ELSEVIER

# The neurogeometry of pinwheels as a sub-Riemannian contact structure

## Jean Petitot [*]

*CREA, École Polytechnique, 1 Rue Descartes, 75005 Paris, France*

## Abstract

We present a geometrical model of the functional architecture of the primary visual cortex (V1) and, more precisely, of its pinwheel structure. The problem is to understand from within how the internal "immanent" geometry of the visual cortex can produce the "transcendent" geometry of the external space. We use first the concept of blowing up to model V1 as a discrete approximation of a continuous fibration $\pi : R \times P \to P$ with base space the space of the retina $R$ and fiber the projective line $P$ of the orientations of the plane. The core of the paper consists first in showing that the horizontal cortico-cortical connections of V1 implement what the geometers call the contact structure of the fibration $\pi$, and secondly in introducing an integrability condition and the integral curves associated with it. The paper develops then three applications: (i) to Field's, Hayes', and Hess' psychophysical concept of association field, (ii) to a variational model of curved modal illusory contours (in the spirit of previous models due to Ullman, Horn, and Mumford), (iii) to Ermentrout's, Cowan's, Bressloff's, Golubitsky's models of visual hallucinations.
© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Contact structure; Sub-Riemannian geometry; Pinwheels; Functional architecture; Blowing up; Fibration; Horizontal connections; Jet space; Connection; Lie group; Association field; Illusory contour; Variational model

## 1. Introduction

Fine grained experimental results concerning the structure of the first retinotopic areas of the visual cortex are now sufficiently accurate to justify a modeling of their functional architecture. The purpose of this article is to present such a model for the most elementary level of early vision, that of the pinwheel structure of the primary visual cortex (V1). We want to show that this functional architecture implements *geometrical* algorithms which explain how the neural calculus is able to integrate contours.

As neurons are local processors, we will have to solve two main problems:

1. how are neurally implemented local geometrical data such as "points" $a$, pairs $(a, p)$ of a position $a$ and of an orientation $p$, etc.;

2. how are these local data integrated in global geometrical structures such as lines, figures, shapes, etc.

To be clear from the outset and avoid any misunderstanding, let us make some preliminary remarks.

1. This research is based on what we call *neurogeometry*. The idea is that the geometrical structures which are constitutive of perceptively construed objects are neurally implemented in geometrically well behaved functional architectures. As was often pointed out by the logician Giuseppe Longo (see e.g. [44]) the brain cannot be a universal Turing machine because neural computations are highly dependent upon their specific neural coding. The classical functionalist hypothesis according to which cognitive "softwares" are essentially independent from the neural "hardware" they are implemented in seems to be quite wrong, as if, in neurocomputing, cognition was in the coding and *computation was the hardware*.
2. The internal geometry of low level vision is very different from the classical Euclidean geometry of the external objective world. It is a representation (a coding) of the proximal information transduced by the

photoreceptors. This formatting of the external optical signal is a kind of filtering (already at the retinal level the ganglion cells process a wavelet analysis of the signal), but, due to the functional architecture, *these local measures of the signal by arrays of parallel point processors are equivalent to a global geometrization.*

3. We apply the analytic ("cartesian") methodology which is classical in mathematical physics: simplify and idealize first a basic relevant phenomenon and try to completely reconstruct it mathematically, and then, but only in a second step, to reconstruct other aspects of it and other phenomena depending upon it. This methodology is rather rare in neurosciences where the complexity of biological phenomena are often argued against the very idea of a mathematical modeling.

4. We will restrict our analysis to the area V1. Of course the model would have to be considerably complexified to take into account the other areas (V2, V4, etc.) and their top-down feedbacks [79]. But we will see that the neurogeometry of primitive geometrical concepts such as that of a line is already rather complex at the V1 level. In fact, we will adopt Mumford–Lee's "high-resolution buffer hypothesis" according to which V1 is not a mere bottom-up "early-module", but participates via a lot of top-down connections to any higher levels of processing that requires high resolution, its functional architecture being therefore essential to the entire visual system (see [42,51]). Many specialists consider (see e.g. Robert Hess' contribution to this volume) that V1 is more important than previously thought.

5. The mathematical elements we will use belong to domains of modern (post-Riemannian) differential geometry since Elie Cartan's and Hermann Weyl's pioneering works: differential forms, integrals of integrable systems of differential forms, contact structures, fibrations, jet spaces, connections, Lie groups, sub-Riemannian metrics, etc. Some readers will perhaps be surprised by this recourse to relatively modern and sophisticated mathematical structures, even if some of them were already present in the seminal works of William Hoffman and Jan Koenderink. But we must emphasize again that one of the main problem of early vision is to integrate local measures in global Gestalts. Now, this problem has also been a fundamental one in pure geometry and in physics for a long time, and it is precisely for solving it that geometers worked out such tools. It is why we think relevant to introduce them into that new context.

6. Our perspective being foundational, it runs also into philosophy of mathematics. Its horizon is to understand in the framework of a neural materialism the origin of the spatial intuition formalized by geometry and, moreover, to bridge the gap between neural

activity and ideality of space. This aspect of the classical immanent/trancendent opposition is quite hard to clarify: as far as the visual system can access only its immanent structure we must understand "from within" the "transcendent" geometry of the external space. Since the deep works of Riemann, Helmholtz, Poincaré, Hilbert, Cartan, Husserl, and Weyl, this problem of the cognitive origin of space is one of the main enigma of epistemology. We think that, at the beginning of this new century, neurogeometry can become an alternative to the foundational axiomatic approach. Neural materialism will be perhaps the future of logical idealism.

7. It is quite a long time (the early seventies) since I began to work on the applications of differential geometry and of singularity theory (in the sense of Hassler Whitney, René Thom, Vladimir Arnold, James Damon, etc.) to perception. During the nineties I have been particularly interested in the works of David Mumford, Jean-Michel Morel, Steven Zucker, and Jan Koenderink who all strongly emphasized the interest of such geometrical tools for the study of natural vision and image analysis. It is the importance of many new exciting experimental results (such as those of Gregory De Angelis or William Bosking) which persuaded me to develop for them new neurogeometrical models.

After having presented briefly a model for the receptive profiles of the simple cells of V1 and for the way they act by convolution on the optical signal, we work out a first elementary model of the functional architecture of V1 as a *fibration* (in the geometric sense) $\pi : R \times P \to P$ with base space the space of the retina $R$ and with fiber the projective line $P$ of the orientations of the plane.

Then we present a geometrical model of the pinwheel structure of V1 using the geometrical concept of *blowing up* and we show that this model can be considered as a discrete approximation of the fibration model.

The core of the paper consists in showing that the horizontal cortico-cortical connections of V1 implement what the geometers call the *contact structure* of the fibration $\pi$. We can therefore introduce the *Frobenius integrability condition* canonically associated with this contact structure, condition which defines *integral curves* and explains how to globalize into global curves local data $(a, p)$ of a retinal position $a$ and an orientation $p$. The metric computed as the minimal length of integral curves joining two points $(a, p)$ and $(b, q)$ of V1 is an example of what is called a *sub-Riemannian Carnot-Carathéodory* metric.

As a first application, we show that the integrability condition—which gives a rigorous geometrical status to the old Gestalt principle of "good continuation"—corresponds exactly to the psychophysical concept of

*association field* introduced by David Field, Anthony Hayes, and Robert Hess.

We show next that the same integrability condition allows to work out a *variational* model of *curved modal illusory contours* in Kanizsa's sense. This improves previous models due to Ullman and Horn and is akin to Mumford's elastica model for amodal occluded contours.

Finally, we show that the striking Ermentrout's, Cowan's, Bressloff's, Golubitsky's model for visual hallucinations consists in encoding the contact structure of the fibration into the synaptic weights of a continuous neural net and analyzing the stability properties and the bifurcations of the associated partial differential equation.

## 2. Receptive fields and wavelet analysis

### 2.1. Receptive profiles

The receptive field (RF) of a visual neuron is classically defined as the domain of the retina to which it is connected through the neural connections of the retino-geniculo-cortical pathways (projecting from the retina to the cortex through the thalamic way) and whose stimulation elicitates a spike response. As was stressed by Yves Frégnac and others such as Lamme and Maffei (see e.g. [41,46]), this concept of "minimal discharge field" (MDF) has to be refined to take into account the sub-threshold activity of the neurons [20].

In the following we will restrict our models to the RF in the narrow sense of MDF.

Classically, an RF is decomposed into ON (positive contrast) and OFF (negative contrast) zones according to the type of response to light and dark luminance Dirac stimulations. There exists therefore a receptive profile (RP) of the visual neuron, which is simply its transfer function as a filter. It is a function $\varphi(x, y)$ (where $x, y$ are retinal coordinates) $\varphi : D \to \mathbb{R}$ which is defined on the RF $D$ and measures the response $(+ = \text{ON}, - = \text{OFF})$ of the neuron to stimulations at the point $(x, y)$.

Sophisticated techniques enable the recording of the *level curves* of the RPs (see e.g. De Angelis [12]). A light and dark spot or bar is switched ON and OFF at different positions of the RF and the mean response is measured. One uses for instance random sequences of flashes (of 50 ms) distributed over a lattice of $20 \times 20$ positions, with 100 ms–1 s for each response after each flash, and takes the mean value on 10 flashes at each position (white noise analysis). The correlation of the inputs (flashes) with the outputs (spikes) yields the transfer function of the neuron.

It is a classical result of neurophysiology, already strongly emphasized by David Marr in the late 1970s, that the RF of the retinal ganglion cells are like Lapla-

cians of Gaussians $\Delta G$. It is the same thing for the cells of the lateral geniculate nucleus (Figs. 1–3).

More generally, there exist in the visual system RPs which have the form of *partial derivatives of Gaussians DG* up to order (at least) 3, and to order (at least) 4 if one takes into account the temporal evolutions of the RPs due to fast synaptic plasticity and adaptability to stimuli. The simple orientation cells of V1 have an RP in $\partial^3 G / \partial x^3$ (Figs. 4–6).

Many neurophysiologists interpret these RPs as Gabor patches (trigonometric functions modulated by a Gaussian). Qualitatively it is effectively the same thing, but there are many evidences in favor of the *DG* interpretation (see Johnston's, McOwan's, Benton's paper, this volume). We prefer the *DG* interpretation for the following reason.

### 2.2. Visual neurons as convolution operators

How do such visual neurons act on the signal? Let $I(x, y)$ be the optical signal defined on the retina $R$. Let $\varphi(x - x_0, y - y_0)$ be the RP of a neuron $N$ whose RF is defined on a domain $D$ of the retina centered on $(x_0, y_0)$. $N$ acts on the signal $I$ as a filter. It computes the mean value of $I$ on $D$ weighted by the weight $\varphi$

$$I_\varphi(x_0, y_0) = \int_D I(x', y') \varphi(x' - x_0, y' - y_0) \, dx' \, dy'. \quad (1)$$

This is a *measure* at $(x_0, y_0)$ of the signal $I$ by the neuron $N$. A field of such neurons covering the whole retina $R$ acts therefore by *convolution* on the signal

$$
\begin{aligned}
I_\varphi(x, y) &= \int_D I(x', y') \varphi(x' - x, y' - y) \, dx' \, dy' \\
&= (I * \varphi)(x, y).
\end{aligned} \quad (2)
$$

In that sense, the processing can be thought of as a neural implementation of a *wavelet analysis* of the signal.

As was emphasized by Florack [17], it is relevant to treat the signal $I$ (which is a geometrically very bad behaved noisy function) as a *distribution*, that is as a continuous linear functional $\langle I | \varphi \rangle$ defined on a space of test functions $\varphi$ (regular $C^\infty$ localized functions with compact support or at least rapidly decreasing). One can then treat the RPs $\varphi(x, y)$—which are highly regular and localized functions—as classes of test functions *wired* in the visual system. The measure of the signal $I$ by a neuron of RP $\varphi$ provides then its representation $\langle I | \varphi \rangle$ in the distributional sense.

Now, it is well known that the Dirac distribution $\delta$ is the basic operator of the distributional differential calculus. Indeed, for every distribution $T$ we have

$$\delta * T = T, \quad \delta' * T = T', \quad \delta^{(m)} * T = T^{(m)} \quad (3)$$

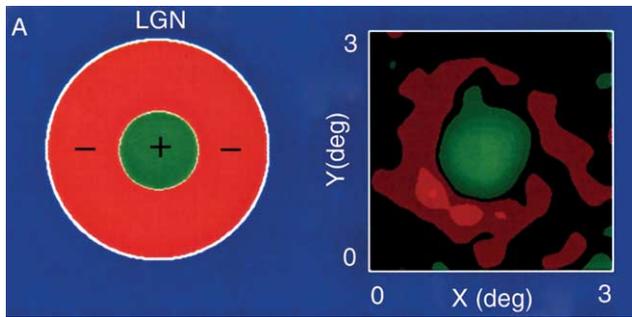(where $T^{(m)}$ is the $m$th derivative of $T$) and, more generally

Fig. 1. The receptive profile of a (ON-center) LGN cell. (Left) Structure with the + (ON) and − (OFF) domains and (Right) record of the level curves. It would be the same for the ganglion cells of the retina (from [12]).
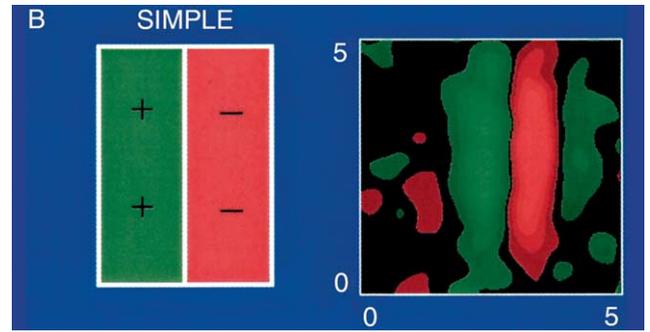


Fig. 4. The receptive profile of a simple cell of V1. (Left) Structure with the + (ON) and − (OFF) domains and (Right) record of the level curves (from [12]).
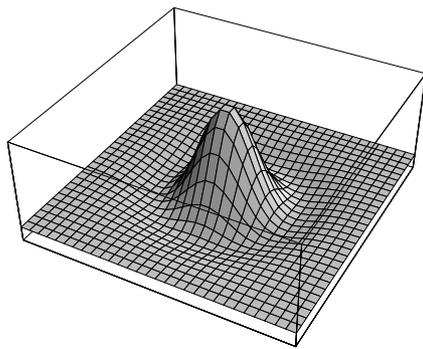


Fig. 2. The Laplacian model $\Delta G$ of the ganglion and LGN (ON-center) cells with the + center and the − periphery.
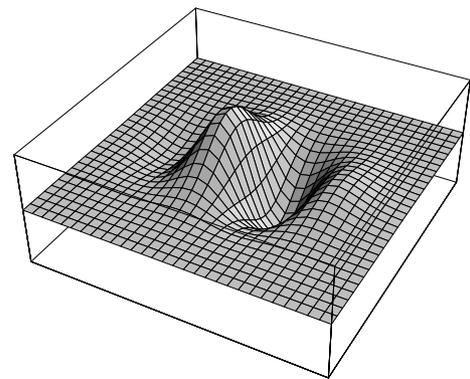


Fig. 5. The $\partial^3 G/\partial x^3$ model of the simple orientation cells of V1.
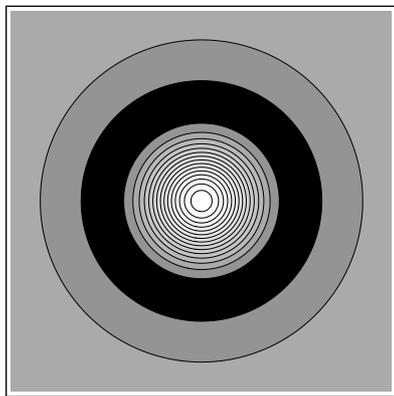


Fig. 3. The level curves of $\Delta G$ fit with the empirical results of Fig. 1.



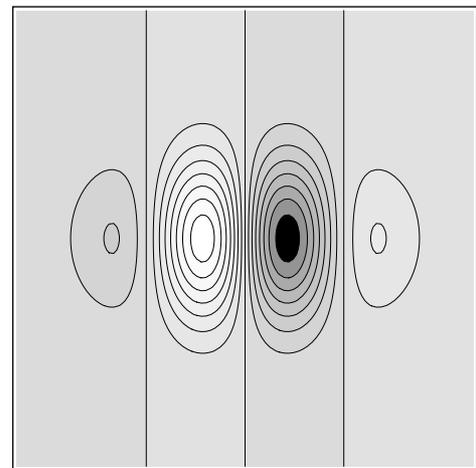Fig. 6. The level curves of $\partial^3 G/\partial x^3$ fit with the empirical results of Fig. 4.

$$D\delta * T = DT \tag{4}$$

for every differential operator $D$ (with constant coefficients).

To substitute a Gaussian $G_\sigma$

$$G_\sigma = G(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \tag{5}$$

for $\delta$ is therefore equivalent to the choice of a certain scale (a certain resolution) defined by the width $\sigma$ of $G_\sigma$. The family $G_\sigma$ expresses what becomes a point in a multi-scale wavelet filtering, and this introduces the key concept of *scale-space* (see e.g. [18]). The convolution

product of Gaussians satisfies $G_\sigma * G_\tau = G_{\sqrt{\sigma^2+\tau^2}}$, and the composition law of scales $\dotplus$ is therefore given by $\sigma \dotplus \tau = \sqrt{\sigma^2 + \tau^2}$; in other words it is the parameter $\sigma^2$ which satisfies an additive law. The Gaussian being the heat kernel, the multi scale wavelet filtering consists in taking the signal $I$ as an initial condition for the *heat equation*

$$\left(\frac{\partial}{\partial s} - \Delta\right)I = 0 \quad \text{(with } 2s = \sigma^2\text{)}. \tag{6}$$

This diffusion equation links the ideal geometry (infinite resolution) with its physical counterpart (finite resolution). It expresses the operational constraint of transforming the physical optical signal into a geometrical observable. It substitutes local multi-scale concrete measures for ideal infinitesimal ones.

If $D$ is a differential operator of order $p$ one has by definition

$$\langle DI | \varphi \rangle = (-1)^p \langle I | D\varphi \rangle. \tag{7}$$

This fundamental equality shows that the main function of RPs which are partial derivatives of Gaussians is to derive the signal $I$ in the distributional sense. More precisely an RP of the form $DG$ applies the differential operator $D$ at a certain scale. In that sense, visual neurons implement *multi-scale differential geometry*.

From the classical formula $I * DG = D(I * G)$, the convolution of the signal $I$ by a $DG$ RF amounts to apply $D$ to the *smoothing* $I * G$ of the signal $I$ at the scale defined by $G$. At the scale 0 (infinite resolution) we find of course the classical derivative $DI$ because $I * D\delta = D(I * \delta) = DI$. For instance, the ganglion cells whose RFs are Laplacians of Gaussian $\Delta G$ compute the Laplacian of the signal $\Delta(I * G) = I * \Delta G$ at the scale defined by $G$. It is well known that these cells detect spatial contrasts (see e.g. [9,49]). This is due to the fact that they perform a typical wavelet analysis whose function is precisely to extract the encoded qualitative discontinuities (see e.g. Mallat [48]). A well known criterion is e.g. Marr's zero crossing: a discontinuity corresponds to a crossing of 0 flanked by two very acute peaks, respectively positive and negative. But this is nothing else than the multi-scale version of the well known dipole structure of the second derivative $\delta''$.

### 2.3. Feature detectors

With these types of methods one can easily construct multi-scale detectors of geometric features more complex than simple discontinuities. For instance a *corner* detector must detect a strong curvature on a boundary. Boundaries are well detected by the gradient norm $|\nabla I|$. The curvature of an isophote (a level line of $I$) being given by the divergence of the normalized gradient

$$\kappa = \text{div}\left(\frac{\nabla I}{|\nabla I|}\right) = \frac{1}{|\nabla I|}\left(\Delta I - \frac{H(\nabla I, \nabla I)}{|\nabla I|^2}\right) \tag{8}$$

(where $H$ is the Hessian of $I$), a good corner detector is the invariant $\kappa |\nabla I|^3$. When one implements it, one gets RPs which respond only to places where the signal $I$ smoothed by $G$ presents a curve of discontinuity with a discontinuity of the tangent (see e.g. Florack [17] and Hamy [26]).

## 3. The functional architecture of V1

But, of course, filtering is radically insufficient, and a functional architecture is absolutely necessary. As multi-scale feature detectors remain local, the fundamental problem of a geometric formatting is the *continuation from the local level to the global one* and it requires a whole structuration of the detectors arrays. It is the global *coherence* of the architecture which generates geometry.

### 3.1. Hypercolumns

We will focus on area V1. Neurophysiological studies have discriminated three types of structures in V1, respectively layered, retinotopic and (hyper)columnar.

(i) The layered structure (about 1, 8 mm thick) is constituted of 6 "horizontal" layers, the most important for our purpose being the layer 4 and more precisely the sublayer 4$C$ where most of the fibers from the lateral geniculate body project.

(ii) The retinotopy means that there exist mappings from the retina to the cortical layers which preserve retinal topography. If we note by $R$ the retina and by $M$ the cortical layer, the retinotopy is then described by a map $\rho : R \rightarrow M$ which is an isomorphism for a certain level of geometric structure (Fig. 7). A model well fitting the empirical data is given by the logarithmic conformal map $\text{Log}[(z + 0.333)/(z + 6.66)]$ [1]. [1] (Fig. 8).

(iii) The columnar and hypercolumnar structure organizes the cells of V1 in columns corresponding to parameters such as orientation, ocular dominance, and color (blobs). It is the simple orientation cells [74] which are of interest here. Due to their RP in third derivative of Gaussian $\frac{\partial^3 G}{\partial x^3}$, they detect preferred orientations, that is, at a certain scale, pairs $(a, p)$ of a spatial (retinal) position $a$ and of a local orientation $p$ at $a$. The hypercolumnar organization means essentially that to each position $a$ of the

---

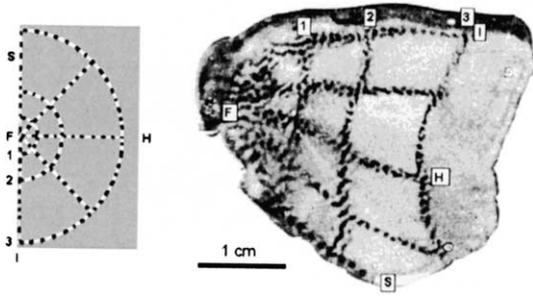[1] When we do not need to distinguish between $R$ and $M$ we will put $R = M$ and $\rho = Id$.

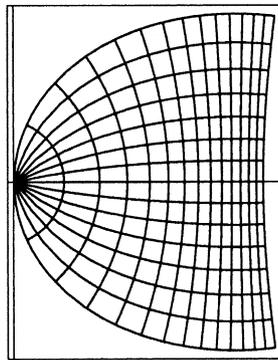Fig. 7. The retinotopic map between the retina and the cortical layer 4C.



Fig. 8. The conformal map $\text{Log}[(z + 0.333)/(z + 6.66)]$ as a model of the retinotopic map between the retina and the cortical layer 4C.

retina $R$ is associated a full exemplar $P_a$ of the space of orientations $p$ at $a$. [2]

Cells grouped in one column are strongly redundant: their orientational preferences are similar and their receptive fields strongly overlap. This redundancy allows two things:

(i) the variation of the phase (see [13], where De Angelis and his co-workers show that in a column "spatial phase is the single parameter that accounts for most of the difference between receptive fields of nearby neurons");
(ii) a "population coding" providing a greater accuracy in the representation of the orientation and spatial position of the stimulus, the resolution of the network being greater than that of the individual detectors (see e.g. Snippe and Koenderink [63]).

_____

[2] To every point in the visual field is associated the set of orientations of a bar, the set of values of ocular dominance, the set of wavelengths of a stimulus, and possibly other parameters (such as movement direction, curvature, binocular disparity, etc.). For more technical details about these parameters, and especially curvature detection, see e.g. Zucker et al. [81].

### 3.2. The first geometrical model of V1 as a fibration

Given the model of the visual space as a 2-manifold, how to take into account the "secundary" variables: orientation, ocular dominance, color, direction of movement (and possibly others)? The set of all possible values of such variables are represented in each cortical hypercolumn and a hypercolumn covers a small chart of the visual field. We have to model the concept of "engrafting" variables on the basic retinal variables $(x, y)$, in the sense of Hubel:

> What the cortex does is to map not just two but many variables on its two-dimensional surface. It does so by selecting as the basic parameters the two variables that specify the visual field coordinates (distance out and up or down from the fovea), and on this map it engrafts other variables, such as orientation and eye preference, by finer subdivisions. [32, p. 131]

A hypercolumn can be modeled as the cartesian product of its chart with the space of secundary "engrafted" variables. The overlap of receptive fields is interpreted as *gluing* operators and the structural projection is implemented by the vertical connections

(i) from the retina to the hypercolumns, and
(ii) within the hypercolumn itself.

The most simple and idealized model of V1 uses the fundamental geometrical concept of a *fibration* (or fiber bundle).

### 3.2.1. Fiber bundles and "engrafted" variables

The key notion of a fiber bundle was worked out by mathematicians for rather deep reasons: how to deal with processes which associate to every point of a manifold $M$ an entity of a certain type $F$ (a scalar, a vector, a covector, a tensor, an exterior form, a direction, a phase, a quantum number, a color, a texture, etc.), depending on that point? A first (evident) solution is to use maps $\varphi : M \to F$, the result being called *fields* on $M$. But in many cases, it is necessary to consider that to every point of $M$ is associated the complete set of values in $F$. For example, in a very concrete technical domain, at every pixel of a computer screen, the whole set of grey levels (1 byte) or of RGB colors (3 bytes) is represented.

Intuitively speaking, a fiber bundle is constituted by a base space $M$ (a differentiable manifold) and by a copy of the same manifold $F$, called the fiber, "above" each of its points. Globally, the total space $E$ of the fiber bundle (the fibers "glued" together) is not necessarily a trivial cartesian product $M \times F$. It results from the gluing of several cartesian products $U_i \times F$ defined on local (open)

domains $U_i$ of $M$. This *local triviality* has until now interested geometers only to deal with situations where the base $M$ was not a globally trivial space $\mathbb{R}^n$ but a manifold that might not be simply connected and might present homotopy and homology. In our case, the fibrations are globally trivial, but *their local structure is imposed by neurophysiology* (receptive fields).

