vision, In: Bajaj, C.L. (eds.) Algebraic Geometry and its Applications, pp. 491-506,Springer-Verlag,1994.

[31] A. Nagel, E.M. Stein, S. Wainger, Balls and metrics defined by vector fields I: Basic properties. Acta Math. 155, 103-147, 1985.

[32] B. Øksendal, Stochastic Differential Equations. An Introduction with Applications. Berlin-Heidelberg-New York-Tokyo, Springer-Verlag 1985.

[33] J. Petitot, The neurogeometry of pinwheels as a sub-Riemannian contact structure, Journal of Physiology, Paris,97, 265-309, 2003.

[34] J. Petitot,Y. Tondut, Vers une Neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux, Mathematiques, Informatique et Sciences Humaines, 145, EHESS, CAMS, Paris,5-101, 1999.

[35] L.P. Rothschild, E.M.Stein, Hypoelliptic differential operators and nilpotent groups. Acta Math., 137, 347320, 1976.

[36] L.V. Sabinin, Smooth quasigroups and loops. Kluwer, 1999.

[37] A. Sanchez-Calle, Fundamental solutions and geometry of the sum of squares of vector fields, Inventiones mathematicae 78, 143-160, 1984.

[38] A. Sarti,G. Citti,J. Petitot, The symplectic structure of the primary visual cortex. Biol. Cybern. 98, 33-48, 2008.

[39] P. Sorba, The Galilei group and its connected subgroups. J. Math. Phys 17(6), 941-953, 1976.

[40] L.R. Williams, D.W. Jacobs, Stochastic completion fields, ICCV Proceedings, 1995.

[41] U. Von Luxburg, A tutorial on spectral clustering. Technical Report No. TR-149, 2006.

[42] S.W. Zucker, Differential geometry from the Frenet point of view:boundary detection, stereo, texture and color. In: Paragios, N., Chen, Y., Faugeras, O. (eds.) Handbook of Mathematical Models in Computer Vision, pp. 357-373, Springer, US, 2006.