By definition, a fiber bundle (or fibration) is a quadruple $(E, M, F, \pi)$ such that

1. $E$, $M$ and $F$ are differentiable manifolds, called respectively the total space, the base space and the fiber of the bundle.
2. $\pi : E \to M$ is a surjective differentiable map, called the structural projection of the bundle.
3. All inverse images $E_x = \pi^{-1}(x)$ $(x \in M)$ are isomorphic to $F$. $E_x \simeq F$ is called the fiber above the point $x$ (Fig. 9).
4. For every point $x \in M$, there exists a neighborhood $U$ of $x$ such that $\pi^{-1}(U)$ is diffeomorphic to the product $U \times F$ endowed with the canonical projection $\pi : U \times F \to U$, $(x, p) \mapsto x$ (local triviality, see Fig. 10).

A *section* of a fibration is a differentiable map which lifts up the projection $\pi$, associating to every point $x$ of the base manifold $M$ an element of the fiber $E_x$. If $s : M \to E$ is a section, we have therefore: $\pi \circ s = Id_M$ (see Fig. 11). Sections can of course be defined only locally over open subsets $U \subset M$.

In the case of a globally trivial bundle $\pi : E = M \times F \to M$, a section over $U$ is nothing else than a map $s : U \to F$.



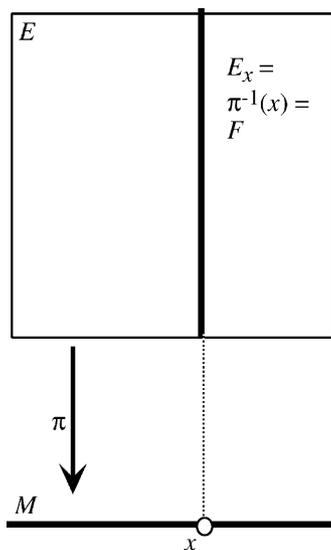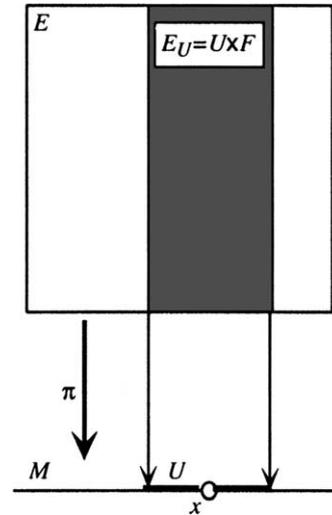Fig. 10. The local triviality of a fibration. For every point $x$ of $M$ there exists a neighborhood $U$ of $x$ whose inverse image $\pi^{-1}(U) = E_U$ is the direct product $U \times F$ with $\pi$ the projection on the first factor.
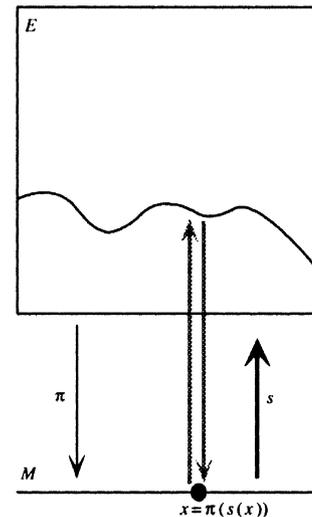


Fig. 11. A section of a fibration defined on an open set $U$ of $M$ associates to every point $x$ of $U$ a value $s(x)$ in the fiber $E_x$ above $x$.

### 3.2.2. V1 as a fiber bundle

If we idealize mathematically the functional architecture of the retino-geniculo-cortical way, the retinotopic and hypercolumnar structure of V1 will be naturally modeled by the fibration $\pi : V \to R$ which associates to every point $a$ of the retina $R$ a copy $P_a$ of the space $P$ of the directions of the plane. $P_a$ is isomorphic to the unit circle $\mathbb{S}^1$ if we take into account the sense of the directions, and to the projective line $\mathbb{P}^1$ (the quotient of $\mathbb{S}^1$ by the equivalence relation identifying pairs of diametrally opposed points) if we do not take it into account. The total space $V$ of these copies of



Fig. 9. The general schema of a fibration with base space $M$, fiber $F$ and total space $E$. Above every point $x$ of $M$ the fiber $\pi^{-1}(x) = E_x$ is isomorphic to $F$.
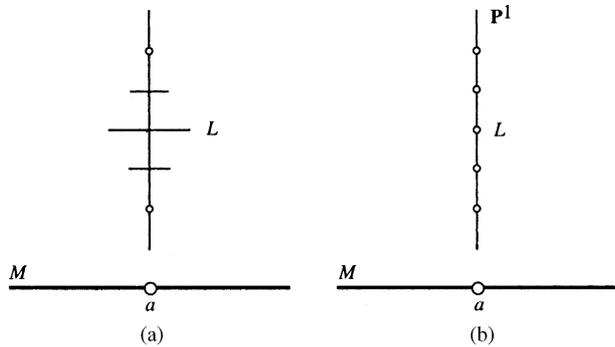
Fig. 12. The fiber bundle $E = V$ with base space the retinal plane $M = R$ (represented as a line for simplicity) and fiber the projective line $\mathbb{P}^1$ of directions in $M$: (a) the elements of the fiber above $a$ are represented as rotating horizontal segments in perspective and (b) the elements of the fiber above $a$ are represented as points (the coordinate in the fiber codes the angle $\theta$ of the direction $p$).

$P_a$, glued together by local coordinate changes in the base space $R$, is a fiber bundle. We will see that it is in fact the *contact* bundle $CR$ of $R$, or in other words the projectivization of the tangent bundle $TR$ of $R$. The points of $V$, namely pairs $(a, p)$ of a point $a$ of $R$ and of a direction $p$ at $a$, are called *contact elements* of $R$ (see Fig. 12).

Via this functional architecture, a (discretized) copy of the projective line $\mathbb{P}^1$ is retinotopically associated to each retinal position $a$. There exists therefore a neural implementation of the fibration $\pi : R \times \mathbb{P}^1 \to R$. The set of feed forward "projections" (in the neurophysiological sense) of the retino-geniculo-cortical ways implements the projection $\pi$ (in the geometrical sense). [3]

### 3.2.3. V1 as a 1-jet bundle

As we will see later, V1 can also be viewed as the bundle of what are called 1-*jets* of curves in $R$. The idea of a jet generalizes the classical notion of Taylor expansion, and gives it an intrinsic geometric meaning. Let us suppose that in a coordinate system $(x, y)$ of $R$ a curve $\gamma$ is the graph $\{x, f(x)\}$ of a real-valued function $f$ on $\mathbb{R}$. The 1st-order jet of $f$, referred to as $j^1 f(x)$, is characterized by three slots: the coordinate $x$, the value of $f$ at $x$, $y = f(x)$, and the value of its derivative $p = f'(x)$. The latter is the slope of the tangent to the graph of $f$ at the point $a = (x, f(x))$ of $R$. So, if we identify $R$ with a domain in $\mathbb{R} \times \mathbb{R}$, a 1-jet is nothing else but a pair $c = (a, p)$. Conversely, to every pair $c = (a, p)$, we can associate the set of functions whose graph is

tangent to $c$ at $a$, that is a 1-jet of curves. In this paper, the bundle of 1-jets of curves in $R$ will be referred to as $J^1 R$. [4]

Jets are feature detectors specialized in the detection of tangents. The fact that V1 can be viewed as $J^1 R$ explains why V1 is functionally relevant for *contour integration*. On the 2-dimensional manifold $R$, to determine the direction $p$ of the tangent to a contour at a point $a$ requires to compare the values of the curve within a neighborhood of that point. But the system can access directly this geometrical information as a single numerical value in the 3-dimensional jet space. This spares a local computation which would be very expensive in terms of wiring.

Jan Koenderink [39] strongly emphasized this fundamental importance of the concept of jet. Without jets, it is difficult to understand how the visual system could extract geometric features such as the tangent or the curvature of a curve at some point.

> Geometrical features become multilocal objects, i.e. in order to compute boundary curvature the processor would have to look at different positions simultaneously, whereas in the case of jets it could establish a format that provides the information by addressing a single location. Routines accessing a single location may aptly be called *point processors*, those accessing multiple locations *array processors*. The difference is crucial in the sense that point processors need no geometrical expertise at all, whereas array processors do (e.g. they have to know the environment or neighbors of a given location). [39, p. 374]

It is effectively the central point. The supplementary variable $p$ allows to convert a geometrical expertise into a functional architecture of $J^1 R$. Neurons are point processors (at the scale defined by the size of their RFs) and can only measure a value at a point. To do differential geometry with point processors, one needs *new variables* evaluating partial derivatives of appropriate degree.

Koenderink emphasized the fact that (hyper)columns implement jet spaces (see also Johnston's paper in this volume):

> The modules (like "cortical columns" in the physiological domain) of the sensorium are local approximations (*Nth order jets*) of the retinal illuminance that can be addressed as a *single datum* by the point processors. [39, p. 374]

---

[3] In the geometrical modeling of neural functional architectures, we meet lexical conflicts. Lexical terms such "fiber", "projection", "connection", etc. are used with different meanings by mathematicians and neurophysiologists. The context allows in general to clear the ambiguity.

[4] The exact classical notation would be $J^1(\mathbb{R}, \mathbb{R})$.

### 3.2.4. Generalization of the model

It is possible to extend this model to other characteristic variables of the visual signal which are represented in the hypercolumns. For that purpose, we need to consider the manifolds in which these variables vary: $[0, 1]$ for the rate of ocular dominance, the projective plane $\mathbb{P}^2$ for color, the circle $\mathbb{S}^1$ for the set of the directions of movement (for further information see e.g. [80] or [76]). V1 will then be modeled as a fiber bundle with base $R$ and with fiber the *cartesian product* of the spaces of secondary variables.

### 3.2.5. Some differences between neurophysiology and its geometrical idealization

There exist important differences between the neurophysiological data and the geometrical idealization of the fibration. Let us evoke three of them.

- To begin with, the RFs introduce a scale of resolution and we need therefore a *multi-scale* theory of fiber bundles. Moreover the RFs are adaptive and modulated by the stimuli.
- Next, there exists a strong *redundancy* of the columns: a "point" $(a, p)$ of the fibration corresponds in fact to an entire column. This "population coding" is essential for the adaptive property and the refinement of the resolution. It allows also an *oscillatory* response and therefore synchronizing and "phase locking" phenomena.
- Finally, we have to tackle a fundamental *dimensional* constraint. The fibration $\pi : R \times P \to R$ is intrinsically of dimension 3 (2 degrees of freedom for the retinal position $a = (x, y)$, 1 degree of freedom for the orientation $p$), while the cortical layers are essentially of dimension 2. There exists therefore a problem of dimensional collapse. The visual system solved it via the fascinating "pinwheel" structure analyzed (in particular) by Bonhöffer, Blasdel and Grinvald.

### 3.3. Pinwheels

### 3.3.1. Experimental data on pinwheels

The method introduced by Bonhöffer and Grinvald [4] in the early nineties of "in vivo optical imaging based on activity-dependent intrinsic signals" allowed to acquire direct images of the activity of the superficial cortical layers. High contrast gratings are presented many times (20–80) with, e.g., a width of 6.25° for the dark strips and of 1.25° for the light ones, a velocity of 22.5°/s, and eight different orientations. A window is opened above V1 and the cortex is illuminated with an infrared light. Differential absorption patterns, resulting from spatial non-homogeneities in the local deoxyhemoglobin/oxyhemoglobin ratio, are observed according to the grating's orientation. One does the summation of the images of V1's activity for the different gratings and constructs differential maps (differences between orthogonal gratings). The low frequency noise is then eliminated and the maps are normalized by dividing the deviation relative to the mean value at each pixel by the global mean deviation. One gets that way precise maps of V1 orientation cells (Fig. 13).

In the following picture (Fig. 14) taken from the seminal paper of William Bosking et al. [5], the orientations are coded by colors and iso-orientation lines are therefore represented by isocolor lines.

We note that there are three classes of qualitatively different points:

1. Regular points, where the orientation field is locally trivial in the sense that the iso-orientation lines are approximatively parallel.
2. Singular points at the center of the pinwheels, where all orientations converge. We see clearly the opposed chirality (clockwise and counterclockwise) of adjacent singular points.
3. Saddle points, where iso-orientation lines bifurcate: two near iso-orientation lines start from the same singular point but end at two different singular points.

If we idealize the situation, we get the model of a lattice of singular points which are the centers of local pinwheels globally glued together (Fig. 15).

We must emphasize the fact that in general the direction of a ray of a pinwheel is *not* the orientation associated with it. When the ray spins around the singular center with an angle $\theta$, the associated orientation rotates with an angle $\theta/2$. This implies that two diametrically opposed rays correspond to orthogonal orientations.

But how can this 2D pinwheel structure fit with the more abstract 3D fibration model? We will see now, using the geometrical concept of a "*blowing up*", that it can be considered as a *discrete* approximation of it.

### 3.3.2. A "blowing up" model of pinwheels

In algebraic geometry, the blowing up of a manifold at a point, e.g. of the plane $M = \mathbb{R}^2$ at the origin $O = (0, 0)$, is the following operation. Let $a = (x, y) \neq (0, 0)$ be a point in $\mathbb{R}^2$. We can associate the direction $Oa$ to it and define a map $\delta$

$$\delta : \mathbb{R}^2 - \{O\} \to \mathbb{P}^1,$$
$$a = (x, y) \mapsto \delta(a) = p = \frac{y}{x}.$$

The graph of $\delta$ is a 2D algebraic ruled surface $H$ in the 3D fibration $V = \mathbb{R}^2 \times \mathbb{P}^1$ whose topological closure is an helicoïd $\overline{H}$. The restriction to $\overline{H}$ of the projection $\pi : \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$ is an isomorphism of $\overline{H}$ to $\mathbb{R}^2$ outside $O = (0, 0)$. If $d$ is the straight line $Oa$ in $\mathbb{R}^2$ then its inverse image $d' = \pi^{-1}(d - \{O\})$ is constituted by the
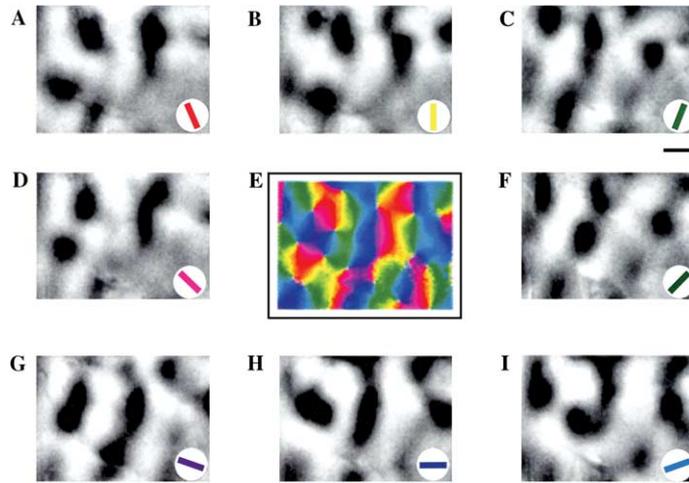
Fig. 13. Maps of the activity of V1 for different orientations. The superposition of the maps provides a color map where colors code the orientation (from [10]).
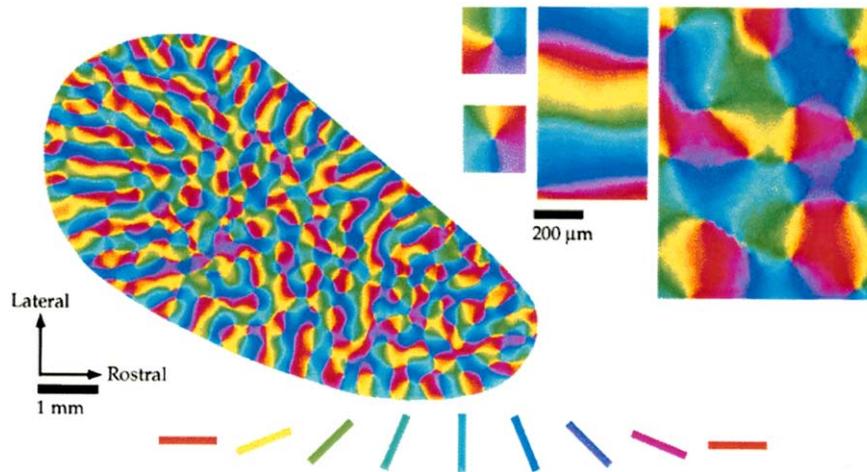


Fig. 14. The pinwheel structure of V1 for a tree shrew. The different orientations are coded by colors. Examples of regular points and of singularities with opposed chiralities are zoomed in (from [5]).
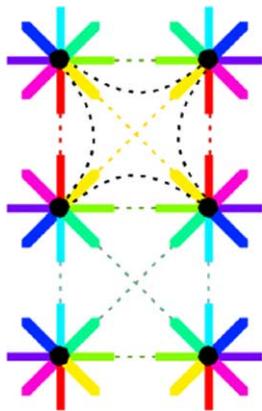


Fig. 15. An idealized "cristal" model of pinwheels centered on a regular lattice of singular points. Some iso-orientations lines are represented. The saddle points in the centers of the domains are well visible.

points $(\lambda a, \delta(a) = p)$ of $V = \mathbb{R}^2 \times \mathbb{P}^1$, that is by the line $d$ at the height $\delta(a) = p = \frac{y}{x}$. When the line $d$ rotates in the plane $\mathbb{R}^2$, $d'$ rotates in $\mathbb{R}^2$ but translates also in the $\mathbb{P}^1$ direction, hence the helicoïdal movement. The inverse image of the blown up point $O$ is the *entire* projective line $\mathbb{P}^1$ and therefore $\pi$ is by no means an isomorphism at $O$ but a projection *collapsing a 1D fibre onto a 0D point*. In that sense, the blowing up of the plane at a point provides a geometrical structure which is, in some sense, *intermediary* between a 2D plane and a 3D fibration over it: it is the fibration $V$ over $O$ and the plane $\mathbb{R}^2$ outside $O$. We can say that the blowing up $\pi : \overline{H} \to \mathbb{R}^2$ unfolds in a third dimension the orientation wheel centered on $O = (0,0)$ (Figs. 16 and 17).

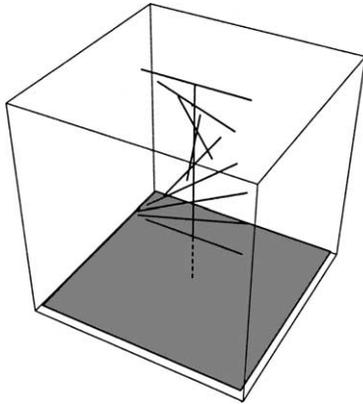This construction can be viewed as an interpretation of the polar coordinates in terms of the fibration $\pi_1$ :

Fig. 16. The blowing of a plane at a point $a$. The directions at $a$ are unfolded in a third dimension and constitute an helicoidal surface.
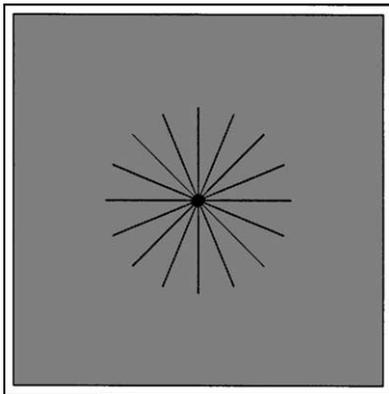


Fig. 17. When the third dimension collapses, the blowing up of a plane at a point becomes a pinwheel.

$\mathbb{R}^2 \times \mathbb{S}^1 \to \mathbb{R}^2$. Indeed, on the punctuated plane $\mathbb{R}^2 - \{O\}$, the argument $\theta(a) \in [0, 2\pi]$ of a point $a$ is well defined and we can therefore consider the section $\vartheta_1$ of $\pi_1$ defined by $\vartheta_1 : a \to \vartheta_1(a) = (a, e^{i\theta(a)})$. The fibration $\pi : \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$ is the quotient of the fibration $\pi_1 : \mathbb{R}^2 \times \mathbb{S}^1 \to \mathbb{R}^2$ by the identification of $\theta$ with $\theta + \pi$ (that is of $e^{i\theta}$ with $-e^{i\theta}$) and $\vartheta_1$ lifts to $\pi_1$ the section of $\pi$ $\vartheta : a \to \vartheta(a) = (a, e^{i\theta(a)})$ where $\theta(a) \in [0, \pi]$ is now considered modulo $\pi$. $\vartheta_1(a)$ is constant on the rays $\theta = \text{cst}$ and, when it is lifted from $\mathbb{R}^2 \times \mathbb{P}^1$ to $\mathbb{R}^2 \times \mathbb{S}^1$, the surface $H$ becomes the image of $\vartheta_1$.

The concept of blowing up was introduced at the beginning of the 20th century by specialists of projective algebraic geometry under the name of "quadratic transformation". It is one of the simplest case of *birational transformation* and is particularly useful for *desingularizing* singular curves. If a curve $\gamma$ in $\mathbb{R}^2$ has a singular point in $O$ presenting many branches with different tangents, then its lifting to $\overline{H}$, $\Gamma = \pi^{-1}(\gamma)$, presents branches of differents heights and therefore the crossing has been removed.

We can *localize* this algebraic model and consider its infinitesimal version where we look only at points $a = (dx, dy)$ infinitely close to the blown up point $O = (0, 0)$. This amounts to take what is called the *germ* of the structure at $O$. As was pointed out to me by Pierre Cartier, [5] to blow up a point is like to cut out an infinitesimal disk centered on it. In the infinitesimal model we have $p = dy/dx$ and the surface $H$ is therefore included *in the kernel of the differential* 1-*form* $\omega = dy - p\,dx$ defined on $V = \mathbb{R}^2 \times \mathbb{P}^1$ (we will return on this key point later in Section 4.2.3). Reciprocally, the algebraic model can be considered as the tangent structure of the infinitesimal one where infinitesimal vectors are substituted by tangent vectors.

We will use the blowing up model to represent the empirical data of the previous section. As a first approximation, we presented there the pinwheel structure omitting the redundancy of the cortical columns orthogonal to the layer surface. This was legitimate since the simple cells of a same column detect essentially the same pair $(a, p)$ (we do not take into account the variation of phase). *But it is definitely no longer the case at the center of a pinwheel*. Maldonado et al. [47] have analyzed the fine-grained structure of orientation maps at the singularities in the cat striate cortex. They found that

> orientation columns contain sharply tuned neurons of different orientation preference lying in close proximity.

In a certain sense, "all" orientations are present at the singular point. Moreover, the orientation associated to a ray of a pinwheel selects this orientation at the center, which means that the operation of closure $\overline{H}$ of $H$ is neurally implemented. This result can be interpreted saying that the singular point is blown-up inside the thickness of the cortical layer. The existence of *complex* cells without a specific orientation tuning is also an argument in favor of the blowing up model.

We must emphasize again that this model is a model of the *functional* geometry of a pinwheel and not of the orientations associated to its rays. Indeed, as we stressed at the end of the previous section, two diametrally opposed rays correspond to orthogonal orientations. Therefore, orientations do not correspond to the section $\vartheta : a \to \vartheta(a) = (a, e^{i\theta(a)})$, $\theta(a) \in [0, \pi]$, of the fibration $\mathbb{R}^2 - \{O\} \times \mathbb{P}^1 \to \mathbb{R}^2 - \{O\}$ but to the section $\psi : a \to \psi(a) = (a, e^{i\theta(a)/2})$ of the fibration $\mathbb{R}^2 - \{O\} \times \mathbb{P}^1 \to \mathbb{R}^2 - \{O\}$. We need *two* $\pi$-rotations (that is a $2\pi$-rotation of $\theta$) to get the same orientation again. Daniel Bennequin pointed out to me this could mean that pinwheels implement in fact what is called in geometry a *spin structure*.
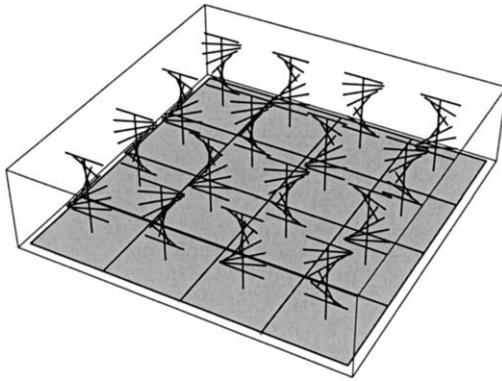
---

[5] Private communication.

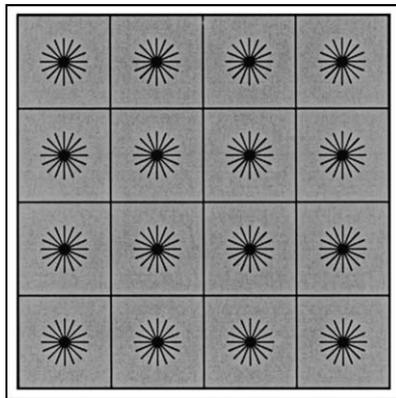Fig. 18. The parallel blowing up of a lattice of points in the plane.



Fig. 19. When the third dimension collapses, a parallel blowing up of a lattice of points in the plane becomes a lattice of pinwheels.

### 3.3.3. From the blowing up model to the fibration model

To recover the fibration model from the blowing up model of a single pinwheel, we have to blow-up in parallel *all* the points of a lattice $L$. This is not possible in the framework of *algebraic* geometry because, to iterate the algebraic model, we have to embed successive blowings up into embedding spaces of greater and greater dimension. But we can do it in the framework of differential geometry by gluing together the local models for the different points of the lattice. We get that way a rather exact model of the pinwheel structure (Figs. 18 and 19).

Now, we can consider that, as the mesh of the lattice $L$ tends to 0, the limit of this parallel multi-blowing-up is the fibration $\pi : \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$ gluing together the infinitesimal model of all the points of $\mathbb{R}^2$. In that sense, the pinwheel structure can effectively be considered as a discrete approximation of the fibration $\pi$ and, reciprocally, $\pi$ can be considered as the simultaneous blowing up of *all* the points in the plane. In fact, we get that way not only the fibration $\pi : \mathbb{R}^2 \times \mathbb{P}^1 \to \mathbb{R}^2$ but also a supplementary infinitesimal structure defined on it by the 1-form $\omega = \mathrm{d}y - p\,\mathrm{d}x$. We will see below that it is exactly the contact structure of $\pi$.

But if we do not take the limit and keep the lattice $L$ with a finite mesh, we can consider the field of orienta-

tions $\vartheta(a) = \mathrm{e}^{\mathrm{i}\theta(a)}$ of the simple cells of V1 as a section of the fibration $\pi : \mathbb{R}^2 \times \mathbb{S}^1 \to \mathbb{R}^2$ defined over $\mathbb{R}^2 - L$. We can use the Fourier transform and look at linear superpositions of the form

$$\sum_{k=1}^{k=N} c_k \mathrm{e}^{2\mathrm{i}\pi(x\cos(2\pi k/N) + y\sin(2\pi k/N))}.$$

Fig. 20 (see p. 289), shows two examples for $N = 256$, the coefficients $c_k$ being random numbers in $[0, 1]$.

We can also notice that the rays of the pinwheels leave space for coding a supplementary parameter. It seems (De Angelis et al. [13], Bressloff et al. [82]) that it could be spatial frequency. At the limit, this parameter disappears and we recover the fact that classical geometry is an idealization having infinite resolution, no scale, and no spatial frequency.

### 3.4. The neural implementation of the topological and differentiable manifold structure

Let us be a little more precise concerning the implementation of the global geometric coherence of the subjective, immanent, internal, phenomenal space (which has to be carefully distinguished from the objective, trancendant, external, physical space).

#### 3.4.1. The problem of an immanent global space

The visual system constructs successive retinotopic representations of the retinal space whose global "spatial" structure is rather problematic. Indeed, when two neighboring ganglion cells have overlapping RFs they share common photoreceptors through intermediary (horizontal and amacrine) cells, but when their RFs are disconnected, it becomes impossible to compare their activities directly. In that sense, the immanent retinal space is very different from an Euclidean domain since at the retinal level, *there exists no global geometrical structure*. While two distant parts of the Euclidean plane can always be compared by translations, there is no possibility to compare immanently two non-overlapping receptive fields. In other words, the Euclidean space possesses a *global* structure of homogeneous space (the group of its automorphisms acts transitively upon it) while the retinal space results basically from "*gluing*" *local domains* (the receptive fields). Space is not given. It must be *constituted* as a very special type of representation.

The same phenomenon holds for each of the successive levels of representation of the visual field: two cells with non-overlapping receptive fields and no interconnections would "ignore" each other. Long-range "vertical" connections link successive levels of processing but if "horizontal" connections within a single level were only short-ranged, the continuation from local structures to global ones would be impossible. We have

therefore to carefully distinguish two completely different types of structure:

(i) the local one which defines the differentiable manifold structure, and
(ii) the global one which allows parallel transport and other comparisons between two distant points.

### 3.4.2. The neural implementation of the topology

The idea of passing from local structures to global ones by gluing processes has become fundamental in modern geometry: since a long time [6] mathematicians have been interested in the properties of spaces constructed by gluing local domains, to which they have extended the powerful tools of differential calculus. To this purpose they introduced the concept of a *differentiable manifold* defined on an underlying topological space. Intuitively, a differentiable manifold is a topological space locally homeomorphic to a standard space $\mathbb{R}^n$ through systems of local coordinates (called local charts), the coordinate transformations being differentiable maps. [7]

In what concerns the underlying topological space, Koenderink [38] has shown how it can be constructed from the functional relationships to which the system has internally access. The gluing of local domains (overlapping receptive fields) is coded *by the temporal correlations* of the signals along the associated nervous fibers. This fine grained temporal information (the only one internally available to the system) encodes the
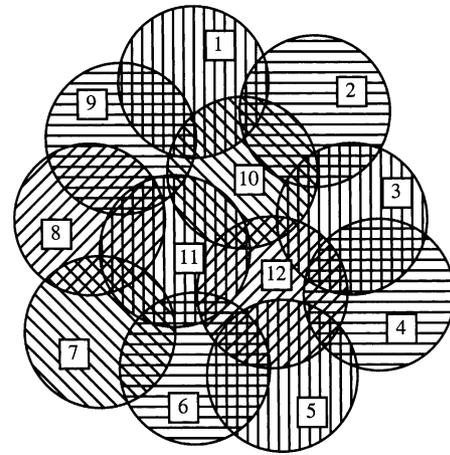


Fig. 21. A covering of a topological space by a set of open subsets. The nerve of the covering is the simplicial structure defined by the intersections.

topological structure resulting from gluing: indeed the intersection structure of any open covering of a space characterizes its topology (it is the base of what is called technically Čech cohomology). More precisely, if $\mathscr{U} = (U_i)_{i \in I}$ is an open covering of a manifold $M$, one associates to it a combinatorial structure $\mathscr{N}(\mathscr{U})$ (a simplicial complex called the "nerve" of $U$) by considering the intersections $U_i \cap U_j$, $U_i \cap U_j \cap U_k$, etc., which are *not empty* (Fig. 21).

In Fig. 21, there are 12 open subsets $U_i$. For the $U_i \cap U_j$ the intersection matrix is the following:

|    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|
| 1  | × | × |   |   |   |   |   |   | × | ×  |    |    |
| 2  | × | × | × |   |   |   |   |   |   | ×  |    |    |
| 3  |   | × | × | × |   |   |   |   |   | ×  |    | ×  |
| 4  |   |   | × | × | × |   |   |   |   |    |    | ×  |
| 5  |   |   |   | × | × | × |   |   |   |    |    | ×  |
| 6  |   |   |   |   | × | × | × |   |   |    | ×  | ×  |
| 7  |   |   |   |   |   | × | × | × |   |    | ×  |    |
| 8  |   |   |   |   |   |   | × | × | × |    | ×  |    |
| 9  | × |   |   |   |   |   |   | × | × | ×  | ×  |    |
| 10 | × | × | × |   |   |   |   |   | × | ×  | ×  | ×  |
| 11 |   |   |   |   |   | × | × | × | × | ×  | ×  | ×  |
| 12 |   |   | × | × | × | × |   |   |   | ×  | ×  | ×  |

For the intersections $U_i \cap U_j \cap U_k$ we get the 13 triples: $(1, 2, 10)$, $(1, 10, 9)$, $(2, 3, 10)$, $(3, 4, 12)$, $(3, 10, 12)$, $(4, 5, 12)$, $(5, 6, 12)$, $(6, 11, 12)$, $(6, 7, 11)$, $(7, 8, 11)$, $(8, 9, 11)$, $(9, 10, 11)$, $(10, 11, 12)$.

$\mathscr{N}(\mathscr{U})$ characterizes to a certain extent the structure of $M$. Koenderink's hypothesis is that, for $M = $ "the visual field" and $U = $ "the covering of $M$ by the receptive fields of ganglion cells", $\mathscr{N}(\mathscr{U})$ is encoded in the

---

[6] The idea comes back to Gauss at the beginning of the 19th century ($\sim$1820–1830). Riemann introduced the concept of a Riemannian manifold (*Mannigfaltigkeit*) in his 1854 *Habilitationsarbeit* "Über die Hypothesen, welche der Geometrie zu Grunde liegen" (see his *Gesammelte Mathematische Werke*), and Hermann Weyl introduced the concept of an abstract differentiable manifold at the beginning of the 20th century (1913) in his masterpiece *Die Idee der Riemannschen Fläche* (Leipzig, Berlin, B.G. Teubner).
[7] We cannot rigorously define here the technical concept of a differentiable manifold: see e.g. Spivak [65].

temporal correlations of the signals along the fibers of the optic nerve. Koenderink and his co-workers have shown, through simulation, how such a structure can be set up during the development of the visual system [67].

### 3.4.3. The neural implementation of the differentiable structure

More structured than the concept of a topological space, the concept of a manifold was introduced in order to provide a class of generalized spaces on which it is possible to do differential geometry. Since every point is surrounded by a neighborhood isomorphic to a domain of a classical space, all the operations of differential geometry, which are *local*, can be transferred to manifolds. Koenderink [40] proposed the idea that the tangent space at one point could be implemented by a cortical hypercolumn above this point. According to him, each of the cells in the hypercolumn would implement a tangent vector, while the connections within the hypercolumn would implement the vectorial operations of addition and scalar multiplication. To evaluate this hypothesis, let us first define these mathematical notions more precisely.

The classical dual notions of tangent and cotangent vectors easily generalize to manifolds. If the local parametric equation of a curve $\gamma$ in the $n$-manifold $M$ is $x(t) = (x_1(t), \ldots, x_n(t))$, then its tangent vector $X(t)$ at $x(t)$ will be

$$X(t) = (X_1(t), \ldots, X_n(t)) = \left( \frac{\mathrm{d}x_1(t)}{\mathrm{d}t}, \ldots, \frac{\mathrm{d}x_n(t)}{\mathrm{d}t} \right)$$
$$= \frac{\mathrm{d}x(t)}{\mathrm{d}t}. \tag{9}$$

It is obvious that these tangent vectors constitute a vector space on $\mathbb{R}$. It is noted $T_x M$ and called the tangent space of $M$ at $x$. The different $T_x M$ glue together and constitute a fiber space $\Pi : TM \to M$ of base $M$ and fiber $\mathbb{R}^n$.

If $\varphi$ is an observable, that is a function $\varphi : M \to \mathbb{R}$, its derivative along the curve $\gamma$ is given by

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{\partial \varphi}{\partial x_i} \frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{i=1}^{n} X_i \frac{\partial \varphi}{\partial x_i}$$
$$= \left( \sum_{i=1}^{n} X_i \frac{\partial}{\partial x_i} \right) (\varphi) = X(\varphi), \tag{10}$$

where $X$ is the derivation operator $X = \sum_{i=1}^{n} X_i \frac{\partial}{\partial x_i}$. Thus tangent vectors can be interpreted as derivation operators on observables. The particular derivations $\frac{\partial}{\partial x_i}$ associated to the local coordinates $x_i$ constitute a basis for the vector space $T_x M$.

Differential forms $\omega$ are the dual entities of tangent vectors and therefore are also called cotangent vectors: they are linear forms on tangent vectors. If $X = \sum_{i=1}^{n} X_i \frac{\partial}{\partial x_i}$ is a tangent vector, then $\omega(X)$ will be a number depending linearly both on $X$ and $\omega$. If $\left( \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} \right)$ is the basis of the tangent space $T_x M$ of $M$ at $x$ associated to the local coordinates $(x_1, \ldots, x_n)$, then the dual basis of the cotangent space is $(\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$. Thus $\mathrm{d}x_j \left( \frac{\partial}{\partial x_i} \right) = \delta_{ji}$ (where $\delta_{ji}$ is the Kronecker symbol $\delta_{ji} = 1$ if $i = j$ and $\delta_{ji} = 0$ otherwise), and if $\omega = \sum_{j=1}^{n} \omega_j \, \mathrm{d}x_j$ we get

$$\omega(X) = \sum_{i,j=1}^{n} \omega_j X_i \delta_{ji} = \sum_{i=1}^{n} \omega_i X_i = \langle \omega, X \rangle. \tag{11}$$

If $\varphi$ is an observable, we can associate to it the differential form $\mathrm{d}\varphi = \sum_{j=1}^{n} \frac{\partial \varphi}{\partial x_j} \mathrm{d}x_j$. Then we have the fundamental *duality*

$$\mathrm{d}\varphi(X) = \sum_{i=1}^{n} \frac{\partial \varphi}{\partial x_i} X_i = X(\varphi). \tag{12}$$

Let us now consider a cell whose receptive profile is (an approximation of) a Gaussian $G$. As we have seen in Section 2.2, it operates on the signal $I$ through the convolution [8] $A = G * I$. We have also seen that a fundamental property of convolution is its behavior relative to derivation

$$\frac{\partial}{\partial x} (\varphi * \psi) = \frac{\partial \varphi}{\partial x} * \psi = \varphi * \frac{\partial \psi}{\partial x} \tag{13}$$

for two distributions $\varphi$ and $\psi$. It follows that a cell with an RP $\frac{\partial G}{\partial x}$ operates via:

$$A' = \frac{\partial G}{\partial x} * I = G * \frac{\partial I}{\partial x} = \frac{\partial}{\partial x} (G * I) = \frac{\partial A}{\partial x}. \tag{14}$$

The profile $\frac{\partial G}{\partial x}$ operating on $I$ provides therefore the directional derivative $X(I)$ (with $X = \frac{\partial}{\partial x}$).

But the neurophysiological data seem to invalidate the hypothesis of a *complete* neural implementation of the tangent and cotangent spaces of the retinal space. Indeed:

1. Although all *directions* are represented within a hypercolumn, there is no such evidence for the vectors themselves.
2. Even if the implementation of the vector structure (vector addition and scalar multiplication) is theoret-

---

[8] The classical formula for a Gaussian $G$ of width $\sigma$ on the plane is $\frac{1}{2\pi\sigma^2} \mathrm{e}^{-(r^2/2\sigma^2)}$ (where $r$ is the length of the radius $Oa$). We have seen in Section 2.2 that in scale-space geometry, the scale $s$ is linked to the width $\sigma$ of $G$ by the formula $2s = \sigma^2$.

ically possible, it would require lots of connections which have not been observed.

3. As we will see, the main function of orientation columns is to represent contours. Now, the tangent vector to a curve at one point depends on the parametrization and has therefore no intrinsic geometric meaning, while the tangent direction is an intrinsic geometrical entity.

What seems to be represented over every point $a$ of the retinal field $R$ via the pinwheels is therefore not the tangent plane $T_aR$ itself, but only its *projectivization*, namely the projective line $\mathbb{P}^1$ of tangent directions.

## 4. Horizontal connections and the contact structure of V1

Up to now, we have modeled the "vertical" structure of connections defining the fibration $\pi : M \times P \to P$ with base space $M = R$ and fiber $P = \mathbb{P}^1$ or $P = \mathbb{S}^1$. [9] But such a "vertical" structure is largely not sufficient. As we already previously emphasized, to implement a *global* coherence, the visual system must be able to *compare* two retinotopically neighboring fibers $P_a$ and $P_b$ over two neighboring points $a$ and $b$ of $M$. This is a problem of *parallel transport*. It has been solved at the empirical level by the discovery of "horizontal" cortico-cortical connections (see e.g. [11]).

### 4.1. Horizontal connections

Horizontal connections are long ranged (up to 6–8 mm) and connect cells of the *same* orientation in distant hypercolumns. To detect them (see e.g. Ts'o et al. [69]) one can

(i) measure the correlations between cells belonging to different hypercolumns;
(ii) compare the preferred orientation of a reference cell with the preferred orientation of other cells met along a cortical penetration;
(iii) compute cross-correlograms.

One verify that cells with neighboring orientations are strongly correlated while cells with sufficiently different orientations are decorrelated (Fig. 22).

The beautiful next figure (Fig. 23, see p. 289) is due to William Bosking [5]. It shows how biocytin injected locally in a zone of specific orientation (green–blue) diffuses via horizontal cortico-cortical connections.
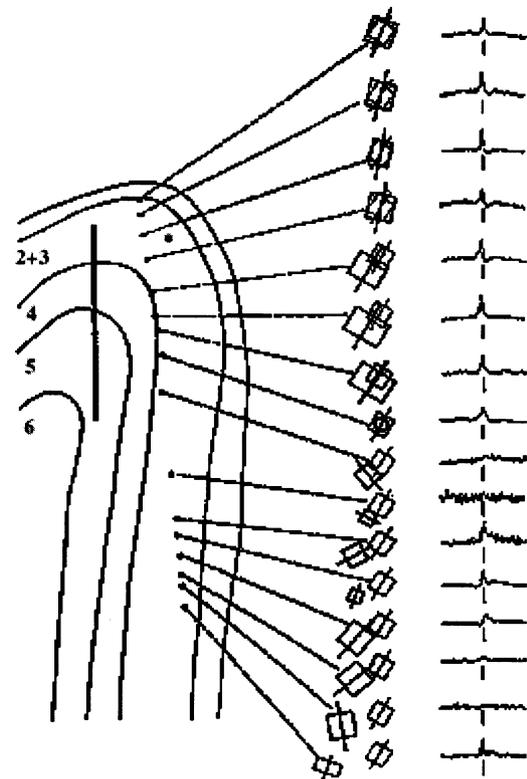


Fig. 22. Examples of cortico-cortical horizontal connections. The orientation cells are symbolized by a square (their receptive field) and a line (their preferred orientation). The ★ represents the reference cell. Electrodes are placed on other cells $N$. For each pair $(\star, N)$ the cross-correlogram is shown. When there exists a peak, this means that the two cells ★ and $N$ are synchronized and therefore connected. We see that the most strongly connected cells are those sharing the same preferred orientation (from [69]).

The key experimental fact is that short range diffusion is isotropic, while long range diffusion is, on the contrary, highly anisotropic and restricted to cortical zones sharing essentially the same orientation (the same color) as the injection site. These different types of connections implement two different levels of structure:

(i) the short range connections implement the local triviality of the fibration $\pi : M \times P \to P$, while
(ii) the long range connections implement a richer structure.

One could think that horizontal cortico-cortical connections violate retinotopy. But in fact they reinforce it by insuring its large scale coherence. Without them, neighboring hypercolumns would become independent and retinotopy would lose any *immanent* reality for the system itself.

That cortico-cortical connections connect neurons of the same orientation in different hypercolumns means that the system is able to know, for $b$ different from $a$, if an orientation $p$ at $a$ is the same as an orientation $q$ at $b$.

---

[9] For greater generality, we use in the following the more generic notations $M$ and $P$. $M$ is the retina $R$ or the cortical layer and $P$ is the space of orientations.
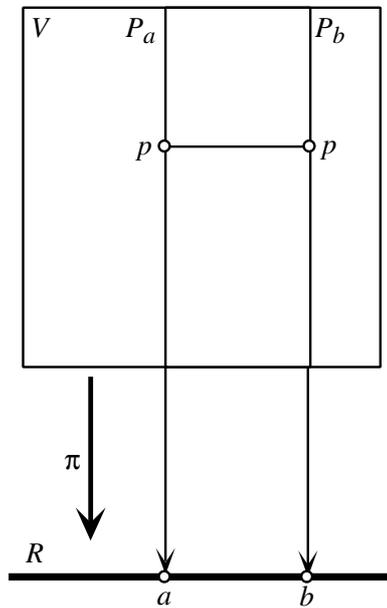
Fig. 24. Cortico-cortical horizontal connections allow the system to compare orientations in two different hypercolumns corresponding to two different retinal positions *a* and *b* (schematic representation).

In other words, while the retino-geniculo-cortical "vertical" connections provide an internal meaning for the system to the relations between pairs $(a,p)$ and $(a,q)$ (different orientations $p$ and $q$ at the same point $a$), the "horizontal" cortico-cortical connections provide an internal meaning for the system to the relations between pairs $(a,p)$ and $(b,p)$ (same orientation $p$ at different points $a$ and $b$) (Figs. 24 and 25).

Moreover, it can be shown that cortico-cortical connections preferentially connect neurons detecting not only parallel (co-oriented) pairs $(a,p)$ and $(b,p)$, but also *co-axial* pairs, that is pairs such that $p$ is the orientation of the axis $ab$ (Fig. 26).

### 4.2. The contact structure of V1

We will now present in this section the most important geometrical structure of the fibration $\pi : M \times P \to P$, namely its *contact structure* $\mathscr{C}$ which models quite exactly the functional architecture of V1. We are aware that the topic is mathematically rather technical but we think that, in order to understand really the concepts at stake, it is necessary to describe their formal content.

### 4.2.1. The contact bundle and the 1-jet space

In general, if $M$ is a $n$-manifold, one can consider at every point $a$ of $M$, not the tangent vector space $T_aM$, but the set of its hyperplanes (linear sub-spaces of co-dimension 1), referred to as $C_aM$. $C_aM$ is isomorphic to the projective space $\mathbb{P}^{n-1}$. The total space gluing these fibers is the contact bundle of $M$ and is referred to as
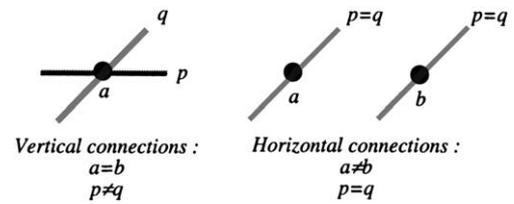


Fig. 25. While the retino-geniculo-cortical "vertical" connections give a meaning to the relations between pairs $(a,p)$ and $(a,q)$ (different orientations $p$ and $q$ at the same point $a$), the "horizontal" cortico-cortical connections give a meaning to the relations between pairs $(a,p)$ and $(b,p)$ (same orientation $p$ at different points $a$ and $b$).
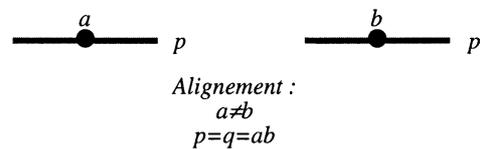


Fig. 26. Cortico-cortical connections connect preferentially neurons detecting not only co-oriented but also co-axial pairs $(a,p)$ and $(b,p)$ where $p$ is the orientation of the axis $ab$.

$CM$. In our case, $M$ is 2-dimensional, the hyperplanes are orientations and $CM = M \times P$. The set of contact elements at $a$, $C_aM$, is isomorphic to the projective line $P = \mathbb{P}^1$ (the set of plane directions).

We can interpret the coordinate $p$ in terms of $T_aM$. If $(x,y)$ are local coordinates at $a$, the tangent plane $T_aM$ is naturally endowed with the associated coordinates $(\xi, \eta)$ in the natural basis $\left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)$. Then, on an open set not containing the "vertical" line $\xi = 0$, a local coordinate of $C_aM$ is $p = \frac{\eta}{\xi}$ (in the neighborhood of $\xi = 0$, we can choose the coordinate $p = \frac{\xi}{\eta}$). An element $c$ of $CM$ is therefore identified by the coordinates $(x,y,p) = (a,p)$, and $CM$ is a 3-manifold isomorphic to $V = M \times P$.

As we already sketched it in Section 3.2.3, in the case where $M$ is a domain of $\mathbb{R} \times \mathbb{R}$, the contact bundle of $M$ is closely related to the bundle of 1-jets of curves in $M$, referred to as $J^1M$ (Fig. 27). $C_aM$ is the compactification
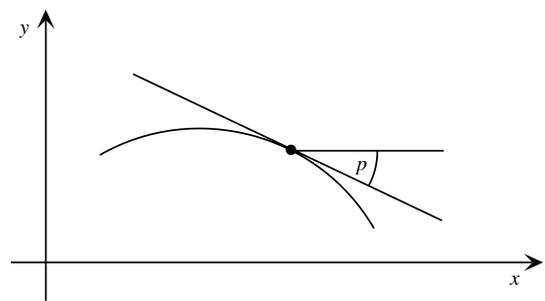


Fig. 27. A 1-jet is a pair $(a,p) = (\text{position}, \text{orientation})$ representing the equivalence class of the differentiable curves going through $a$ with tangent $p$.

of $J_a^1 M$ where the symbols $\infty$ and $-\infty$ are identified i.e. where a point at infinity is added.

Let us emphasize once again that jet spaces are of fundamental importance because they reduce *local* computations to *punctual* ones. The counterpart of this drastic simplification is to increase the number of variables: instead of considering the plane $(x,y)$, and then computing $y' = \frac{dy}{dx}$ (which requires to know not only the value $y = f(x)$ of $f$ at $x$ but also the values of $f$ in a neighborhood of $x$), we consider the 3-dimensional phase space $(x,y,p)$ *endowed with the constraint* $y' = p$. This deep idea is ubiquitous in mathematical physics and comes back to Hamilton.

### 4.2.2. E(2)-invariance

To choose a frame $(x,y,p)$ of $CM$ and identify $CM$ with (a compactification of ) $J^1 M$ is equivalent to choose a point $O$ of $M$ (which becomes the origin) and the contact plane $C_0$ (which becomes the plane $(x,p)$), the third axis $y$ being orthogonal to $C_0$. But, when we choose such a frame $(x,y)$, we break the symmetry of $R$. This symmetry breaking is compensated by the fact that the structure of $J^1 M$ is invariant under the action of the Euclidean group $E(2) = SO(2) \rtimes \mathbb{R}^2$ of rigid motions in the plane ($E(2)$ is the semi-direct product $\rtimes$ of the rotation group $SO(2)$ and of the translation group $\mathbb{R}^2$).

Let $(p, r_\theta)$ be an element of $E(2)$ where $p$ is a point of $M$ and $r_\theta$ the rotation of angle $\theta$. [10] $(p, r_\theta)$ acts on a point $a$ of $M$ by

$$(p, r_\theta)(a) = p + r_\theta(a). \tag{15}$$

If $(p, r_\theta)$ and $(q, r_\varphi)$ are two elements of $E(2)$, their (non-commutative) product is given by the formula

$$(q, r_\varphi) \circ (p, r_\theta) = (q + r_\varphi(p), r_{\varphi + \theta}). \tag{16}$$

The product is non-commutative for $(p, r_\theta) \circ (q, r_\varphi) = (p + r_\theta(q), r_{\theta + \varphi})$. Of course $r_{\varphi + \theta} = r_{\theta + \varphi}$, but $q + r_\varphi(p) \neq p + r_\theta(q)$ (Fig. 28).

The rotation $r_\theta$ acts on the fibration $J^1 M \to M$ by

$$r_\theta(a, \varphi) = (r_\theta(a), \varphi + \theta) \tag{17}$$

(where $\varphi$ is the angular coordinate corresponding to $p$). This particular form of action warrants the fact that the alignment of preferred directions is $E(2)$-invariant (Fig. 29).

### 4.2.3. Legendrian lifts and the condition of integrability

Let now $\gamma$ be a smooth curve plotted in the manifold $M$. It can be *lifted* to a curve $\Gamma$ in $V = J^1 M$. Indeed, let us consider the 1-jet map $j^1 : \gamma \subset M \to J^1 M$ which associates to every point $a$ of $\gamma$ the 1-jet of $\gamma$ at that point, that is the pair $(a, p_a)$ where $p_a$ is the tangent of $\gamma$ at $a$. $\Gamma = j^1 \gamma$ is the image of $\gamma$ by $j^1$ and is called the *Legendrian lift* of $\gamma$ into $J^1 M$ (or $CM$). $\Gamma$ represents $\gamma$ as *the envelope of its tangents* (Fig. 30).

If $a(s)$ is a parametrization of $\gamma$, we note $a'(s) = y'(s)/x'(s)$ and therefore $\Gamma = (a(s), p(s)) = (a(s), a'(s))$. In terms of the coordinates $x$ and $y$ and of an equation $y = f(x)$ of $\gamma$, the equation of $\Gamma$ is then $(x, y, p) = (x, y = f, y' = f')$.

Let us assume now that we *do not have access* to what happens in the base $M$, and that we are trying to recover it only from what we observe in the total space $V \simeq J^1 M \simeq \mathbb{R}^3$. To every curve $\gamma$ in $M$ is associated a curve $\Gamma$ in $V$. *But the converse is definitely false.* The crucial question is therefore to characterize, among all the skew curves $\Gamma$ in $V$, those which result from the lifting of smooth curves $\gamma$ lying in the base space $M$. Let $\Gamma = v(s) = (a(s), p(s))$ be a (parametrized) curve in $V$. The projection $a(s)$ of $\Gamma$ is a curve $\gamma$ in $M$ and $\Gamma$ is the lifting of $\gamma$ iff $p(s) = a'(s)$. Equivalently, if $\Gamma$ is locally defined by equations $y = f(x)$, $p = g(x)$, there exists a curve $\gamma$ in $M$ such that $\Gamma = j^1 \gamma$ if and only if $g(x) = f'(x)$, that is $p = y'$.

In differential geometry, this condition is called a *Frobenius integrability condition*. It says that to be a *coherent* curve in $V$, $\Gamma$ must be an *integral curve of the contact structure* of the fibration $\pi$ in the following sense. Let $t = (a, p; \alpha, \pi) = (x, y, p; \xi, \eta, \pi)$ be a tangent vector to $V$ at the point $v = (a, p) = (x, y, p)$. If $t$ is tangent to a curve of equation $y = y(x)$ and $p = p(x)$, we have $t = (x, y, p; 1, y', p')$. If $p = y'$, that is if the integrability condition is satisfied, we have therefore $t = (x, y, p; 1, p, p')$. Due to this very special form, $t$ belongs to the *kernel of the differential form* $\omega = dy - p\, dx$ on $V = J^1 M$. Indeed, from the general formula $\omega(t) = \sum \omega_i t_i$, (where $t_i$ and $\omega_i$ are the respective components of $t$ and $\omega$ relative to the bases of $TJ^1 M$ and $T^* J^1 M$ associated to the local coordinates $(x, y, p)$), we get $\omega(t) = -p \cdot 1 + 1 \cdot p + 0 \cdot p' = -p + p = 0$ (we have $\omega = -p\, dx + 1\, dy + 0\, dp$ and $dx$ (respectively, $dy$, $dp$) applied to $(1, p, p')$ selects the first (respectively, the second, the third) component 1 (respectively, $p$, $p'$)). Now, the kernel of a differential form on a 3-dimensional space is a plane. $t$ belongs therefore to a plane $C_v$, tangent to $V = J^1 M$ at $v$ (i.e. $C_v \subset T_v V$) (Fig. 31). [11]

---
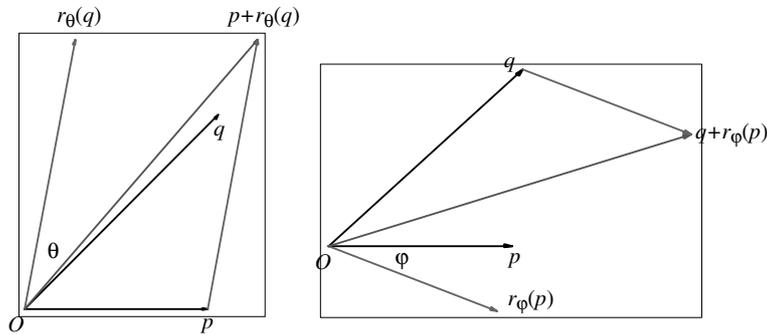
Fig. 28. The non-commutativity of the Euclidean group $E(2)$.
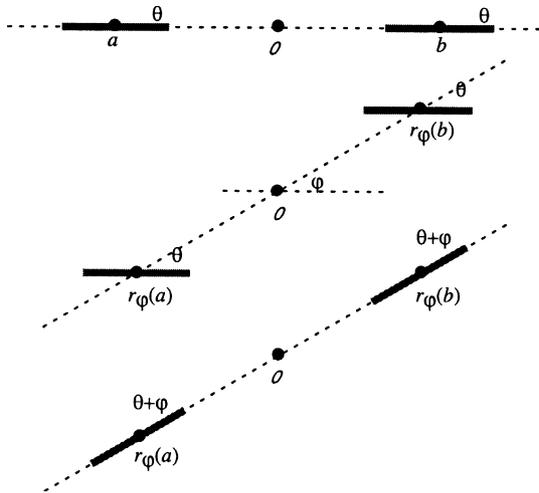


Fig. 29. The $E(2)$-invariance of the contact structure.



Fig. 30. The Legendrian lifting of a curve $\gamma$, $y = f(x)$, in the base space $M$ to the fibration $V = M \times P$. Above every point $(x, y = f(x))$ of $\gamma$ we take the tangent direction $p = f'(x)$.
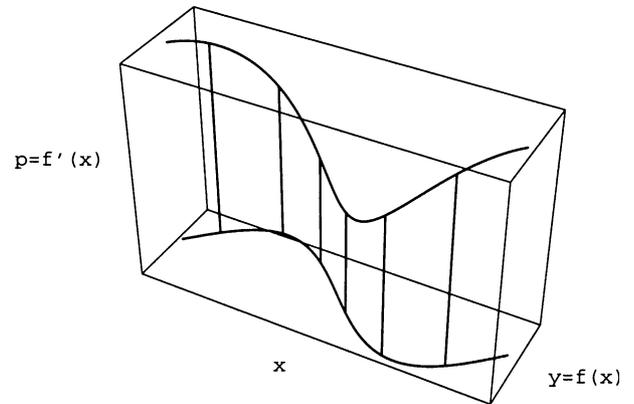
Thus, the tangents to the curves in $V = J^1M$ of the form $j^1\gamma$ belong to the field of planes $\mathscr{C} : v \to C_v$. This field is called the *contact structure* of $J^1M$, and $\omega$ its *contact form*. Since the curves of the form $j^1\gamma$ are tangent at each of their point to this field of planes, they are called *integral curves* of the contact structure (Fig. 31).

Conversely, let us consider a skew curve $\Gamma$ in $J^1M$ of equations $y = f(x)$ and $p = g(x)$. Then, on $\Gamma$, we have $\omega = (f' - g)\,\mathrm{d}x$ (for $\mathrm{d}y = f'\,\mathrm{d}x$ and $p\,\mathrm{d}x = g\,\mathrm{d}x$). If $\Gamma$ is an integral curve of the contact structure, we have $\omega = 0$ on $\Gamma$, then $f' = g$ and $\Gamma$ is effectively of the form $j^1\gamma$.

The contact structure $\mathscr{C}$ is therefore exactly what discriminates the Legendrian lifts of planar curves among the other skew curves in the jet space $V = J^1M$. Through the transformation $u \to -p$, $v \to x$, $w \to y$, it corresponds to the well known classical standard example on $\mathbb{R}^3$ (with coordinates $(u, v, w)$) given by $\omega = u\,\mathrm{d}v + \mathrm{d}w$.

### 4.2.4. The complete non-integrability of the contact structure, non-holonomy, and the Frobenius condition

We must stress a very important fact concerning the contact structure $\mathscr{C}$. It is defined as the field of *planes* $v \in V \mapsto C_v \subset T_vV$ which is the field of kernels of the 1-form $\omega = \mathrm{d}y - p\,\mathrm{d}x$. It is then natural to ask if this 2D field $C_v$ which possesses plenty of 1D integral *curves*, could itself be integrable in the sense that it would exist *surfaces* $S$ of $J^1M$ tangent to $C_v$ at every of their point $v$, so that we would have $T_vS = C_v$. But this is impossible. The field $v \mapsto C_v$ is a prototypical example of a completely *non*-integrable field because it is too "twisted" (Fig. 32).

More technically, one can remark that if $t \in T_vV$ is a tangent vector to $V$ of components $t = (\xi, \eta, \pi)$, the vanishing condition $\omega(t) = 0$ defines the plane $-p\xi + \eta = 0$ i.e. $\eta = p\xi$. Through the identification $V \simeq \mathbb{R}^3$, $C_v$ becomes the "vertical" plane above the "horizontal" line of slope $p$. When one moves along the fiber $P_a = J_a^1M$, this plane rotates with $p$. The non-integrability of the field $C_v$ results from the fact that the 2D *Frobenius integrability condition* $\omega \wedge \mathrm{d}\omega = 0$ (i.e. $\mathrm{d}\omega(t, t') = 0$ for all $t$ and $t'$ such that $\omega(t) = \omega(t') = 0$) is *not* satisfied. The general theorem says that a necessary and sufficient condition of integrability is that, for every basis $\{t_1, t_2\}$ of $C_v$, the Lie bracket $[t_1, t_2]$ belongs to $C_v$, or in other words that $C_v$ is a *Lie sub-algebra* of $T_vJ^1M$. But for $\omega = \mathrm{d}y - p\,\mathrm{d}x$ we have
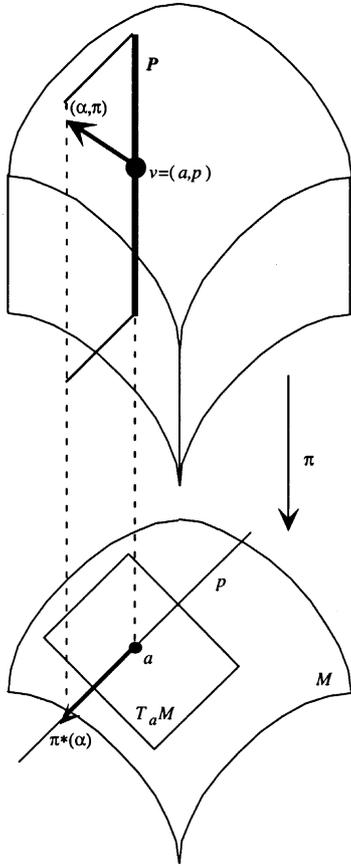
Fig. 31. The contact structure of the contact bundle (or the 1-jet bundle) of a base manifold $M$. Let $a \in M$ and $p$ a hyperplane of the tangent plane $T_a M$. $p$ corresponds to a point in the fiber $P = C_a M$ of $CM$ above $a$. Let $t = (\alpha, \pi)$ be a tangent vector to $CM$ at $v = (a, p)$. $t = (\alpha, \pi)$ belongs to the contact plane $C_v M$ of $CM$ at $v$ iff the projection $\pi_*(\alpha)$ of the "horizontal" component $\alpha$ of $t$ is aligned with $p$.
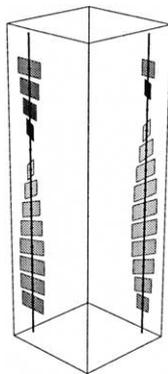


Fig. 32. The field of contact planes is too twisted to be integrable.

$$d\omega = -dp \wedge dx = dx \wedge dp \qquad (18)$$

and therefore

$$\omega \wedge d\omega = (-p\,dx + dy) \wedge dx \wedge dp = dy \wedge dx \wedge dp$$
$$= -dx \wedge dy \wedge dp. \qquad (19)$$

But this 3-form is nothing else than the *volume form* of $J^1 M$ and it is therefore impossible for it to vanish. We can also remark that for the natural basis

$$\left\{ t_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y} = (1, p, 0), t_2 = \frac{\partial}{\partial p} = (0, 0, 1) \right\}$$

of $C_v$ we have $[t_1, t_2] = t_3 = -\frac{\partial}{\partial y} = (0, -1, 0)$ and $t_3 = (0, -1, 0) \notin C_v$ because for this vector $-p\xi + \eta = -0 + (-1) = -1 \neq 0$.

The consequence of the $2D$ non-integrability—also called *non-holonomy*—of the contact structure $\mathscr{C}$ is that its integrals are *necessarily 1-dimensional*. In other words, $\mathscr{C}$ is functionally dedicated to the integration *of curves*. Moreover, the non-integrability of the contact structure expressed by its Lie brackets implies that the geometry of $V$ which models the functional architecture of V1 in terms of $\mathscr{C}$ *is completely different from a classical Euclidean geometry*, even at the infinitesimal level.

### 4.2.5. Contact and Lie structures

It is interesting to note that the contact structure of $V = J^1 M$ can in fact be easily recovered as a translation-invariant structure in an appropriate *Lie group*. Indeed, let us define a product on $V$ by the formula

$$(x, y, p) \cdot (x', y', p') = (x + x', y + y' + px', p + p'). \qquad (20)$$

It is immediate to verify that this composition law is associative, that the origin $(0, 0, 0)$ is a neutral element, and that the inverse of $v = (x, y, p)$ is $v^{-1} = (-x, -y + px, -p)$. But, due to the asymmetry of the term $px'$, the product is *non-commutative*.

The Lie algebra of $V$ is the vector space $\mathfrak{v} = T_0 V$ endowed with the Lie bracket

$$[t, t'] = [(\xi, \eta, \pi), (\xi', \eta', \pi')] = (0, \xi'\pi - \xi\pi', 0). \qquad (21)$$

Let us consider the *left translation* $L_v$ defined by $L_v(v') = v \cdot v'$. It is a non-linear diffeomorphism of $V$ whose tangent application at 0 is the linear map

$$T_0 L_v : T_0 V \to T_v V$$

$$t = (\xi, \eta, \pi) \mapsto T_0 L_v(t) = (\xi, \eta + p\xi, \pi).$$

The matrix of $T_0 L_v$ is

$$T_0 L_v = \begin{pmatrix} 1 & 0 & 0 \\ p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

That shows that the basis $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}$ of the tangent bundle $TV = TJ^1 M$ associated to the coordinates system $\{x, y, p\}$ is not left-invariant. It is the origin of non-holonomy. To get a left-invariant basis we must translate the basis $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}_0$ at 0 and this yields the basis $\left\{ \frac{\partial}{\partial x} + p \frac{\partial}{\partial y}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p} \right\}$ that is $\{t_1, -t_3, t_2\}$.

Let us now consider a vector $t$ of $C_0$. As $\eta = p\xi$ and $p = 0$, we have $\eta = 0$. Its transform $T_0 L_v(t)$ is therefore given by $(\xi, p\xi, \pi)$. As $\eta = p\xi$, $T_0 L_v(t)$ is an element of the

contact plane $C_v$ and *the contact structure* $\mathscr{C} = \{C_v\}$ *is nothing else than the left invariant field of tangent planes generated by left translating* $C_0$. Equivalently, we can say that $\mathscr{C}$ is the field of kernels of the 1-form $\omega$ which is left invariant. Indeed, at the origin 0, $\omega = \mathrm{d}y - p\,\mathrm{d}x$ is simply $\omega_0 = \mathrm{d}y$. If we translate $\omega_0$ to the point $v$ we get $\omega_v = T_0 L_v^*(\omega_0)$ defined by the formula $\omega_v(t) = \omega_0(T_0 L_v^{-1}(t))$ for $t = (\xi, \eta, \pi) \in T_v V$. But

$$T_0 L_v^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$T_0 L_v^{-1}(t) = \begin{pmatrix} 1 & 0 & 0 \\ -p & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \pi \end{pmatrix} = \begin{pmatrix} \xi \\ -p\xi + \eta \\ \pi \end{pmatrix}.$$

So

$$\begin{aligned} \omega_v(t) &= \mathrm{d}y(\xi, -p\xi + \eta, \pi) = -p\xi + \eta \\ &= \mathrm{d}y(t) - p\,\mathrm{d}x(t) = \omega(t). \end{aligned} \tag{22}$$

Let us examplify on this very simple case some general features of general Lie groups on which we will return in Section 8. The left translation $L_v$ translate the situation at 0 to a situation at $v$. We can return to 0 using the *right* translation $R_{v^{-1}}$. We get that way an *inner automorphism* of the Lie group $V = J^1 M$ (it is trivial to verify that it is effectively a morphism of the group structure)

$$\begin{aligned} A_v &: v' \mapsto v \cdot v' \cdot v^{-1}, \\ (x', y', p') &\mapsto (x', y' + px' - p'x, p'). \end{aligned} \tag{23}$$

As 0 is a fixed point of $A_v$, the tangent application $Ad_v = T_0 A_v$ of $A_v$ at 0 is an automorphism of the Lie algebra $\mathfrak{v} = T_0 V$. Its matrix (the Jacobian of $A_v$ at 0) is given by

$$Ad_v = \begin{pmatrix} 1 & 0 & 0 \\ p & 1 & -x \\ 0 & 0 & 1 \end{pmatrix} \tag{24}$$

It is trivial to verify that this map $v \mapsto Ad_v$ from $V$ to $\mathrm{Aut}(\mathfrak{v})$ is a *representation* (that is a morphism of groups). It is referred to as the *adjoint representation* of $V$. Its tangent map is a morphism of Lie algebra, referred to as $ad_t$, from the Lie algebra $\mathfrak{v}$ to the Lie algebra $\mathrm{End}(\mathfrak{v})$ of $\mathrm{Aut}(\mathfrak{v})$. If $t = (\xi, \eta, \pi) \in \mathfrak{v} = T_0 J^1 M$, the matrix of $ad_t$ is

$$ad_t = \begin{pmatrix} 0 & 0 & 0 \\ \pi & 0 & -\xi \\ 0 & 0 & 0 \end{pmatrix}. \tag{25}$$

We have therefore $ad_t(t') = (0, \xi'\pi - \xi\pi', 0) = [t, t']$ and the Lie bracket can be recovered from the adjoint representation.

The *orbits* of the adjoint representation are easy to compute. If $v = (x, y, p)$ *varies* in $V = J^1 M$, and if

$t = (\xi, \eta, \pi) \in \mathfrak{v} = T_0 V$ is *fixed*, $Ad_v(t) = (\xi, p\xi + \eta - x\pi, \pi)$ generates the line $\tilde{t} = (\xi, \mathbb{R}, \pi)$ when $\xi \neq 0$ or $\pi \neq 0$. When $\xi = \pi = 0$, $Ad_v(t) = t$ and all the elements $t = (0, \eta, 0)$ are fixed points: $\tilde{t} = \{t\}$.

It is easy to *dualize* these constructions. Let $\{\mathrm{d}x, \mathrm{d}y, \mathrm{d}p\}$ be the basis of the vector space of 1-forms on $V = J^1 M$ associated to the coordinates system $\{x, y, p\}$. At the point 0 we get a basis of the dual $\mathfrak{v}^*$ of the Lie algebra $\mathfrak{v} = T_0 M$, so if $\theta$ is a 1-form on $\mathfrak{v}$ it can be expressed as $\theta = \alpha\,\mathrm{d}x + \beta\,\mathrm{d}y + \delta\,\mathrm{d}p = (\alpha, \beta, \delta)$. If $t \in \mathfrak{v}$, it is conventional to note $\langle \theta, t \rangle$ the value $\theta(t)$ to emphasize the duality between tangent vectors and 1-forms (also called co-vectors). We define then the *co-adjoint representation* by $\langle Ad_v^*(\theta), t \rangle = \langle \theta, Ad_v(t) \rangle$. It is easy to show that this is a representation of the group $V$ on $\mathfrak{v}^*$. As

$$\begin{aligned} \langle \theta, Ad_v(t) \rangle &= \Big\langle \alpha\,\mathrm{d}x + \beta\,\mathrm{d}y + \delta\,\mathrm{d}p, \\ &\quad \Big( \xi\frac{\partial}{\partial x}, (p\xi + \eta - x\pi)\frac{\partial}{\partial y}, \pi\frac{\partial}{\partial p} \Big) \Big\rangle \\ &= \alpha\xi + \beta(p\xi + \eta - x\pi) + \delta\pi \\ &= (\alpha + \beta p)\xi + \beta\eta + (\delta - \beta x)\pi, \end{aligned} \tag{26}$$

we get $Ad_v^*(\theta) = (\alpha + \beta p, \beta, \delta - \beta x)$.

The orbits of the co-adjoint representation are the planes $(\mathbb{R}, \beta, \mathbb{R})$ if $\beta \neq 0$ (planes parallel to the plane $(\alpha, \delta)$ with ordinate $\beta$). If $\beta = 0$, all the points $(\alpha, 0, \delta)$ of the plane $(\alpha, \delta)$ are fixed points.

Taking the tangent application of the co-adjoint representation we get the adjoint $ad^*$ of the $ad$ map. It is a morphism of algebra of $\mathfrak{v}$ to $\mathrm{End}(\mathfrak{v}^*)$ defined by

$$\begin{aligned} ad_t^*(\theta)(t') &= \langle ad_t^*(\theta), t' \rangle = \langle \theta, ad_t(t') \rangle = \langle \theta, [t, t'] \rangle \\ &= \langle \alpha\,\mathrm{d}x + \beta\,\mathrm{d}y + \delta\,\mathrm{d}p, (0, \xi'\pi - \xi\pi', 0) \rangle \\ &= \beta(\xi'\pi - \xi\pi'). \end{aligned} \tag{27}$$

But as $\xi' = \mathrm{d}x(t')$ and $\pi' = \mathrm{d}p(t')$, we get $ad_t^*(\theta) = (\beta\pi, 0, -\beta\xi)$.

Let us also say a word on what is called *the Maurer–Cartan form* $\Lambda$ of a Lie group. Let $\mathrm{d}v = (\mathrm{d}x, \mathrm{d}y, \mathrm{d}p)$. It can be considered as a 1-form $\mathrm{d}v \in T^*V \otimes \mathfrak{v}$ (where $\otimes$ is the tensorial product) with values in the Lie algebra $\mathfrak{v} = T_0 V$ in the sense that, if $t_v = (\xi, \eta, \pi) \in T_v V$ is a tangent vector of $V$ at $v$, $\mathrm{d}v(v)(t_v)$ is a tangent vector of $V$ at 0. [12] But $\mathrm{d}v$ is not left-invariant. The Maurer–Cartan form $\Lambda$ consists in taking the 1-form $\mathrm{d}v(0)$ on $T_0 V$ and in translating it in order to get a left-invariant 1-form. By definition, $\Lambda(v) = (T_v L_{v^{-1}})^* \mathrm{d}v(0)$ where $T_v L_{v^{-1}} = T_0 L_v^{-1} : T_v V \to T_0 V$. Therefore we have by definition

$$\Lambda(v) : T_v V \overset{T_v L_{v^{-1}}}{\to} T_0 V \overset{\mathrm{d}v(0)}{\to} \mathbb{R}. \tag{28}$$

---

[12] $\mathrm{d}v(v)(t_v)$ is the value on $t_v$ of the 1-form $\mathrm{d}v$ taken at $v$.

But for $t = (\xi, \eta, \pi) \in T_0 V$, $dv(0)(t)$ is nothing else than $t$ itself since $dx$, $dy$, and $dp$ pick up the components of $t$. Therefore, we get $\Lambda(v)(t) = T_0 L_v^{-1}(t) = (\xi, -p\xi + \eta, \pi)$. This shows that

$$\Lambda = (dx, \omega, dp).$$

Using this expression of the Maurer–Cartan form $\Lambda$, it is easy to verify that $\Lambda$ satisfies the Maurer–Cartan equation

$$d\Lambda = -\frac{1}{2}[\Lambda, \Lambda], \tag{29}$$

which is a *universal* equation for Lie groups. Indeed $\Lambda$ can be written

$$\Lambda = dx \otimes \partial_x + \omega \otimes \partial_y + dp \otimes \partial_p \in T^* V \otimes \mathfrak{v}.$$

As $d^2 x = d^2 y = 0$ and $d\omega = dx \wedge dp$ by Eq. (18), we get $d\Lambda = (dx \wedge dp) \otimes \partial_y$. On the other hand, by definition of the exterior product of 1-forms with values in a Lie algebra, we have

$$
\begin{aligned}
[\Lambda, \Lambda] = {} & (dx \wedge dx) \otimes [\partial_x, \partial_x] + (dx \wedge \omega) \otimes [\partial_x, \partial_y] \\
& + (dx \wedge dp) \otimes [\partial_x, \partial_p] + (\omega \wedge dx) \otimes [\partial_y, \partial_x] \\
& + (\omega \wedge \omega) \otimes [\partial_y, \partial_y] + (\omega \wedge dp) \otimes [\partial_y, \partial_p] \\
& + (dp \wedge dx) \otimes [\partial_p, \partial_x] + (dp \wedge \omega) \otimes [\partial_p, \partial_y] \\
& + (dp \wedge dp) \otimes [\partial_p, \partial_p].
\end{aligned}
\tag{30}
$$

But $dx \wedge dx = \omega \wedge \omega = dp \wedge dp = 0$ and $[\partial_x, \partial_x] = [\partial_y, \partial_y] = [\partial_p, \partial_p] = 0$ for general reasons of antisymmetry; $dx \wedge \omega = dx \wedge dy$ and $dp \wedge \omega = dp \wedge dy - p \, dp \wedge dx$ by definition of $\omega$; $[\partial_x, \partial_y] = [\partial_y, \partial_p] = 0$ and $[\partial_x, \partial_p] = -\partial_y$ due to the Lie algebra structure. These equations imply immediately $[\Lambda, \Lambda] = -2d\Lambda$.

### 4.2.6. Connections, sub-Riemannian geometry, Carnot groups, and Carnot-Carathéodory metrics

A very interesting aspect of the interaction between the contact structure $\mathscr{C}$ and a Lie group structure is that one can consider $\mathscr{C}$ and the left invariant 1-form $\omega$ as defining a *connection* (in Elie Cartan's sense) on $V = J^1 M$. In general, a connection on a manifold consists in a way of comparing neighboring tangent planes via a parallel transport of tangent vectors. It can be shown that this connection is compatible with the projection $pr : V \to V/W$, where $W$ is the isotropy subgroup of $\omega_0 = dy = (\alpha = 0, \beta = 1, \delta = 0)$ for the coadjoint representation. [13] By definition, $W$ is the set $W = \{v \in V$ s.t. $Ad_v^*(\omega_0) = \omega_0\}$. It is a 1-dimensional Lie group and its Lie algebra $\mathfrak{w}$ is the set $\mathfrak{w} = \{t \in \mathfrak{v}$ s.t. $ad_t^*(\omega_0) = 0\}$. As $\omega_0 = dy = (\alpha = 0, \beta = 1, \delta = 0)$, for $v = (x, y, p)$, we have $Ad_v^*(\omega_0) = (\alpha + \beta p, \beta, \delta - \beta x) = (p, 1, -x)$. To satisfy the identity $Ad_v^*(\omega_0) = \omega_0 = (0, 1, 0)$ we must have $x = 0$ and $p = 0$. Therefore $W$ is

the $y$ axis, the group law restricted to it being simply the addition $y + y'$. We verify that, as $ad_t^*(\omega_0) = (\pi, 0, -\xi)$ since $\beta = 1$, the identity $ad_t^*(\omega_0) = (\pi, 0, -\xi) = 0$ leads to $\xi = 0$ and $\pi = 0$ which is also the $y$ axis, but considered as the Lie algebra $\mathfrak{w}$ of $W$.

This representation via a connection is in some sense *dual* to the classical one. In the classical case, the base plane is the plane $(x, y)$ and the fiber is the axis of tangents $p$. Curves $\gamma$ are given as function $y = f(x)$ and the tangent $p$ is computed via derivation. In the alternative perspective, the base plane is the plane $(x, p)$ and the fiber is the axis of $y$ values. Curves $\gamma$ are given as functions $p = g(x)$, that is as *envelopes* of tangents, and $y$ is computed via integration since the lifting of $\gamma$ in $V$ is given by $y = \int y' \, dx = \int p \, dx$.

It can be shown that the *curvature* $d\omega$ of the connection $\omega$ is a symplectic form on $V/W$. In our case this is evident since $d\omega = dx \wedge dp$ is the standard symplectic form on the base plane $\{x, p\} = V/W$.

A connection $\mathscr{C}$ on a manifold $V$ allows to redefine differential calculus using the parallel transport defined by the field of hyperplanes $\mathscr{C}$ which are the kernels of the connection 1-form $\omega$. For instance, for the exterior derivative of differential forms, the key idea is to define the new derivative, called the *covariant derivative*, as the exterior derivative restricted to the $\mathscr{C}$-components of the tangent vectors. Namely, if $\theta(t_1, \ldots, t_k)$ is a $k$-form, its covariant derivative $D\theta$ will be given by $D\theta(t_1, \ldots, t_{k+1}) = d\theta(\bar{t}_1, \ldots, \bar{t}_{k+1})$, where $\bar{t}$ is the projection of $t \in T_v V$ onto the plane $C_v$. It is the same thing for other geometrical entities. For instance if $f : V \to \mathbb{R}$ is a real function on $V$, its *gradient* relative to $\mathscr{C}$ will be given by

$$\nabla_{\mathscr{C}}(f) = t_1(f)t_1 + t_2(f)t_2, \tag{31}$$

where $t_1$ and $t_2$ constitute the left-invariant basis of $\mathscr{C}$. By construction, $\nabla_{\mathscr{C}}$ is tangent to the contact structure $\mathscr{C}$ and defines therefore a vector field whose trajectories are all integral curves of $\mathscr{C}$. In the same vein, if $X = \varphi t_1 + \psi t_2$ is a contact vector field on $V$, its *divergence* relative to $\mathscr{C}$ will be given by

$$\mathrm{div}_{\mathscr{C}} X = t_1(\varphi) + t_2(\psi). \tag{32}$$

In fact, a very strong general theory can be developed from the fact that we have a Lie group $V$ with a Lie algebra $\mathfrak{v}$ with only one non-vanishing commutator $[t_1, t_2] = t_3$ (in our case, $t_1 = \frac{\partial}{\partial x} + p \frac{\partial}{\partial y}$, $t_2 = \frac{\partial}{\partial p}$, $t_3 = -\frac{\partial}{\partial y}$) and a left invariant 1-form $\omega$ which is a contact form ($\omega \wedge d\omega \neq 0$). $t_3 = \chi$ is called the *characteristic field* of the field $\mathscr{C}$ of contact planes $\{t_1, t_2\}$.

These data allow to define special metrics on $V$ called by Gromov [24], Lafontaine, and Pansu [54] *Carnot-Carathéodory metrics*. [14] The idea is to consider a metric

---

[13] $\omega_0 = dy = (\alpha = 0, \beta = 1, \delta = 0)$ is the value of the 1-form $\omega = dy - p \, dx = (\alpha = -p, \beta = 1, \delta = 0)$ at the point $0 = (x = 0, y = 0, p = 0)$.

[14] This name come from the theory of adiabatic processes in thermodynamics.

$g_{\mathscr{C}}$ defined only on the planes of $\mathscr{C}$, and not on the complete tangent bundle $TV$, for instance the metric making $t_1$ and $t_2$ a left-invariant orthonormal basis, and to *restrict* the consideration of curves $\Gamma$ in $V$ to curves which are *tangent* to $\mathscr{C}$, that is to *integral curves* of $\mathscr{C}$. We distinguish and select therefore a class of special curves. Let $v$ and $v'$ be 2 points of $V$. To take into account the integrability constraint, it is natural to define their distance $d_{\mathscr{C}}(v, v')$ as the *inf* of the $g_{\mathscr{C}}$-length of the integral curves joining $v$ to $v'$ (it can be shown that such curves always exist due to the fact that the Lie brackets of $\mathscr{C}$ generate all the tangent bundle $TV$). More precisely, we define $d_{\mathscr{C}}$ by the formula

$$d_{\mathscr{C}}(v, v') = \inf \int_I \|\Gamma'(s)\| \, ds \quad \text{for} \quad \begin{array}{l} \Gamma \text{ integral curve of } \mathscr{C}, \\ \Gamma(0) = v, \Gamma(1) = v' \end{array}$$

$$(33)$$

A *geodesic* between $v$ and $v'$ for the Carnot-Carathéodory metric is then an integral curve of $\mathscr{C}$ which realizes the distance $d_{\mathscr{C}}(v, v')$.

The Carnot-Carathéodory metric $g_{\mathscr{C}}$ is a "path-metric" called *sub-Riemannian* for it is defined only on the *sub-bundle* $\mathscr{C}$ of the tangent bundle $TV$ [2,24]. Of course, it can be extended to a Riemannian metric $g$ by considering the characteristic field of $\mathscr{C}$ as unitary and orthogonal to $\mathscr{C}$, the contact structure becoming what is called a *polarization* of $g$. But it is nevertheless completely different from $g$. It can be considered as a limit of Riemannian geometries on $V$ which penalize more and more the defect of integrability of the curves, that is the defect of tangency to $\mathscr{C}$, the non-integral curves becoming prohibited at the limit. It is highly non-isotropic and non-homogeneous, *singular* and even *fractal*. Indeed, it can be shown that, although $V$ is topologically of dimension 3, its Hausdorff dimension relative to $g_{\mathscr{C}}$ is 4.

For our purpose here, it is extremely important to note that the sub-Riemannian metric $g_{\mathscr{C}}$ is a *limit* of more and more anisotropic Riemannian metrics $g_\varepsilon$ defined by introducing a difference of scale between the directions respectively tangent and orthogonal to the $C_v$ planes. Indeed, we can easily implement this difference if the connections connecting neurons $(a, p)$ corresponding to the $C_v$ planes are "strong", while those connecting neurons $(a, p)$ corresponding to the characteristic direction $\chi_v$ are "weak".

### 4.2.7. The problem of "regular" surfaces in V

If $f : V \to \mathbb{R}$ is a regular function, $f = 0$ defines a surface $S$ of $V$. Its unitary normal vector relative to $\mathscr{C}$ at the point $v \in S$ is $n_{\mathscr{C}}(v) = \frac{\nabla_{\mathscr{C}}(f)(v)}{\|\nabla_{\mathscr{C}}(f)(v)\|}$. $n_{\mathscr{C}}$ is the normalized projection on $C_v$ of the classical normal unitary vector $n(v) = \frac{\nabla(f)(v)}{\|\nabla(f)(v)\|}$. It is not defined if $\nabla_{\mathscr{C}}(f)(v) = 0$, that is if $T_v S = C_v$ at $v$. Such points are called *characteristic points*

of $S$. They will exist in general even if $S$ is without any singular points in the classical sense, that is if $\nabla(f)(v) \neq 0$ everywhere. But as the contact structure $\mathscr{C}$ is completely non-integrable they will necessarily be rather exceptional.

### 4.3. Contact structures and vision

Up to now, almost no specialist of vision has really investigated the striking relations existing between the orientation hypercolumns of the primary visual cortex and geometrical notions such as fibrations, jet spaces, or contact structures. But, as far as I know, there are two distinguished exceptions: Jan Koenderink (see Section 3.2.3) and William Hoffman.

William Hoffman has been a pioneer of the applications of differential geometry and of Lie group theory to vision. In his seminal paper "The visual cortex is a contact bundle" [30] he introduced the idea that contours lift discontinuities of the retinal stimulus:

> A path on one manifold [the retina] is "lifted" via a fibering to another manifold [the cortex] in a coherent fashion. (p. 145)

In another paper [29] he claimed

> Fibrations (...) are certainly present and operative in the posterior perceptual system if one takes account of the presence of 'orientation' micro-response fields and the columnar arrangement of cortex. (p. 645)

He developed also the idea that the retinal RFs provide local charts, the transition maps between overlapping charts being implemented by the fine connectivity of the retina. He considered also that the RFs of the simple orientation cells of V1 implement a connection on the fiber bundle.

### 4.4. Pinwheels and ocular dominance

The relation between pinwheels and domains of ocular dominance (DODs) is very interesting (see e.g. Hübener [33]). Many iso-orientation lines cross the boundaries of DODs close to right angles and, as was shown by Michael Crair [10], the peaks of ocular dominance are localized near the centers of the pinwheels in the middle of these DODs (Figs. 33 and 34).

Hübener's results show that there exist *quasi-quadruple* points on the boundaries of the DODs. This morphological cue prompts us to think that if we idealize the structure using, as in Section 3.3, a regular lattice of pinwheels, we will get the geometry of Fig. 35.

It is interesting to note that this geometrical configuration of pinwheels and DODs is as if there existed a
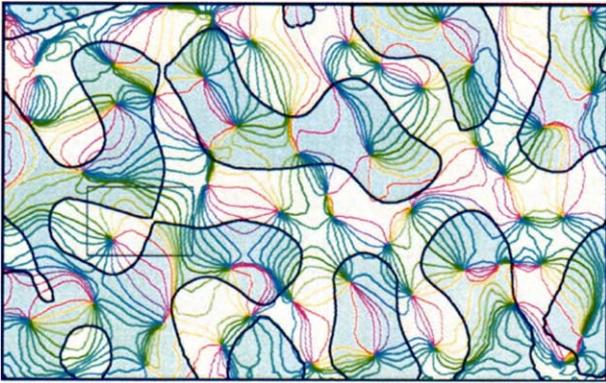
Fig. 33. Relation between the pinwheels and the domains of ocular dominance (DODs). Many iso-orientation lines cross the boundaries of DODs close to right angles (from [33]).
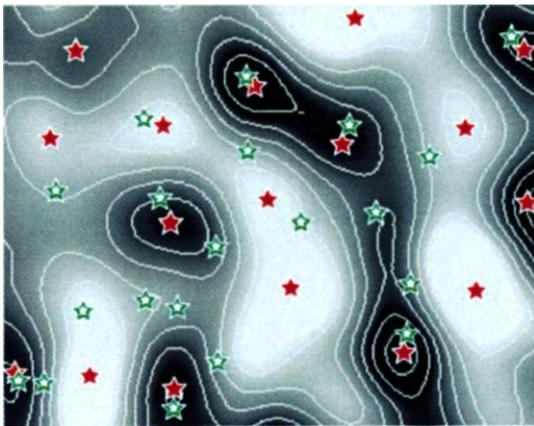


Fig. 34. The peaks of the domains of ocular dominance coincide roughly with the centers of the pinwheels (from [10]).
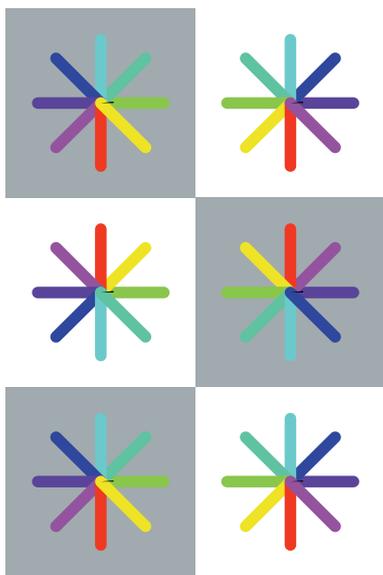


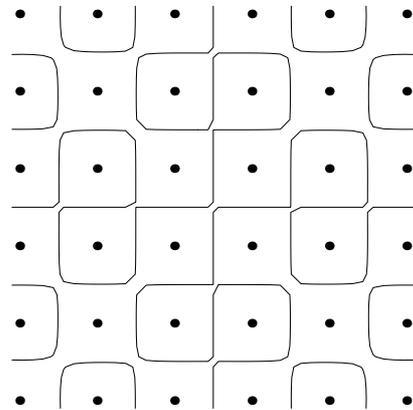Fig. 35. The domains of ocular dominance in the idealized geometrical "crystal" model of Fig. 15.



Fig. 36. The idealized geometrical lattice model of Fig. 35 interpreted as a physical model where the pinwheel singularities generate a scalar field whose gradient lines give the iso-orientation lines and whose level lines give the boundaries of DODs.
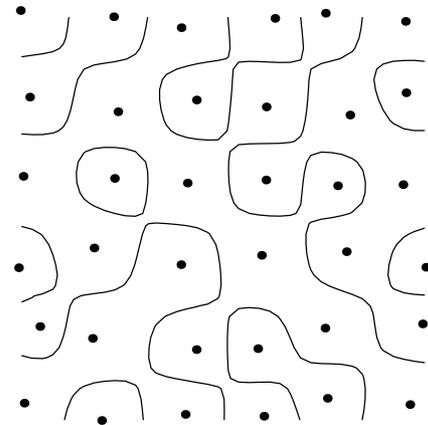


Fig. 37. A deformation of the regular physical model of Fig. 36.

sort of *scalar field* underlying the functional architecture of V1: the centers of the pinwheels of respectively left and right chirality would become respectively + and − topological charges, the level lines would become the boundaries of the DODs and the gradient lines would become the iso-orientation lines, these two families of lines being orthogonal. In Figs. 36 and 37 we show how the idealized geometrical model interpreted as a physical model, deforms into a Hübener type configuration.

## 4.5. The dynamics of pinwheels and the development of functional architecture

Up to now, we have treated the pinwheel neurogeometry as a given fixed architecture and we proposed for it a geometrical and a physical model. But it is of course the result of a *learning process* (see e.g. [66]). We have

seen in Section 3.3.2 that we can consider the field of orientations $\vartheta(a) = e^{i\theta(a)}$ of the simple cells of V1 as a section of the fibration $\pi : M \times P \to P$ defined outside the lattice $L$ of pinwheels' centers. In the same perspective, Fred Wolf and Theo Geisel [78] have worked out a beautiful learning theory of orientation selectivity, starting from an initial state unselective to orientation. They model the columns by a *complex field* $z(a) = \rho(a)e^{i\theta(a)}$ where the spatial phase $\theta(a)$ codes the orientation preference and where the module $\rho(a) = |z(a)|$ codes the selectivity to orientation. $z$ is therefore a section of the fibration of base $M$ and fiber $\mathbb{C}$ and the singular points correspond to the *zeroes* of $z$.

The authors introduce also two other fields

1. $d(a)$ for the ocular dominance: $d(a) > 0$ (respectively, <0 ) if the left (respectively, right) eye is dominant;
2. $r(a)$ for the position of the receptive field on the retina (distorsion of the retinotopy, see Figs. 7 and 8 in Section 3.1).

They show that after a period of proliferation of pinwheels, the number of pinwheels decreases via movements, collisions and annihilations of pairs of pinwheels of opposed chirality, and that DODs slow down and stabilize the process (Fig. 38).

This dynamics of pinwheels is consistant with the physical model of the previous Section 4.4, + and − charges annihilating each other.

## 5. Application 1: the association field of Field, Hayes and Hess

### 5.1. Field's, Hayes' and Hess' experiments

The Frobenius integrability condition in Section 4.2.3 is an idealized mathematical version of the Gestalt principle of good continuation. Its psychophysical empirical counterpart has been studied in great detail by David Field, Anthony Hayes and Robert Hess in their celebrated 1993 paper "Contour Integration by the Human Visual System" [16], where they introduced the key concept of *association field*.

Their experimental protocol consists in briefly presenting (typically during 1s) a grid made up of 256 oriented elements $(a, p)$ to subjects. They use Gabor patches because these spatially oriented bandpass filters select a single spatial frequency (their Fourier transform is a narrow peak). The line segments used by other experimentalists (e.g. [35]) have the drawback of exciting cells responding to different spatial frequencies. In half cases, the grid contains 12 elements whose centers are aligned along a smooth line $\gamma$, the others being randomly oriented (Fig. 39). In the other cases, all the elements are randomly oriented. The task (method of the forced

choice between two alternatives) consists in determining whether there exists or not an alignment in the presented grid. [15] The results show that the subjects perceive quite easily the alignment if the elements $(a, p)$ are aligned "*tangentially*" to $\gamma$ and if the slope variation between two consecutive elements is not too large ($\leqslant 30°$). This phenomenon of perceptual saliency, or *pop-out*, is characteristic.

The key point is that the elements of the grid are too distant from each other to belong to a single RF:

It is clear that this 'association field' covers a considerably wider area than would be covered by the receptive field of a mammalian cortical cell. [16, p. 185]

But despite their distance, subjects group spontaneously the aligned elements. An automatic mechanism connecting several RFs must therefore operate as a *low level integration*. It is why the authors looked for a local to global integration process:

A useful segregation process for real scenes may be based on local (rather than global) integration. (...) The general theme of these algorithms is that the points along the length of a curved edge can be linked together according to a set of local rules that allow the edge to be seen as a whole, even though different components of the edge are detected by independent mechanisms. [16, p. 174]

It is really the core of the problem. Grouping is global in a sense completely different from that associated to large RFs:

In our stimuli, there does not exist any 'global' feature that allows the path to be segregated from the background. It is not possible to segregate the path by filtering along any particular dimension. Our results imply that the path segregation is based on local processes which group features locally. [16, p. 191]

But if feature detectors remain point processors how can their measures be globalized?
Experiments provide two other striking results:

- if the slope variation between two consecutive elements is too large, no alignment is perceived (Fig. 40);
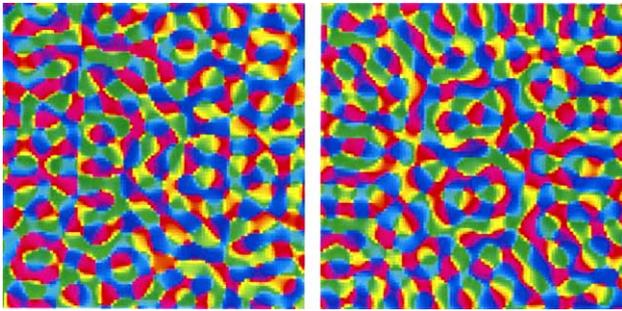
---

Fig. 20. Two examples of a superposition of orientation "waves" yielding a pinwheel structure.
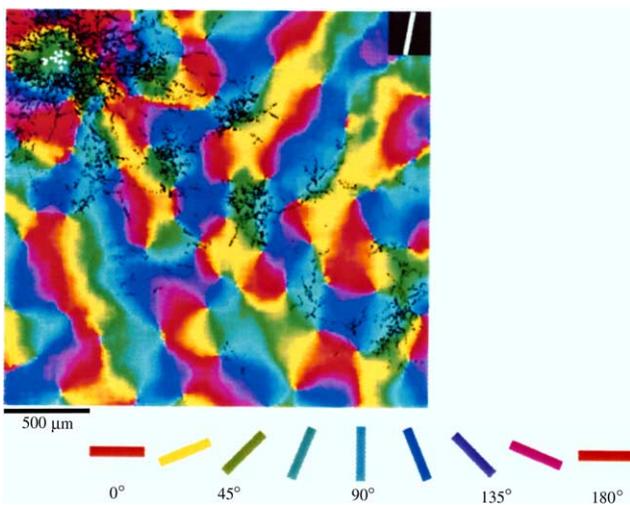


Fig. 23. Connections between cortico-cortical horizontal cells marked by biocytin's diffusion. The short range diffusion is isotropic, while the long range diffusion is anisotropic and restricted to iso-orientation domains (from [5]).

- it is the same if the orientations $p$ of the elements are no longer tangent but transverse (e.g. orthogonal) to the curve $\gamma$ described by their centres (Fig. 41).

In these two cases the pop-out vanishes and the paths are no longer perceptively but only cognitively and inferentially detected.

## 5.2. Explanation through the concept of association field

According to Field, Hayes and Hess, the tendency of elements $c_i = (a_i, p_i)$ to be perceived as aligned is due to the existence, around each element, of a region in which other elements tend to be perceived as grouped. This region, called the *association field*, is defined by *joint conditions of position and orientation*. Its structure is described in Figs. 42 and 43.

The authors have described very acutely the geometrical nature of the association field. First, association is not simply

a general spread of activation, linking together all types of features within the field. [16, p. 185]

It manifests a correlation between position and orientation:

Elements are associated according to joint constraints of position and orientation. [16, p. 187]

Whence the key conclusion:

There is a *unique link* between the relative positions of the elements and their relative orientations. (...) *The orientation of the elements is locked to the orientation of the path*; a *smooth curve* passing through the long axis can be drawn between any two successive elements. [16, p. 181]

This affirmation is *a discrete formulation of the integrability condition*. The contact elements $c_i = (a_i, p_i)$ embedded in a background of distractors generate a perceptively salient curve (pop-out) iff the orientations $p_i$ are tangent to a curve $\gamma$ interpolating between the positions $a_i$. This is due to the fact that the co-activation of approximatively co-axial simple orientation cells (detecting pairs $(a, p)$ and $(b, q)$ with $b$ roughly aligned with $a$ in the direction $p$ and $q$ close to $p$), co-activates, via the horizontal cortico-cortical connections, intermediary cells (Fig. 44).

When the distance between the positions $a_i$ tends to 0, the "joint constraints" and the "unique link" between
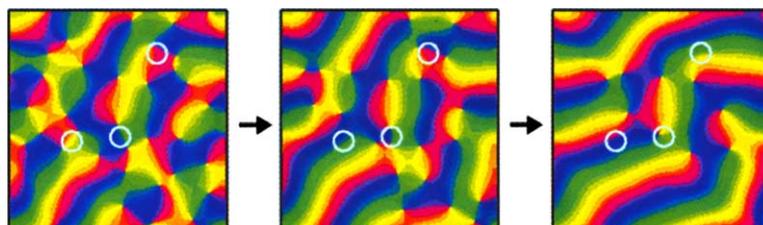


Fig. 38. Annihilations of pinwheels of opposed chirality simplify pinwheel geometry (from [78]).
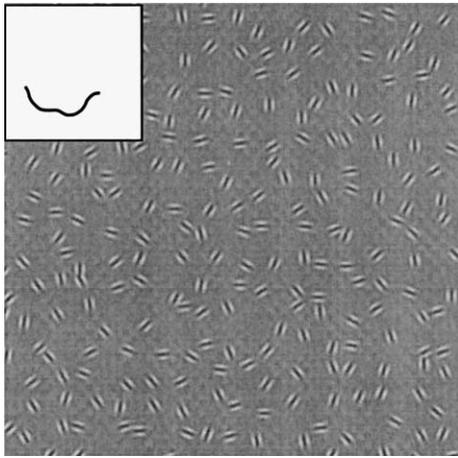
Fig. 39. The path on the upper left corner is embedded in a background of randomly oriented distractors. In such an array of Gabor patches $(a, p)$, the subject can observe the pop-out of the components $(a_i, p_i)$ of the path if they are suitably aligned (from [16]).
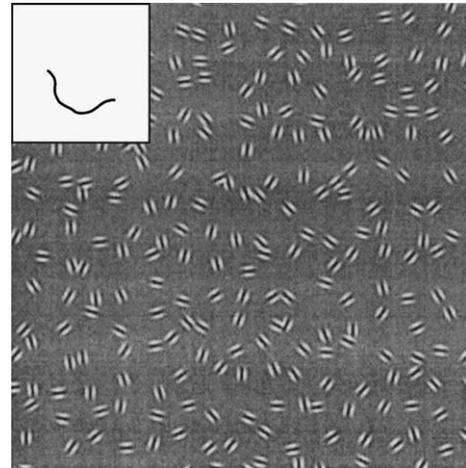


Fig. 41. If the orientations $p_i$ of the elements $(a_i, p_i)$ of the path are not tangent but transverse (e.g. orthogonal) to the curve $\gamma$ interpolating between their centers $a_i$, the subjects do not perceive the alignment anymore.
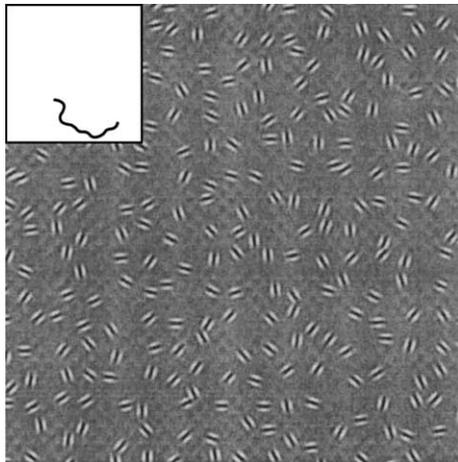


Fig. 40. If the variation of slope between consecutive elements $(a_i, p_i)$ of the path is too large, then the subjects do not perceive the alignment anymore.



Fig. 42. Schema of the association field. The elements are pairs $(a, p) = $ (position, orientation). Two elements $(a_1, p_1)$ and $(a_2, p_2)$ are connected (thick strokes) if one can interpolate between positions $a_1$ and $a_2$ a curve $\gamma$ tangent to $p_1$ at $a_1$ and to $p_2$ at $a_2$. If it is not the case, the two elements are not connected (thin strokes).



Fig. 43. The field lines of the association field.

positions $a$ and orientations $p$ become exactly the Frobenius integrability constraint $p = f'(x)$ of Section 4.2.3.

### 5.3. The association field as a discrete version of the contact structure

Yannick Tondut [68] has shown that the association field can effectively be interpreted as a *discretized* version of the contact structure $\mathscr{C}$ of the contact bundle $V$. As the contact structure is an idealization of the pinwheel structure corresponding to a 0 scale, it is natural to discretize it.

We choose then a particular *scale*, a step of discretization approximating d$x$ (infinitesimal variation) by $\Delta x$ (finite variation). But what scale? The visual system processes the retinal signal at several scales simulta-
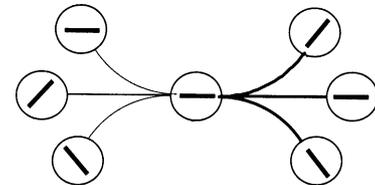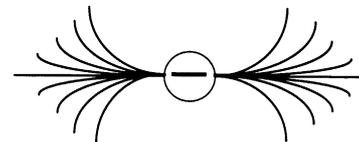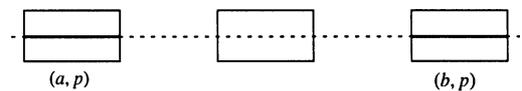


Fig. 44. The co-activation of two co-axial cells $(a, p)$ and $(b, q)$ (with $p = q = ab$ by definition of co-axiality), co-activates intermediary co-axial cells.

neously. Now, the local information (for example the direction of the tangent to a curve) can be completely different from scale to scale (e.g. in the case of a fractal boundary). Field and his co-workers made an interesting hypothesis: the visual system could solve the continuity problem separately at each scale. The association field will then correspond to the solution of the problem *at a given scale*.
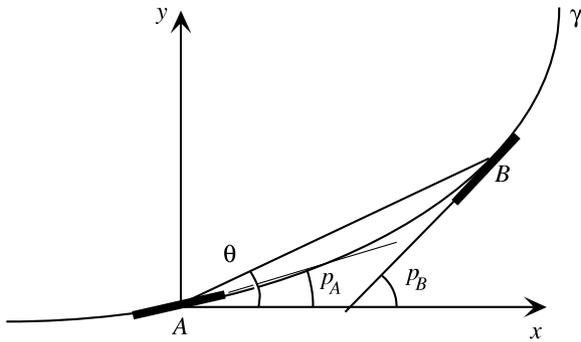
Fig. 45. Discretization of the Frobenius integrability condition.

In our model, this hypothesis amounts to fix the step of discretization. We will refer to it as $\Delta s$. We also discretize the orientation, using a step $\Delta p$. We thus consider the contact bundle with local coordinates $(x, y, p)$, and a curve $\Gamma$ in it, locally defined as before by equations $y = f(x)$, $p = g(x)$. As we have seen in Section 4.2.3, $\Gamma$ lifts the base curve $\gamma$ of equation $y = f(x)$ iff the Frobenius condition $p = g(x) = f'(x) = \frac{dy}{dx}$ is satisfied.

In order to discretize this equation, let us consider two consecutive points $A$ and $B$ on $\gamma$. [16] They satisfy the metrical relation $d(A, B) = \Delta s$ ($d$ being the Euclidean distance on the visual field [17]). We will refer by $\theta$ to the angle of the vector $AB$ with the $x$-axis, so that (Fig. 45)

$$x_B - x_A = \Delta s \cdot \cos \theta \quad \text{and} \quad y_B - y_A = \Delta s \cdot \sin \theta. \quad (34)$$

In this discrete framework, the tangent to $\gamma$ at $A$ can be approximated by the line $AB$, which amounts to approximate the derivative $f'(x)$ by the ratio $\frac{y_B - y_A}{x_B - x_A} = \tan \theta$. The equation corresponding to the contact structure at $A$ takes then the following discretized form:

$$|p_A - \tan \theta| \leqslant \Delta p.$$

Symmetrically, we have also: $|p_B - \tan \theta| \leqslant \Delta p$. By adding these two inequations, we get the symmetrical form:

$$|p_A - \tan \theta| + |p_B - \tan \theta| \leqslant 2\Delta p. \quad (35)$$

If we graphically represent the oriented elements which are admissible, in the sense of inequality (35), as consecutive to a given one, we get the expected result: they are laid out according to the structure of Field, Hayes and Hess association field (Fig. 45). The value of the threshold $2\Delta p$ is about $\tan(\pi/6)$. Moreover, inequality (35) accounts for the tolerance experimentally observed as the orientation of the bar is slightly deviated from that of the field. We conclude therefore that the association field is a discrete version of the Frobenius integrability condition and results from the fact that the pinwheel structure discretizes the contact structure.

Let us remark that an important effect of discretizing the contact structure is to induce *a limitation of the curvature* that the visual system may admit at a given scale.

We can also give a probabilistic interpretation to the association field. Association is more or less probable depending on the relative positions and orientations of the consecutive elements. Instead of being compared to a threshold, the function $|p_A - \tan \theta| + |p_B - \tan \theta|$ can be seen as a measure of the probability of association between two elements (psychophysical aspect), or as the activity of one of the elements induced by the other (physiological aspect).

It is very easy to explain the third experiment in terms of the contact structure (Fig. 41): the curve $\Gamma$ in $V$ is *non-integrable*. Fig. 46(a) shows a curve in $J^1M$ which is a Legendrian lift. Fig. 46(b) corresponds to the case of the third experiment where one adds to $p = f'(x)$ a constant $p_0$ (here a $\pi/2$ angle, the orientations $p$ are therefore perpendicular to the curve $\gamma$ described by the positions $a$). Fig. 46(c) and (d) show two other examples of non-integrable curves in $V$, the first because $p$ is constant while $f'$ is not, the second because $p$ varies much faster than $f'$. In each of these three cases the pop-out of the curve $\gamma$ described by the positions $a$ is impossible because the integrability condition (the "joint constraints of position and orientation") is not satisfied.
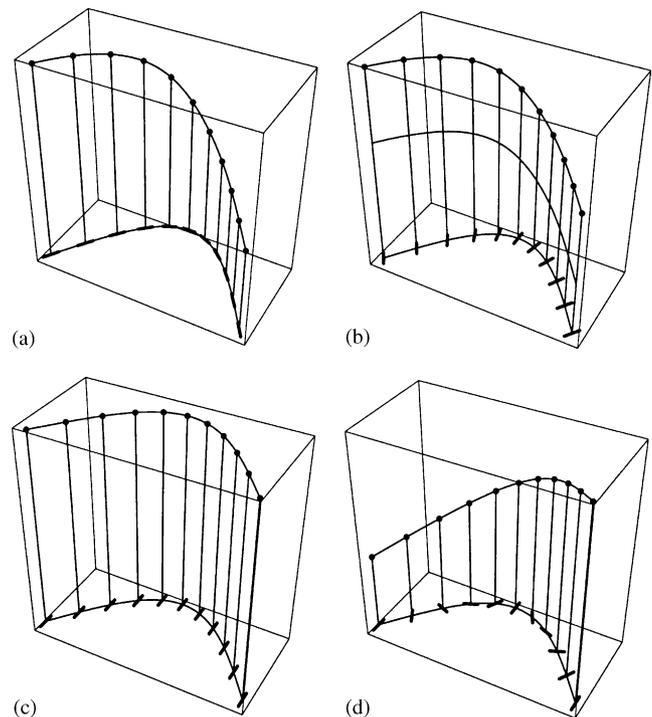


Fig. 46. The association field as a condition of integrability. (a) The integrability condition is satisfied and (b)–(d) the condition is not satisfied. (b) We add a constant angle to the tangent (i.e. $p = f'(x) + p_0$). (c) $p$ is constant while $f'$ is not. (d) $p$ rotates faster than $f'$.

---

[16] In the following, caps $A$ and $B$ refer to end points of curves.
[17] The system accesses this distance through the covariance of neural firings.

### 5.4. The link with horizontal connections in Lorenceau's experiments

In their paper, Field, Hayes and Hess present some "physiological speculations" concerning the implementation of the association field via horizontal connections. They have been confirmed by Jean Lorenceau and colleagues [21] using the method of apparent speed of fast sequences (speed = 64°/s) of oriented Gabor patches. The apparent velocity is greater (overestimated) when the successive elements are aligned in the direction of the motion path (collinear sequences), and smaller (underestimated) when the motion is orthogonal to the orientation of the elements. Moreover, in the first case, the increase of the apparent velocity measured with *psychophysical* methods is approximatively the same as the velocity of propagation of horizontal activation in the cortico-cortical connections measured with *electrophysiological* methods (about 0.2 m/s) [19].

### 5.5. Binding

The pop-out of the curve generated by the contact elements $c_i = (a_i, p_i)$ is a typical Gestalt phenomenon which results from a *binding* induced by their co-activation. Neurons detecting approximatively aligned contact elements *synchronize* their firings through the horizontal connections. The temporal coherence of the correlated firings results in the binding of the features they code, and this explains why elements aligned along a smooth contour would be perceived as a whole (for a critique of binding see [71]). In other words, the integrability condition is the geometrical condition of a binding. As was emphasized by Lee [43], we could think that a V2-feedback is necessary for binding to be efficient. But according to the high-resolution buffer hypothesis, it is the underlying geometry of V1 which is essential.

### 5.6. Comparison with other data

As was stressed by Tondut, this model of the association field also accounts for other results obtained by Polat, Sagi, and Westheimer which show that the facilitation zone induced by an oriented element $(a, p)$ implements *the contact plane* at the corresponding point of the contact bundle.

Polat and Sagi [60] considered phenomena of facilitation and suppression in target detection tasks. Using also Gabor patches, they studied the detection of a low-contrast stimulus (the target) when surrounded by two high-contrast stimuli (masks), aligned with it. Varying the distance between the target and the masks, they observed a facilitation effect for distances between $2\lambda$ and $10\lambda$ where $\lambda$ is the wavelength of the stimuli,

therefore far outside the receptive field. [18] They also compared different configurations of relative positions and orientations of the stimuli: facilitation mainly occurs when the orientation of the three stimuli coincides with that of their alignment, in agreement with the results obtained by Field and his co-workers.

The experiments carried out by Westheimer co-workers [35,36] have confirmed these results. Using simple oriented bars as stimuli, the authors observed facilitated detection of a low-contrast target when aligned with a similar high-contrast bar. Two results are especially interesting:

1. when the two bars are parallel, a very small deviation from collinearity rapidly decreases the facilitation effect, until it reverses it;
2. the facilitation effect persists when the orientation of the inducing bar is modified while preserving the continuity of the path constitued by the two bars; but it decreases with the discrepancy between the orientations of the two bars (or, rather, with the curvature of the path), and reduces to zero beyond 30°.

Another interesting comparison could be drawn with Steven Zucker's works [55,81]. In order to detect curves, Zucker uses a coarse estimate of their geometrical coherence based on the compatibility of neighboring tangents. This compatibility is estimated either directly from the difference of their orientations or through a systematical measurement of the local curvature by means of specialized detectors. Zucker's model is rather rich, and goes beyond the association field. It is nevertheless possible to compare both models in what concerns the compatibility of line elements, that is their probability to be tangent to a same curve, and therefore to be "associated". Indeed, Zucker's compatibility criterion is very similar to the association field.

It would also be interesting to study the relationships between the association field and the cooperation/competition processes used in Grossberg and Mingolla's celebrated models [25]. These authors introduce a cooperation between aligned elements, materialized by oriented "dipole cells" with large receptive fields taking into account approximate alignments. The weight of an element in the receptive field of a dipole cell depends on its relative position and orientation. The set of weights defines *a cooperation field*, whose similarity with Field, Hayes and Hess association field is striking.

Another application of the contact structure of V1 concerns the classical problem of Kanizsa *subjective contours* [34] which are typical examples of filling-in based on the Gestalt law of "good continuation". The local mechanisms of association have indeed deep con-

---

[18] The typical size of a receptive field tuned to a wavelength $\lambda$ is $2\lambda$.

sequences upon the *global* shape of contours. Principles of interpolation can be conceived of by considering curves *having a natural geometrical significance in the bundle of 1-jets*. We will tackle this point in the following section.

## 6. Application 2: Kanizsa subjective modal contours

The objects we investigate now are not classical straight Kanizsa contours but *curved* ones, when the sides of the internal angles of the pacmen are not aligned (Fig. 47).

### 6.1. Some experimental data

The influence of the various parameters controlling the formation of illusory contours began to be investigated in the late 1980s, but the results were rather scarse as is shown by the surveys of Peterhans and von der Heydt [57] or Spillman and Dresp [64].

There are different methods for measuring the extremum of a Kanizsa contour (*K*-contour). For instance we can use the Dresp-Bonnet "sub-threshold summation" method: the threshold for the detection of a small subliminar segment parallel to the *K*-contour decreases suddenly when the segment is located exactly on the *K*-contour [14]. These summation effects, as well as other results, prove that real and illusory modal contours share common neural mechanisms and are in part functionally equivalent.

In what concerns the neural underlying processes, one must distinguish between V1 and V2. For V2, Rüdiger von der Heydt and Esther Peterhans [72,73] have shown that a neuron of V2 can respond to a virtual contour created by a series of end-points aligned with its preferred orientation (the lines are therefore orthogonal to its preferred orientation). For V1 it is not the case (Figs. 48 and 49).

But this does not entail that V1 is not active in the construction of illusory contours. If one compares systematically the responses in V1 and V2 using optical techniques (see [62] for the cat), one observes reponses
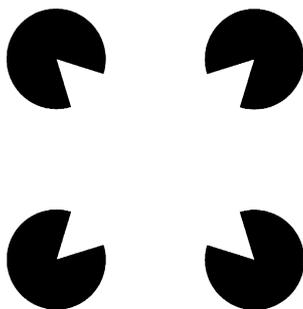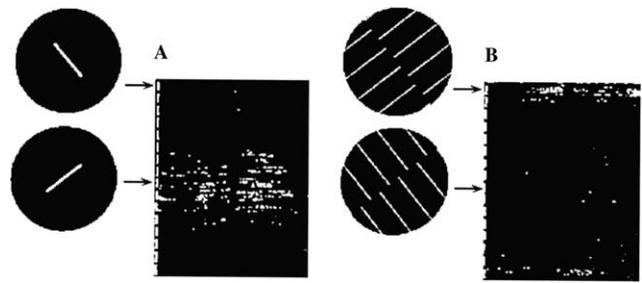


Fig. 48. In the following figures, the 2 discs represent the stimulus (bar or grating) with orientation 0° and 90°. The lines displayed in the rectangles record the responses of the neuron (time = *X*-axis) to the stimulus rotated from 0° to 180° (*Y*-axis). V1 neurons respond only to the real orientation of the bars and not to the illusory contour generated by their ends (from [72]).
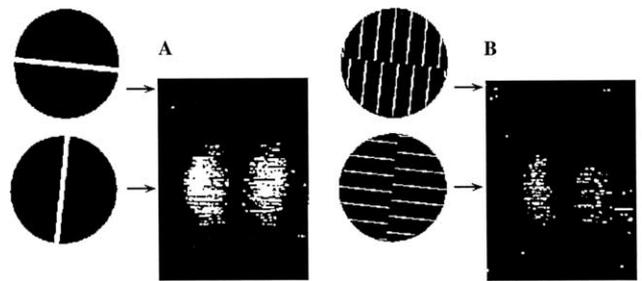


Fig. 49. A V2 neuron can respond to the illusory contour generated by the ends of a set of bars if the illusory orientation is its preferred one (*B*). Its response is as good as for an isolated real bar (*A*). This shows that in case (*B*) it does not respond to the real orientations even if they correspond to its preferred one (from [72]).

to illusory contours in the two areas. One of the difference between V1 and V2 is that the response of V1 to illusory contours is masked by the response to real ones while that of V2 is more salient. One can make the hypothesis of a progressive construction of illusory contours, starting at the V1 level and becoming richer and richer in the further areas with top down feedbacks. According to the high-resolution buffer hypothesis, V1 is strongly implied in illusory contours.

### 6.2. Classical models for subjective contours [19]

We will use the contact structure of V1 for working out a model of *K*-contours. But let us first recall some previous models inspired by the classical models of Grossberg and Mingolla [25], and von der Heydt and Peterhans [72].

In the first model, an illusory contour is generated by the interaction between approximatively collinear boundaries and is oriented along their common



Fig. 47. An example of a Kanizsa curved illusory modal contour.

---

[19] This survey section is due to Yannick Tondut.

direction. With regard to them, Heitger [27] speaks of "paragrouping". For explaining the illusory contours constituted by the end points of parallel segments, Grossberg and Mingolla suppose that the end points induce a piece of boundary orthogonal to the segments, these induced elements being able to cooperate.

For von der Heydt and Peterhans, and for those inspired by their model [37], the illusory contours result from the interaction between "end-stopped" cells (see [27]), and are formed orthogonally to the orientation of boundaries. With regard to them, Heitger [27] speaks of "orthogrouping".

All these models were conceived of to explain straight contours. To adapt them to the case of curved contours is rather problematic since it would imply a combinatorial explosion of the detectors.

Another interesting model is that of Heitger and von der Heydt [28] which uses at the same time "para grouping" and "ortho grouping". It works also for curved contours. Their principle is to detect singularities. They are constituted by oriented cells endowed with two large lobes. When each lobe contains a singularity, the cell sends a local signal for the construction of an illusory contour and generates a small oriented segment. This signal adds to the outputs of the detectors of real contours at the same location. The contour is then determined by the local maxima of activation. This "grouping field" is quite similar to the association field.

All these models can be easily interpreted geometrically in terms of fibrations if one considers fibers whose elements are no longer simple orientations $p$ but *pairs* of orthogonal orientations $(p, p^\perp)$, that is *frames* in the base space $M$. The projective fiber bundle $\pi : M \times P \to M$ is then substituted for by the bundle of frames $\rho : M \times F \to M$. We will return to this point later in Section 8.1.

### 6.3. An experiment on K-curves

In collaboration with our colleague Jacques Ninio (a specialist of vision, see [52]), we carried out an experiment on curved *K*-contours. Our purpose was to measure the precise position of the extremal point of the contour and to compare it with the prediction of simple models. We used families of *K*-curves with

two configurations: triangle, square;
two sizes of configurations;
two sizes of pacmen;
four orientations;
five angles (Figs. 50 and 51).

As for method of detection we asked the subjects to place correctly a marker (the extremity of an orthogonal line, a small segment, the axis of a small stripe) at the extremum of the contour (Fig. 52).
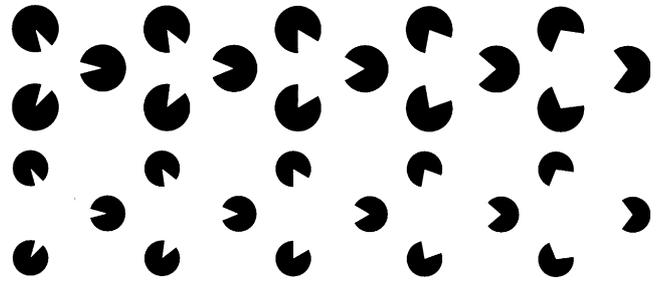

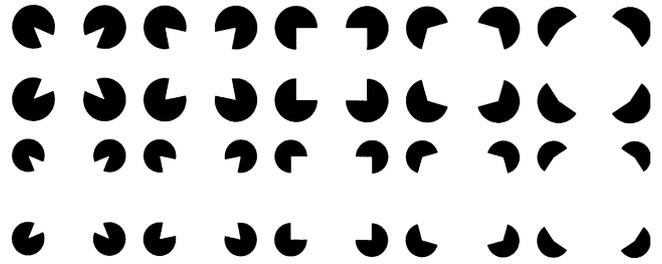Fig. 50. Curved Kanizsa triangles used for the experiment with Ninio.


Fig. 51. Curved Kanizsa squares used for the experiment with Ninio.

For different cases (triangle/square and small/large pacmen size) we compared three positions:

- the piecewise linear position (intersection of the sides of the two pacmen);
- the position chosen by the subjects;
- the circle position (extremum of the arc of circle tangent to the sides of the pacmen) (Fig. 53).

We show in Fig. 54 the results of the experiment for the case of the squares with small pacmen. We see that the model of the arc of circle becomes rapidly quite bad.

### 6.4. The first variational models

We look for models of *K*-contours based on the functional architecture of V1. We can suppose that the induction of activity propagated along the horizontal connections decreases when the angle between the preferred orientations $p_A$ and $p_B$ of the source and the target neurons increases, that is when the local curvature of the contour increases. The two properties of the propagation of activity: decay and curvature dependance, lead to look for *variational principles*. The models must be formulated in the fiber bundle $V = J^1 M$ or $V = CM$ and have to satisfy the two constraints:

1. a "geodesic" principle of length minimization;
2. a principle imposing the decreasing of the induced activity when the discrepancy between the two boundary angles $\theta_A$ and $\theta_B$ increases.
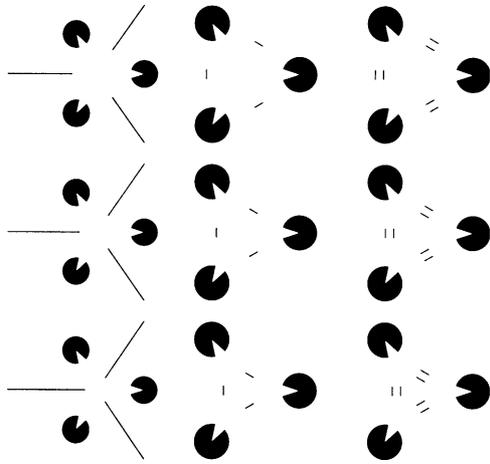
Fig. 52. The method of detection of the extremal point of a curved $K$-contour. The subject is asked to place a marker (the extremity of an orthogonal line, a small segment, the symmetry axis of a small stripe) as exactly as possible at the extremum.
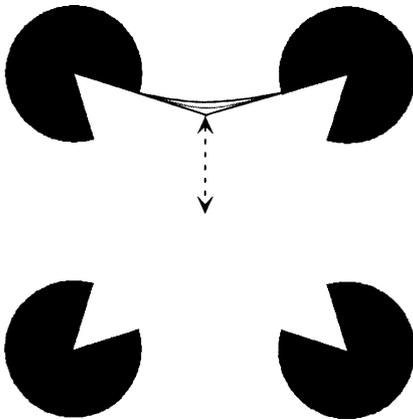


Fig. 53. Comparison of 3 $K$-contours: the piecewise rectilinear one (intersection of the corresponding sides of the two pacmen), the one chosen by the subjects, the circular one (arc of circle tangent to the sides of the pacmen).

After having evoked the early models proposed by Ullman and Horn, we will present three classes of models:

1. the "elastica" models introduced by David Mumford where the constraint 1 is realized through a geodesic principle in $M$ and the constraint 2 through a principle of minimization of the curvature (also in $M$);
2. our model of "Legendrian geodesics" where the constraint 1 is formulated in the fiber bundle $J^1M$ and where the constraint 2 corresponds to the choice of a metric in $J^1M$ which penalizes the discrepancy between $\theta_A$ and $\theta_B$;
3. analog models in the fiber bundle $M \times \mathbb{S}^1$.

As far as I know, Shimon Ullman [70] was the first to introduce the key idea of variational models in his 1976 seminal paper "Filling-in the gaps: the shape of subjective contours and a model for their generation". He remarked first that the problem of the global shape of modal illusory contours had not yet (at that time) being tackled:

An important but hitherto neglected problem posed by the filling-in phenomena concerns the *shape* of the filled-in contours and trajectories.

He developed then the variational hypothesis

that a network with the local property of trying to keep the contours 'as straight as possible' can produce curves possessing the global property of minimizing total curvature.

His conclusion was that the contour would be composed of two joined arcs of circle.

Among Ullman's hypotheses, the hypothesis of *locality* is rather problematic. It says that the continuation of
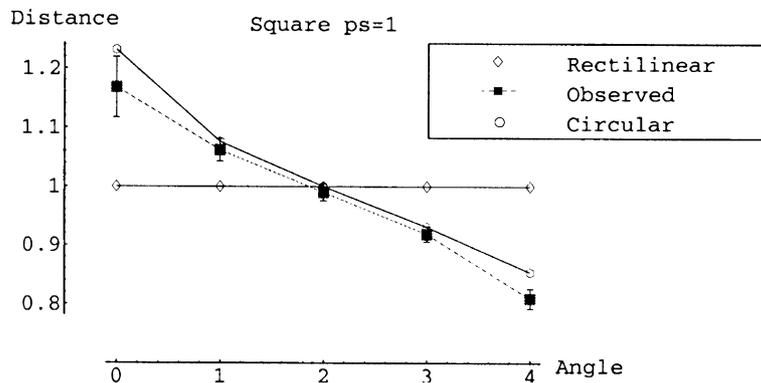


Fig. 54. The results of the experiment for the case of squares with small pacmen (parameter $ps$ = pacmen size = 1). The graphic plots the distance $d$ of the extremum of the $K$-contour to the center of the configuration as a function of the aperture angle. $d$ is measured by its ratio to the piecewise rectilinear case. Five aperture angles are considered: #2 corresponds to the classical case of a straight $K$-contour ($d_2 = 1$), #1 to a slightly concave one ($d_1 > d_2 = 1$), #0 to a more concave one ($d_0 > d_1 > d_2 = 1$), #3 to a slightly convex one ($d_3 < d_2 = 1$), #4 to a more convex one ($d_4 < d_3 < d_2 = 1$). We see that the observed empirical $K$-contour is clearly situated between the piecewise rectilinear one and the circular one.

a contour at a given point depends only on the orientation of the contour at this point, and not on the global shape of the contour. Brady et al. [6] renounced it and looked for contours minimizing the total curvature. By approximating the curvature $\kappa$ of a curve $y = f(x)$

$$\kappa = \frac{f''(x)}{(1 + f'(x)^2)^{3/2}} \qquad (36)$$

by $f''(x)$ when the contours deviate little from the straight line $f'(x) = 0$, they found classical interpolating curves, namely cubic splines. When the configuration is symmetric, the solution is degenerate and gives a parabola. Webb and Pervin [75] ended also at a parabola in generalizing Ullman's locality hypothesis to a problem of parallel transport: starting from a contour extrapolated between two limits, this contour must not change if the limits are moved along it.

In 1983, Horn [31] introduced the curve of least energy. But it was only in later works of Mumford, Nitzberg, Shiota, Williams and Jacobs [50,53,77] that non-trivial and theoretically well based models were worked out.

### 6.5. David Mumford's elastica model

In his 1992 celebrated paper on "Elastica and Computer Vision" [50] David Mumford introduced the elastica which are curves minimizing the integral of the square of the curvature $\kappa$ i.e. the energy

$$E = \int_\gamma (\alpha \kappa^2 + \beta) \, ds, \qquad (37)$$

where $\gamma$ is a curve with element of arc length $ds$.

This model can be justified in the following way. Due to the cortico-cortical horizontal connections, the inductive ends of the pacmen induce a propagation of activity along chains of neurons corresponding to approximatively co-axial contact elements $(a_i, p_i)$. The effective virtual contour will then correspond to the chain whose "leakages" are the weakest. But such a loss of activity has essentially two causes:

- a linear weakening equal to the number $N$ of the neurons of the chain, with a constant factor $\beta$.
- a curvature weakening, equal to the sum of the deviations of orientation between consecutive neurons, with a constant factor $\alpha$. If $\theta_i$ is the angle of the slope $p_i$, we can take for instance $\sum_{i=1}^{i=N-1} (\theta_{i+1} - \theta_i)^2$.

At the continuous limit, the number of neurons in the first term becomes the length $\int_\gamma ds$, and the sum of the deviations $\Delta\theta$ in the second term becomes the integral of curvature $\int_\gamma \kappa^2 ds$ for we have by definition $\kappa = \frac{d\theta}{ds}$. Minimizing the "leakage" terms leads therefore to the variational problem
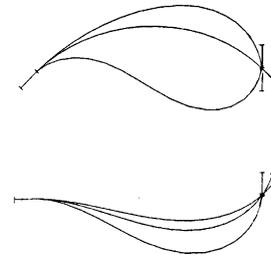


Fig. 55. Some examples of elastica.

$$\min \left( \int_\gamma (\alpha \kappa^2 + \beta) \, ds \right) \qquad (38)$$

with the boundary conditions

$$\begin{cases} f(x_A) = y_A, & f'(x_A) = \tan \theta_A, \\ f(x_B) = y_B, & f'(x_B) = \tan \theta_B, \end{cases} \qquad (39)$$

where $A$ and $B$ are the ends of $\gamma$.

This variational problem is well known in elasticity theory and comes back to Euler. Its solutions are called *elastica*. Elastica are not simple algebraic curves but they can be given an explicit form using elliptic functions. Fig. 55 represents some examples.

David Mumford developed a deep *stochastic* explanation of the role of elastica in vision. Let us suppose that the curvature $\kappa(s)$ of the curve $\gamma$ (parametrized by its arc length $s$) is a *white noise*. As $\kappa(s) = \dot\theta(s)$ (where $\dot\theta(s) = \frac{d\theta(s)}{ds}$), this entails that $\theta(s)$ is a *Brownian movement*, and that at each moment the movement is a Gaussian random variable with vanishing mean and variance $\sigma$. If we suppose that the length $l$ of $\gamma$ is a random variable obeying an exponential law $\lambda e^{-\lambda l} \, dl$ ($l$ is therefore constant for $\lambda = 0$), the probability $\Pr(\gamma)$ of a curve $\gamma$ is given by

$$\Pr(\gamma) = e^{-\int (\alpha \kappa^2 + \beta) \, ds} \qquad (40)$$

with $\alpha = \frac{1}{2\sigma^2}$ and $\beta = \lambda$. Elastica are therefore the most probable curves. As explains Mumford [50, p. 496].

Thus we see that elastica have the interpretation of being the *mode* of the probability distribution underlying this stochastic process restricted to curves with prescribed boundary behavior, e.g., the maximum likelihood curve with which to reconstruct hidden contours.

Elastica are solutions of a second order differential equation in the curvature $\kappa$. Let us give a sketch of Mumford's proof. One starts from the well known fact that if $t(s)$ and $n(s)$ are respectively the unitary tangent and normal vectors to $\gamma$ at the point $a(s)$, then

$$\begin{cases} \dot{a}(s) = t(s), \\ \dot{t}(s) = \kappa(s)n(s), \\ \dot{n}(s) = -\kappa(s)t(s). \end{cases} \qquad (41)$$

One considers then a deformation $a_\delta(s) = a(s) + \delta(s)n(s)$ of $a(s)$ (i.e. of $\gamma$) where $\delta(s)$ is a small perturbation. Taking the derivative and expressing the condition of preservation of lengths, one gets a first-order estimation of the perturbed curvature $\kappa_\delta(s)$

$$\kappa_\delta(s) = \kappa(s) + \ddot{\delta}(s) + \delta(s)\kappa(s)^2. \qquad (42)$$

By computing anew the integral $\int \kappa(s)^2 \, ds$, developing at first order, and making an integration by parts with the condition $\int \delta(s)\kappa(s) \, ds = 0$ (expressing the constancy of the length $\int ds$), one finally gets the constraint

$$\int (2\ddot{\kappa}(s) + \kappa(s)^3)\delta(s) \, ds = 0 \quad \text{if} \quad \int \delta(s)\kappa(s) \, ds = 0. \qquad (43)$$

As $\delta(s)$ is an arbitrary perturbation, this implies that $2\ddot{\kappa} + \kappa^3$ must be proportional to $\kappa$. Hence the differential equation

$$2\ddot{\kappa}(s) + \kappa(s)^3 = b\kappa(s). \qquad (44)$$

Multiplying by $\dot{\kappa}(s)$ and integrating, one gets

$$\dot{\kappa}(s)^2 + \frac{1}{4}\kappa(s)^4 = \frac{b}{2}\kappa(s)^2 + c, \qquad (45)$$

where $c$ is an integration constant.

The relation between $\kappa$ and $s$ is therefore given by the elliptic integral

$$s = \int \frac{2d\kappa}{\sqrt{-\kappa^4 + 2b\kappa^2 + 4c}}. \qquad (46)$$

If $E$ is the elliptic curve of equation

$$v^2 = -u^4 + 2bu^2 + 4c, \qquad (47)$$

the map $s \mapsto (\kappa(s), 2\dot{\kappa}(s))$ sends the elastica $\gamma$ onto $E$. Using the classical theory of elliptic curves, Mumford has shown how elastica can be parametrized by theta functions.

## 6.6. Illusory contours and the contact bundle

### 6.6.1. From elastica to variational problems in jet spaces

Up to now, we have considered modal illusory contours as *plane* curves in the base space $M$. But V1 implements a contact bundle. It is therefore natural to
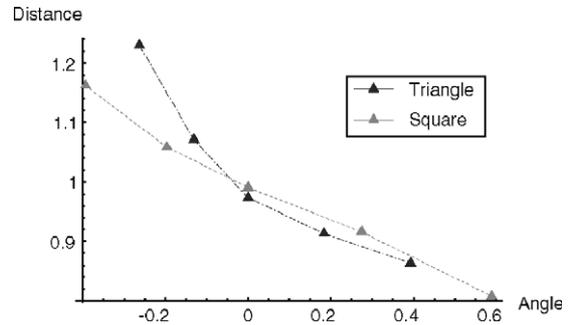


Fig. 56. Comparison of the extremal point of $K$-contours for triangles and squares with the same boundary conditions. Their difference shows that the visual system computes rather minimal surfaces than minimal arcs.

look for curves minimizing some functional defined on that space.

In the geometrical framework of the fibration $\pi : V = M \times P \to R$ idealizing V1, we can explain the Kanizsa problem in a principled way. Two pacmen of respective centers $A$ and $B$ with a specific aperture angle define two contact elements $(a, p) = (A, p_A)$ and $(b, q) = (B, p_B)$ of $V$. A $K$-curve interpolating between $(a, p)$ and $(b, q)$ is

1. a curve $\gamma$ from $a$ to $b$ in $M$ with tangent $p$ at $a$ and tangent $q$ at $b$ (we will suppose that $\gamma$ is defined by an equation $y = f(x)$);
2. a curve minimizing some Lagrangian, that is a functional defining some sort of "energy" for $\gamma$ (variational problem).

If we lift the problem to $V$, we must find in $V$ a curve $\Gamma$ of the form $(x, y = f(x), p = g(x))$ interpolating between $(a, p)$ and $(b, q)$ in $V$, which is at the same time:

1. "as straight as possible", that is "geodesic" in $V$; as the variation of $p$ measures the curvature $\kappa$ of $\gamma$, this is a condition on minimizing in some way the curvature;
2. *an integral curve of the contact structure*, that is which satisfies the integrability condition $g(x) = f'(x)$.

### 6.6.2. Illusory contours and minimal surfaces

In fact, the problem of modeling modal illusory contours is rather more complex. Indeed, the experiment with Ninio shows that the deflections for the triangle and for the square *are not the same* (Fig. 56).

This is a very interesting global effect which shows that the visual system constructs not only virtual curves but also virtual *surfaces* [45,61], which are solutions of a far more complex variational model defined on $V$ endowed with its contact-Lie structure $\mathscr{C}$ and its sub-Riemannian Carnot-Carathéodory metric $d_{\mathscr{C}}$ (see Sections 4.2.5 and 4.2.6). Now, even the problem of minimal

surfaces with predefined boundaries is quite difficult. Some results are already available, especially those of Scott Pauls [56]. The first point is to define an appropriate area measure for surfaces in $V$. Misha Gromov [23] proposed to take the 3-dimensional Hausdorff measure $H^3$ (associated to $d_\mathscr{C}$) of the surfaces. Pansu [54] has shown that

$$H^3(S) = \int \|\nabla_\mathscr{C}(f)\| \, d\sigma,$$

where $f = 0$ is an equation for $S$, and $d\sigma$ is the Riemannian area element induced by the left-invariant metric $g$ making the left-invariant basis $\{t_1, t_2, t_3\}$ orthonormal.

The reader will find in Pauls [56] a variational setup for minimal surfaces in $V$ which uses as for Lagrangian the projection $n_\mathscr{C}^0$ of the normal vector $n$ of $S$ onto $\mathscr{C}$ (the vector $n_\mathscr{C}$ of Section 4.2.6 is the normalization of $n_\mathscr{C}^0$). Here we will simplify the problem and restrict ourselves to minimal *curves*.

### 6.6.3. Illusory contours and minimal curves

We must define appropriate Lagrangians on $V$ and study the curves $\Gamma$ which are solutions of the associated Euler–Lagrange equations. We will work out three cases, one very natural but analytically complex, two less natural but easier to formulate analytically. We will see that the projections of these curves on the base space $M$ present strong analogies with elastica.

To define Lagrangians in the contact bundle $CM$ or the 1-jet space $V = J^1M$ (or any other fiber bundle coding the orientation of tangents to curves $\gamma$ in $M$) we must define first a Riemannian metric which reflects the weakening of the horizontal cortico-cortical connections when the discrepancy between the boundary values $\theta_A$ and $\theta_B$ increases. If $\theta$ is measured relatively to the axis $AB$ ($\theta$ has an *intrinsic* geometric meaning), the weakening must vanish for $\theta = 0$ and $\theta = \pi$, and diverge for $\theta = \frac{\pi}{2}$. The function $p = f' = \tan \theta$ being the simplest function sharing this properties, it seems justified to test first the Euclidean metric of $V = J^1M$. We will therefore use a frame $Oxy$ of $M$ where the $x$ axis is identified to $AB$. The invariance under a change of frame is then expressed by the action of the Euclidean group $E(2)$ on $V$ (see Section 4.2.2).

### 6.6.4. "Legendrian geodesics" in the contact bundle and sub-Riemannian geometry

According to the sub-Riemannian setup associated to a class of distinguished curves, we look then for curves of minimal length in $V$ *among those which are Legendrian lifts*, that is which satisfy the Frobenius integrability condition and are integrals of the contact structure $\mathscr{C}$. We will call "Legendrian geodesics" the solutions of this constrained variational problem.

Let $(x, y, p; \xi, \eta, \pi)$ be local coordinates in the tangent space $TV$ of $V = J^1M \simeq \mathbb{R}^3$. We have to minimize the length of $\gamma$ expressed by the functional $\int_{x_A}^{x_B} ds$ where $ds$ is given by

$$ds^2 = dx^2 + dy^2 + dp^2. \tag{48}$$

The energy is therefore $E = \int_{x_A}^{x_B} L(x) \, dx$, where the Lagrangian $L$ is given, for a curve $\Gamma$ of the form $(x, f(x), f'(x))$, by the formula $L(x) \, dx = ds$, that is

$$L(x) = \sqrt{\xi^2 + \eta^2 + \pi^2} = \sqrt{1 + f'(x)^2 + f''(x)^2}. \tag{49}$$

We have to solve the *constrained* Euler–Lagrange (E–L) equations. The *non*-constrained E–L equations are

$$\begin{cases} \dfrac{\partial L}{\partial y} - \dfrac{d}{dx}\left(\dfrac{\partial L}{\partial \eta}\right) = 0, \\[2mm] \dfrac{\partial L}{\partial p} - \dfrac{d}{dx}\left(\dfrac{\partial L}{\partial \pi}\right) = 0. \end{cases} \tag{50}$$

They cannot be applied as such for we must take into account the integrability constraint $p = f'(x)$, i.e. $p = \eta$, which can be written: $\Sigma = 0$, with $\Sigma = p - \eta$ (this is also equivalent to the integral constraint $\int_{x_A}^{x_B} (p - \eta)^2 \, dx = 0$). The constrained E–L equations are then

$$\begin{cases} \left(\dfrac{\partial}{\partial y} - \dfrac{d}{dx}\dfrac{\partial}{\partial \eta}\right)(L + \lambda\Sigma) = 0, \\[2mm] \left(\dfrac{\partial}{\partial p} - \dfrac{d}{dx}\dfrac{\partial}{\partial \pi}\right)(L + \lambda\Sigma) = 0, \end{cases} \tag{51}$$

where $\lambda(x)$ is a function, called a *Lagrange multiplier*. The idea is that the E–L equations with the constraint $\Sigma = 0$ are the same as the non-constrained E–L equations for the Lagrangian $L + \lambda\Sigma$.

If we substitute their expression for $L$ and $\Sigma$ and if we express the variables $y$, $p$, $\eta$, $\pi$ as functions of $x$, we get

$$\begin{cases} \dfrac{d}{dx}\left[\dfrac{\partial L}{\partial \eta} - \lambda(x)\right] = 0, \\[2mm] \lambda(x) - \dfrac{d}{dx}\dfrac{\partial L}{\partial \pi} = 0, \end{cases} \tag{52}$$

that is

$$\begin{cases} \dfrac{d}{dx}\left[\dfrac{f'(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} - \lambda(x)\right] = 0, \\[2mm] \lambda(x) - \dfrac{d}{dx}\dfrac{f''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} = 0. \end{cases} \tag{53}$$

Let us integrate the first differential equation, and eliminate $\lambda(x)$ with the second equation. There exists a constant $A$ such that

$$\frac{\partial L}{\partial f'} = A + \lambda(x) = A + \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial L}{\partial f''}, \tag{54}$$

that is

$$\frac{f'(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} = A + \frac{\mathrm{d}}{\mathrm{d}x}\frac{f''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}}. \tag{55}$$

Let us develop the total derivative

$$\frac{\mathrm{d}L}{\mathrm{d}x} = \frac{\partial L}{\partial x} + \frac{\partial L}{\partial f}f' + \frac{\partial L}{\partial f'}f'' + \frac{\partial L}{\partial f''}f'''$$
$$= \frac{\partial L}{\partial f'}f'' + \frac{\partial L}{\partial f''}f'''. \tag{56}$$

As we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(f''\frac{\partial L}{\partial f''}\right) = f'''\frac{\partial L}{\partial f''} + f''\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial f''}\right)$$
$$= f'''\frac{\partial L}{\partial f''} + f''\left(\frac{\partial L}{\partial f'} - A\right), \tag{57}$$

according to Eq. (54) we get

$$\frac{\mathrm{d}L}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\left(f''\frac{\partial L}{\partial f''}\right) + Af'', \tag{58}$$

that is

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(L - f''\frac{\partial L}{\partial f''}\right) = Af'' \tag{59}$$

and, after a second integration

$$L - f''\frac{\partial L}{\partial f''} = Af' + B. \tag{60}$$

In our case, we get

$$\frac{f'(x)f''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}}$$
$$= Af''(x) + \frac{\mathrm{d}}{\mathrm{d}x}\frac{f''(x)^2}{\sqrt{1 + f'(x)^2 + f''(x)^2}}$$
$$- \frac{f''(x)f'''(x)}{\sqrt{1 + f'(x)^2 + f''(x)^2}} \tag{61}$$

and

$$\sqrt{1 + f'(x)^2 + f''(x)^2}$$
$$= Af'(x) + B + \frac{f''(x)^2}{\sqrt{1 + f'(x)^2 + f''(x)^2}}, \tag{62}$$

that is

$$1 + f'(x)^2 = (Af'(x) + B)\sqrt{1 + f'(x)^2 + f''(x)^2}. \tag{63}$$

If we take the square of this equation and put $g = f'$, we get the final equation

$$(g')^2 = \frac{(1 + g^2)^2 - (1 + g^2)(Ag + B)^2}{(Ag + B)^2}, \tag{64}$$

whose solution is given by the *elliptic integral* (expressing $x$ as a function of $g$ and not $g$ as a function of $x$)

$$x = C + \int_a^{g(x)} \frac{At + B}{\sqrt{(1 + t^2)[(1 + t^2) - (At + B)^2]}}\, \mathrm{d}t. \tag{65}$$

The solutions $f(x)$ of the E–L equations for the Legendrian geodesics are therefore integrals of elliptic functions associated to elliptic curves of equation

$$v^2 = \frac{(1 + u^2)[(1 + u^2) - (Au + B)^2]}{(Au + B)^2}. \tag{66}$$

We can greatly simplify the solution of the equation when the function $f$ is *even*, and the curve $\gamma$ *symmetric* under the symmetry $x \leftrightarrow -x$. Indeed, this condition implies immediately $A = 0$, whence, putting $k = 1/B$, the simpler differential equation for $g = f'$

$$(g')^2 = (1 + g^2)[k^2(1 + g^2) - 1]. \tag{67}$$

The parameter $k$ is correlated to curvature. In fact (see below), $k^2 - 1 = \kappa(0)^2$. As (if $f$ is differentiable at 0) symmetry implies $f'(0) = g(0) = 0$, we must have $f''(0)^2 = g'(0)^2 = k^2 - 1 \geqslant 0$, whence $|k| \geqslant 1$. We get therefore

$$x = C + \int_0^{g(x)} \frac{1}{\sqrt{(1 + t^2)\left(1 + \frac{k^2}{k^2 - 1}t^2\right)}}\, \mathrm{d}t, \tag{68}$$

which is a well-known elliptic integral of the first kind.

Fig. 57 shows how the solution $g = f'$ evolves when $k$ varies from 1 to 1.65 by steps of 0.5. We see that the module of the slopes of the tangents at the end points
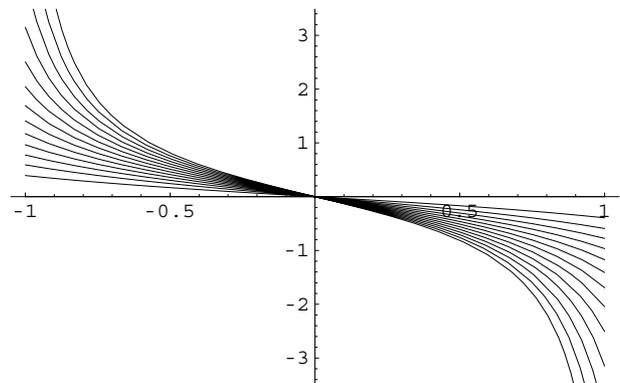


Fig. 57. Evolution of the derivative $g = f'$ of Legendrian geodesics when the parameter $k$ varies from 1 to 1.65 by steps of 0.5.
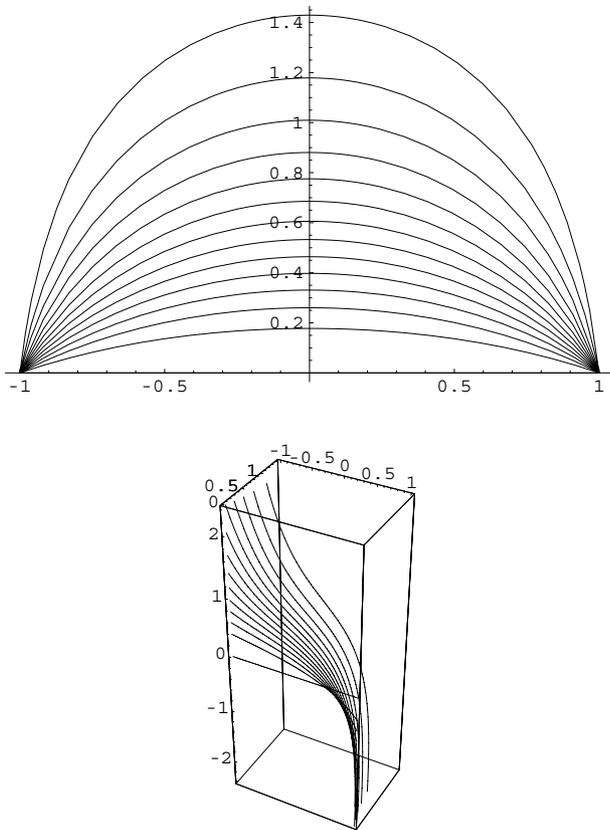
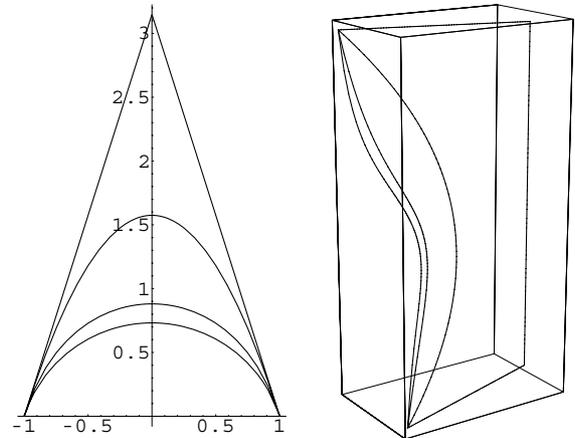Fig. 58. Evolution of Legendrian geodesics $f$ when the boundary tangents become more and more vertical.



Fig. 59. The Legendrian geodesic is intermediary between, on the one hand, the arc of circle and, on the other hand, the arc of parabola and the piecewise linear solution.

$p_A = \tan \theta_A$ and $p_B = \tan \theta_B$ increases. Fig. 58 shows the evolution of the integral $f$ of $g$.

Fig. 59 shows in the base space $M$ and in the bundle $J^1 M$ how the Legendrian geodesic corresponding to $k = 1.5$ is situated relatively to the arc of circle, the arc of parabola and the piecewise linear solution defined by the same boundary conditions. The following table shows that the geodesic minimizes the length:

| Curves | Geodesic | Circle arc | Parabole arc | Peathwise linear |
|--------|----------|------------|--------------|------------------|
| Length | 7.02277 | 7.04481 | 7.50298 | 12.9054 |

As for elastica, we can try to deduce a differential equation for the curvature. Let us recall (see Section 6.5) that, if $s$ is the arc length and $a = (x, y = f(x))$ a point of $\gamma$, the unitary tangent vector at $a$ is given by $t = \frac{da}{ds} = \dot{a} = \frac{(1, f')}{\sqrt{1 + (f')^2}}$ (because $ds = \sqrt{1 + (f')^2} \, dx$). The unitary normal vector $n$ at $a$ is therefore given by $n = \frac{(-f', 1)}{\sqrt{1 + (f')^2}}$, and, from the formula $\frac{dt}{ds} = \kappa n = \frac{(-f'f'', f'')}{(1 + (f')^2)^2}$, we deduce the curvature

$$\kappa = \frac{f''}{(1 + (f')^2)^{3/2}}. \tag{69}$$

If $g = f'$, we get therefore the system of equations: [20]

$$\begin{cases} \kappa = \frac{g'}{(1 + g^2)^{3/2}}, \\ (g')^2 = (1 + g^2)[k^2(1 + g^2) - 1]. \end{cases} \tag{70}$$

To eliminate $g$ and $g'$, we replace first $(g')^2$ by its value in $\kappa^2$, whence $\kappa^2 = \frac{k^2(1+g^2)-1}{(1+g^2)^2}$. Next, we derive $\kappa$ and the second equation, which allows to express $\kappa'$ and $\dot{\kappa}$ as functions of $g$. We get $\dot{\kappa} = \pm \frac{g[2 - k^2(1+g^2)]}{(1+g^2)^2}$. Finally, we eliminate $g$ between $\kappa^2$ and $\dot{\kappa}$ and we get the equation for $\dot{\kappa}^2 = \left(\frac{d\kappa}{ds}\right)^2$:

$$\dot{\kappa}^4 + \dot{\kappa}^2(2\kappa + k^2)(2\kappa - k^2)(2\kappa^2 - k^2(k^2 - 1))$$
$$+ \kappa^2(2\kappa + k^2)^2(2\kappa - k^2)^2(\kappa^2 - (k^2 - 1))$$
$$= 0, \tag{71}$$

whose solutions are

$$\dot{\kappa}^2 = \frac{1}{2}\left[k^2(k^2 - 1) - 2\kappa^2 \pm (k^2 - 1)\sqrt{(k^2 - 2\kappa)(k^2 + 2\kappa)}\right], \tag{72}$$

to be compared to that of elastica presented in Section 6.5 (Eq. (45))

$$\dot{\kappa}^2 + \frac{1}{4}\kappa^4 = \frac{b}{2}\kappa^2 + c. \tag{73}$$

For $k^2 = 2$, we get the simplified equation

$$\dot{\kappa}^2 = (1 - \kappa^2) \pm \sqrt{1 - \kappa^2}. \tag{74}$$

---

[20] As $g(0) = 0$, these equations give $g'(0)^2 = k^2 - 1$ and the relation $\kappa(0)^2 = k^2 - 1$ evoked above.

### 6.6.5. The circle bundle model

We can apply the precedent computations to the case of the fibration $M \times \mathbb{S}^1$ whose tangent bundle has local coordinates $(x, y, \theta; \xi, \eta, \varphi)$ and whose metric is given by

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}\theta^2. \tag{75}$$

The Lagrangian is now

$$
\begin{aligned}
L(x) &= \sqrt{\xi^2 + \eta^2 + \varphi^2} \\
&= \sqrt{1 + f'(x)^2 + \frac{f''(x)^2}{(1 + f'(x)^2)}}.
\end{aligned} \tag{76}
$$

Indeed, $\theta = \mathrm{Arctan}(f')$ and $\varphi = \theta' = \frac{f''(x)}{1+f'(x)^2}$. The constraint is now $\Sigma = 0$, with $\Sigma = \theta - \mathrm{Arctan}(\eta)$, and the E–L equations are

$$
\begin{cases}
\dfrac{\mathrm{d}}{\mathrm{d}x}\left[\dfrac{\partial L}{\partial \eta} + \lambda \dfrac{\partial \Sigma}{\partial \eta}\right] = 0, \\[2mm]
\lambda \dfrac{\partial \Sigma}{\partial \theta} - \dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{\partial L}{\partial \varphi} = 0
\end{cases} \tag{77}
$$

with $\frac{\partial \Sigma}{\partial \theta} = 1$ and $\frac{\partial \Sigma}{\partial \eta} = -\frac{1}{1+\eta^2}$. We get therefore

$$\frac{\partial L}{\partial f'} = A - \lambda \frac{\partial \Sigma}{\partial \eta} = A - \left(\frac{\partial \Sigma}{\partial \theta}\right)^{-1} \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial \varphi}\right)\frac{\partial \Sigma}{\partial \eta}. \tag{78}$$

We develop the total derivative

$$
\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}x} &= \frac{\partial L}{\partial x} + \frac{\partial L}{\partial f}f' + \frac{\partial L}{\partial f'}f'' + \frac{\partial L}{\partial \varphi}\varphi' \\
&= \frac{\partial L}{\partial f'}f'' + \frac{\partial L}{\partial \varphi}\varphi',
\end{aligned} \tag{79}
$$

we write

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\left(\varphi \frac{\partial L}{\partial \varphi}\right) &= \varphi' \frac{\partial L}{\partial \varphi} + \varphi \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial \varphi}\right) \\
&= \varphi' \frac{\partial L}{\partial \varphi} + \varphi\left(A - \frac{\partial L}{\partial f'}\right)\left(\frac{\partial \Sigma}{\partial \eta}\right)^{-1},
\end{aligned} \tag{80}
$$

whence

$$
\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}x} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\varphi \frac{\partial L}{\partial \varphi}\right) + \frac{\partial L}{\partial f'}f'' \\
&\quad - \varphi\left(A - \frac{\partial L}{\partial f'}\right)\left(\frac{\partial \Sigma}{\partial \eta}\right)^{-1}.
\end{aligned} \tag{81}
$$

But as $\frac{\partial \Sigma}{\partial \eta} = -\frac{1}{1+\eta^2}$, $\left(\frac{\partial \Sigma}{\partial \eta}\right)^{-1} = -(1+\eta^2)$, we have $-\varphi\left(\frac{\partial \Sigma}{\partial \eta}\right)^{-1} = f''$ and therefore

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(L - \varphi \frac{\partial L}{\partial \varphi}\right) = Af'', \tag{82}$$

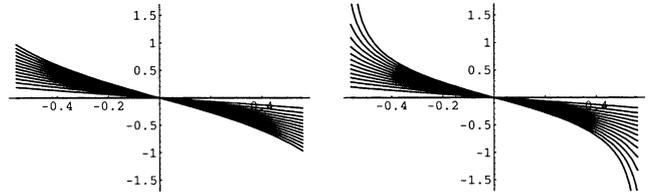equation to be compared with Eq. (59). An integration yields



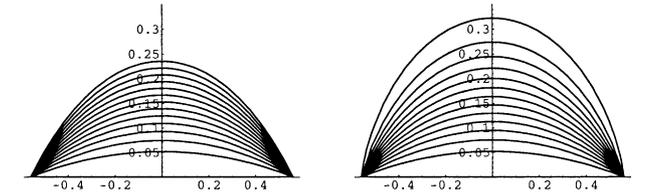Fig. 60. Comparison between the families $g = f'$ solutions of equations (67) and (85).



Fig. 61. Comparison between the solutions $f$ of Eqs. (67) and (85).

$$L - \varphi \frac{\partial L}{\partial \varphi} = Af' + B. \tag{83}$$

The same computation as before gives (with $g = f'$)

$$(g')^2 = \frac{(1+g^2)^4 - (1+g^2)^3(Ag+B)^2}{(Ag+B)^2}, \tag{84}$$

equation to be compared with the previous equation (64). In the symmetric case $A = 0$, we get (with $k = 1/B$)

$$(g')^2 = (1+g^2)^3[k^2(1+g^2) - 1], \tag{85}$$

to be compared with Eq. (67).

Figs. 60 and 61 compare the solutions of this equation (on the interval $(-0.56, 0.56)$ for singularities appear on larger intervals) to those of Eq. (67).

We see that for admissible boundary conditions (not too large boundary slopes), the solutions of the two equations are very close. But they depart clearly from each other when the boundary slopes become more pronounced, those of the first equation rising higher than those of the second equation.

In what concerns the differential equation for the curvature, it is simpler than in the previous case. We get

$$
\begin{cases}
\kappa^2 = \dfrac{(g')^2}{(1+g^2)^3} = \dfrac{(1+g^2)^3[k^2(1+g^2)-1]}{(1+g^2)^3} = k^2(1+g^2) - 1, \\[2mm]
\dot{\kappa} = k^2 g(1+g^2)
\end{cases} \tag{86}
$$

and the elimination of $g$ gives the equation

$$\dot{\kappa}^2 k^2 = (\kappa^2 + 1 - k^2)(1 + \kappa^2)^2 \tag{87}$$

to be compared with Eqs. (74) and (75).

This equation can be explicitly integrated. We get (with $C$ an integration constant)

$$\kappa(s) = \pm \left[ \frac{(e^{2s} - e^{2C})^2 (k^2 - 1)}{(e^{2s} - e^{2C})^2 - k^2 (e^{2s} + e^{2C})^2} \right]^{1/2}. \quad (88)$$

### 6.6.6. Curves of least energy

The integration of the differential equation for Legendrian geodesics is rather difficult due to the presence of square roots. This obstacle disappears if we consider another functional whose Lagrangian is the square of that of geodesics.

The new Lagrangian is

$$L(x) = \xi^2 + \eta^2 + \pi^2 = 1 + f'(x)^2 + f''(x)^2. \quad (89)$$

The constrained variational problem gives the E–L equations

$$\begin{cases} \dfrac{d}{dx}[2f'(x) - \lambda(x)] = 0, \\ \lambda(x) - \dfrac{d}{dx}[2f''(x)] = 0. \end{cases} \quad (90)$$

If we integrate the first equation and eliminate $\lambda(x)$, we get (with $C$ an integration constant)

$$g = f'(x) = \frac{1}{2}\lambda(x) + C = f'''(x) + C. \quad (91)$$

This second-order differential equation in $f'$ is easy to integrate and gives for $f$ functions of the form

$$f(x) = Ae^x + Be^{-x} + Cx + D, \quad (92)$$

where $A, B, C, D$ are constants of integration fixed by the boundary conditions (which impose 4 independant equations corresponding to the values of $f$ and $f'$ at the two end points of $\gamma$). Fig. 62 gives an example.

Analog curves has been evoked by David Mumford in his study of elastica where he uses in dimension 3 the Lagrangian

$$L(x) = \gamma + \beta(1 + zf'(x)^2) + \alpha f''(x)^2. \quad (93)$$

### 6.6.7. Comparison between the different models

We have seen that the Legendrian geodesics satisfy a differential equation (72) for the curvature $\kappa$ which is
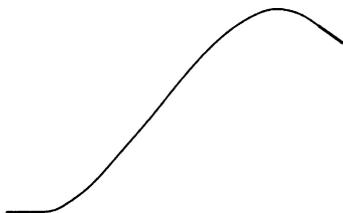


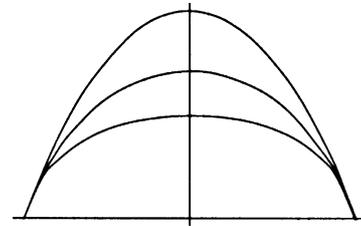Fig. 62. An example of a curve of least energy.



Fig. 63. Comparison between the solutions of three variational models. From top to bottom: elastica computed numerically with the algorithm of Nitzberg et al. [53] ($\alpha = 1/10$), projection of the Legendrian geodesic in the jet space $J^1M$, projection of the curve of least energy in $J^1M$.

rather complex compared to that of elastica. But, reciprocally, elastica are solutions of a variational problem which becomes in its turn rather complex when expressed in the contact bundle $CM$ or the 1-jet bundle $J^1M$.

As we have seen before, the local curvature of a plane curve of equation $y = f(x)$ is given by $\kappa = \frac{f''(x)}{(1+f'(x)^2)^{3/2}}$. If we take for $\alpha$ the quotient $\frac{\alpha}{\beta}$ in Eq. (38) (i.e. if we take $\beta = 1$), elastica minimize the functional:

$$\int_{s_A}^{s_B} (1 + \alpha\kappa^2)\, ds$$

$$= \int_{x_A}^{x_B} \left( 1 + \alpha \frac{f''(x)^2}{(1 + f'(x)^2)^3} \right) \sqrt{1 + f'(x)^2}\, dx$$

$$= \int_{x_A}^{x_B} \left( \sqrt{1 + f'(x)^2} + \alpha \frac{f''(x)^2}{(1 + f'(x)^2)^{5/2}} \right) dx. \quad (94)$$

The lifting of an elastica in the 1-jet bundle minimizes a functional whose Lagrangian presents strong analogies with the Lagrangians studied above. Fig. 63 compares, in the symmetric case and for the same boundary conditions, the elastica, the projection of the Legendrian geodesic and the projection of the minimal energy curve.

## 7. Application 3: Spontaneous geometric visual patterns

As a third application of the contact structure of V1, we will establish a link with the beautiful results worked out by Bressloff et al. [7] concerning the explanation of some geometric visual hallucinations from the functional architecture of the striate cortex.

Extending previous results of Ermentrout and Cowan [15], the authors work in the fibration $\pi : V = \mathbb{R}^2 \times \mathbb{S}^1 \to \mathbb{R}^2$ with local coordinates $(a, \theta)$ ($\theta =$ the angular coordinate of the orientation $p$). Neurons parametrized by pairs $(a, \theta)$ present an activity $z(a, \theta)$ and are connected by connections with weights $w\langle a, \theta | a', \theta' \rangle$. Their dynamics select then specific patterns of activity. As in physics, such models include two parts

(i) a geometrical understructure;
(ii) an activity field governed by a field equation.

Let $z(a, \theta, t)$ be the function expressing the activity of V1 at time $t$. Generalizing the classical Hopfield equations of a neural net, the authors show that the partial differential equation (PDE) governing the evolution of $z$ has the form

$$\frac{\partial z(a, \theta, t)}{\partial t} = -\alpha z(a, \theta, t)$$
$$+ \frac{\mu}{\pi} \int_0^\pi \int_{\mathbb{R}} w\langle a, \theta | a', \theta' \rangle \sigma(z(a', \theta', t)) \, da' \, d\theta'$$
$$+ h(a, \theta, t), \tag{95}$$

where $\sigma$ is a gain function (with $\sigma(0) = 0$), $h$ an external input, and $w\langle a, \theta | a', \theta' \rangle$ the weight of the connection linking $(a, \theta)$ to $(a', \theta')$.

The key point is that *the functional architecture of V1—that is its contact structure—can be expressed by the weights $w$.*

1. The local vertical connections inside a single hypercolumn yield a term

   $$w\langle a, \theta | a', \theta' \rangle = w_{\text{ver}}(\theta - \theta') \delta(a - a'), \tag{96}$$

   where $\delta$ is the Dirac function and the factor $\delta(a - a')$ imposes therefore $a = a'$.
2. The lateral horizontal connections between different hypercolumns yield a term

   $$w\langle a, \theta | a', \theta' \rangle = w_{\text{hor}}(a - a') \delta(\theta - \theta'), \tag{97}$$

   where the factor $\delta(\theta - \theta')$ imposes $\theta = \theta'$ and expresses the fact that the horizontal cortico-cortical connections connect pairs $(a, p)$ and $(b, q)$ with $p = q$.
3. Moreover, the coaxiality condition $p = q = ab$ is expressed by the fact that

   $$w_{\text{hor}}(a - a', \theta) = w_{\text{hor}}(s) \delta(a - a' - se_\theta), \tag{98}$$

   where $e_\theta$ is the unit vector in the direction $\theta$.

This is a good example of what we explain in Section 4.2.6 concerning sub-Riemannian Carnot-Carathéodory metrics $d_{\mathscr{C}}$ as limit of Riemannian metrics $d_\varepsilon$. If we take instead of a Dirac function $\delta$ Gaussians $G_\varepsilon$ we can implement a metric $d_\varepsilon$. When at the limit the Gaussians $G_\varepsilon$ become $\delta$ functions, we implement the sub-Riemannian metric $d_{\mathscr{C}}$.

As the contact structure, the weights $w$ are $E(2)$-invariant and the PDE is therefore $E(2)$-equivariant. Its spectral analysis shows that the initial activation state $z \equiv 0$ (which is a stable solution for $\mu = 0$ if the external input $h = 0$) can become *unstable* and *bifurcate* for critical values of the parameter $\mu$. The new stable activation states present highly structured spatial patterns generated by a $E(2)$-symmetry breaking.

In more recent papers (see this volume), the authors use more complex fibrations to take into account the dimension of spatial frequency.

## 8. "Geodesic" models and Lie groups

As we have shown in Petitot [58], the variational model for Legendrian geodesics can be formulated in a deeper way on the *Lie group* $G = E(2) = SO(2) \rtimes \mathbb{R}^2$ of displacements in $M$. This amounts to use the celebrated *moving frame method* due to Elie Cartan (see Fig. 64), a method geometrically relevant since the group $G$ characterizes the Euclidean geometry of the plane, but also a method neurophysiologically relevant since:

1. the functional architecture of V1 is $G$-invariant;
2. it is probably $G$ which is neurally implemented if we take into account areas V1 and V2; we have seen that when an element of contour is activated it is also the case for the *orthogonal direction*; we can think therefore that V1 and V2 implement together the fibration $G \to M$ with fiber $SO(2) \simeq \mathbb{S}^1$;
3. the geometry of $G$ is universal for the contour problem; it idealizes geometrically a functional architecture which, as was deeply anticipated by Poincaré and Husserl, couples perception with the *kinesthesic sense of movement*;[21] concretely, the translations of $G$ can be kinesthesically interpreted, the motor control of vision allowing a change of moving frame.

The topic is mathematically too technical to be treated here, but let us nevertheless sketch some of its aspects.

### 8.1. Moving frames and the principal bundle of the Euclidean group

Let us first say some words concerning the structure of the Lie group $G$. Let $R_0 = (O, \varepsilon_1, \varepsilon_2)$ be a fixed orthonormal frame of the plane $M$ (coordinates $(x, y)$). Let $\gamma$ be a curve in $M$. At every point $p = (x, y)$ of $M$,[22] we can consider an orthonormal moving frame—also called a Frénet frame—$R = (p, e_1, e_2) = (p, r_\theta)$ with

---

[21] Concerning the "sense of movement", see Alain Berthoz's fundamental work [3]. In what concerns geometric models for the phenomenological description of perception and kinesthesy in Husserl, see Petitot [59].

[22] As in Section 4.2.2, we indicate here by $p$ or $q$ a point of $M$ which is the origin of a moving frame. A current point of $M$ will remain indicated by $a$.
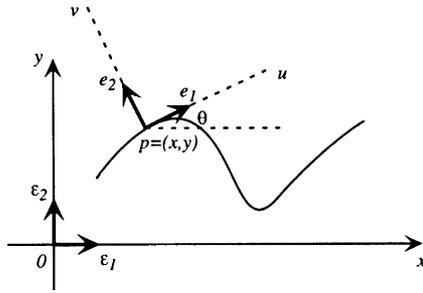
Fig. 64. Elie Cartan's concept of moving frame.

center $p$ and basis $(e_1, e_2)$, the unitary vectors respectively tangent and normal to $\gamma$ at $p$ (coordinates $(u, v)$) (Fig. 64). [23] $r_\theta$ is the rotation of angle $\theta$ transporting, with the translation $p$, the fixed frame $R_0$ onto the moving frame $R$. In terms of coordinates $(x, y, \theta)$ we have therefore

$$\begin{cases} p = (x, y), \\ e_1 = (\cos\theta, \sin\theta), \\ e_2 = (-\sin\theta, \cos\theta). \end{cases} \qquad (99)$$

The moving frame $R = (p, r_\theta)$ is a *displacement* of the plane $M$, and then an element of $G$, acting on a point $a$ of $M$ by the formula

$$R(a) = (p, r_\theta)(a) = p + r_\theta(a), \qquad (100)$$

where $r_\theta(a) = r_\theta(Oa)$. If $R = (p, r_\theta)$ and $S = (q, r_\varphi)$ are two displacements, their composition is given by the *semi-direct product*

$$S \circ R = (q, r_\varphi) \circ (p, r_\theta) = (q + r_\varphi(p), r_{\theta+\varphi}). \qquad (101)$$

This non-commutative group law can easily be expressed as a matrix multiplication. Let us consider the vector space $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R}^2 = \mathbb{R} \times \mathbb{C}$ and the complex forms $p = x + iy$, $w = u + iv$, $r_\theta =$ (complex multiplication by $e^{i\theta}$). [24] It is trivial to verify that the displacement $R = (p, r_\theta)$ of $M$ can be identified with the restriction to vectors $\begin{pmatrix} 1 \\ w \end{pmatrix}$ of the linear endomorphism of $\mathbb{R} \times \mathbb{C}$ whose matrix is $g = \begin{pmatrix} 1 & 0 \\ p & e^{i\theta} \end{pmatrix}$, [25] the semi-direct law becoming simply matrix multiplication.

It is also easy to verify that the inverse $g^{-1}$ of $g$ is given by the formula

$$g^{-1} = \begin{pmatrix} 1 & 0 \\ -pe^{-i\theta} & e^{-i\theta} \end{pmatrix}. \qquad (102)$$

In other words, if we interpret $\mathbb{R}^2$ as the affine plane of the vector space $\mathbb{R}^3 = \mathbb{R} \times \mathbb{C}$ which is parallel to the base $\mathbb{C}$ at height 1, the non-commutativity of the semi-direct product $G = E(2) = SO(2) \ltimes \mathbb{R}^2$ becomes simply induced by that of the rotation group $SO(3)$.

We will therefore identify $G$ to the group of matrices $g$ and $R_0 = (0, r_0)$ to the identity element $e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The stabilizer of $O$ (i.e. the set of $g$ leaving the origin $O$ invariant) is the subgroup $H = SO(2)$ of matrices $g$ with $p = 0$ (pure plane rotations without translation), i.e. of the form $g = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$. The quotient $G/H$ is isomorphic to $\mathbb{R}^2$ (i.e. the base space $M$) and $G$ is the semi-direct product $H \ltimes (G/H)$. As $H = SO(2)$, we recover a fibration $\rho: G \to M = G/H$ having as fiber a group operating on the unit circle $\mathbb{S}^1$ of orientations. $\rho$ is called a *principal* bundle (on the plane $M$). Above every point $p$ of $M$, there is an exemplar of the rotation group $SO(2)$. The fibration $\rho$ *acts* on the fibration $\pi: V = M \times \mathbb{S}^1 \to M$ having as fiber the orientations of the plane: if $p \in M$, the exemplar of $SO(2)$ above $p$ acts on the fiber $V_p = \pi^{-1}(p)$ by rotating the direction. One says that the fiber bundle $\pi$ is *associated* with the principal bundle $\rho$.

### 8.2. The Lie algebra 𝔤 of G and the adjoint and co-adjoint representations

A Lie group $G$ is by definition a differentiable manifold endowed with a group structure whose operations are differentiable maps. There exists subtle interactions between the algebraic and differentiable structures. The group law of $G$ admits an infinitesimal version, which endows the tangent plane $T_e G$ of $G$ at the unit element $e$ with a structure of *Lie algebra* 𝔤. Moreover, $G$ acts on itself according to its own group law, and the infinitesimal version of this action provides a natural representation of $G$ on 𝔤, called the adjoint representation. By duality, one gets the Lie co-algebra 𝔤* of $G$ and a natural representation of $G$ on 𝔤*, called the co-adjoint representation. We already investigated an example in Section 4.2.5.

These remarkable properties are due to the fact that any Lie group $G$ is "homogeneous"—identical at every point—because it acts on itself by left and right translations. Let $g \in G$; one associates to $g$ the left translation

---

[23] We suppose that frames share a positive orientation.

[24] Complex multiplication by $e^{i\theta}$ in $\mathbb{C}$ corresponds in $\mathbb{R}^2$ to the rotation matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. The link with de Moivre formula $e^{i\theta} = \cos\theta + i\sin\theta$ is done through the standard identification of i with the Pauli matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (whose square is $-1$).

[25] $g$ is a $3 \times 3$ real matrix, $p$ being the translation vector $\begin{pmatrix} x \\ y \end{pmatrix}$ and $e^{i\theta}$ the rotation matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

$L_g$ in $G$ defined by $L_g : h \mapsto gh$. [26] $L_g$ is simply a change of frame, $h$ being considered no longer in the fixed frame $e$ but in the moving frame $g$. If $g = (p, r_\theta)$ and $h = (q, r_\varphi)$, we have

$$L_g(h) = (p + r_\theta(q), r_{\theta+\varphi}) = \begin{pmatrix} 1 & 0 \\ p + e^{i\theta}q & e^{i(\theta+\varphi)} \end{pmatrix}. \quad (103)$$

Left translations are diffeomorphisms of $G$, but they are not automorphisms of its group structure since they do not preserve the identity $e$. They are nevertheless compatible with the group law since their composition satisfies $L_g \circ L_f = L_{gf}$. They provide a *global canonical trivialization* of the tangent bundle $TG$. This bundle is not given as such as the direct product $G \times T$ with $T = T_e G$, but it can be canonically identified with $G \times T$ by means of the $L_g$. Many of the geometric properties of Lie groups proceed from this fact. In particular, by translating everywhere in $G$ a frame of $\mathfrak{g} = T_e G$ we get a global $G$-invariant frame of $G$. One says that $G$ is a *parallelizable* manifold.

The translations in $G$ are essential. $G$ is not a flat space since its $G$-invariant metric presents curvature. There exists therefore what is called a problem of *holonomy*. If we start from an orthonormal basis of $\mathfrak{g} = T_e G$ associated with local coordinates at $e$, the bases of the neighboring tangent planes $T_g G$ associated to the same local coordinates will no longer be orthonormal, and, reciprocally, a local field of orthonormal bases cannot derive from a system of local coordinates. But translations allow the construction of such fields.

## 8.3. Elie Cartan's formalism

Let $g = \begin{pmatrix} 1 & 0 \\ p & e^{i\theta} \end{pmatrix}$ be the current element of $G$. We consider its differential $dg = \begin{pmatrix} 0 & 0 \\ dp & ie^{i\theta}d\theta \end{pmatrix}$ and we interpret it as *a 1-form on $G$ with values in* $\mathfrak{g}$, that is as an element $dg \in T^*G \otimes \mathfrak{g}$. This means that the components of $dg$ are 1-forms on $G$, but that the values $dg(h)(\varsigma_h)$ of $dg$ at a point $h$ of $G$ on the tangent vector $\varsigma_h \in T_h G$ to $G$ at $h$ shares the type of an element of $\mathfrak{g}$. It is easy to show that, for $h = e$, $dg(e)$ is the *identity* of $\mathfrak{g} = T_e G$. $dg(e)$ is a 1-form on $T_e G = \mathfrak{g}$ which takes as inputs vectors $\varsigma \in \mathfrak{g}$ and, as it is $\mathfrak{g}$-valued, outputs also vectors $dg(e)(\varsigma) \in \mathfrak{g}$. Identity means simply that $dg(e)(\varsigma) = \varsigma$.

The fundamental idea of Elie Cartan was to start from the natural $\mathfrak{g}$-valued 1-form $dg$ on $G$ and to make it *G-invariant* under the action of the left translations $L_g$. Indeed, $dg$ is not $G$-invariant because, as we have seen, $TG$ is not given as such as the direct product

$G \times T$. $dg$ is the identity map of $TG$ but not of $TG$ *as globally trivialized by the left translations $L_g$*. A $G$-invariant 1-form on $G$ must be *constant* in the bases of $T_g^* G$ dual to the $G$-invariant bases of the tangent spaces $T_g G$. But it is easy to see that it is not the case for $dg$.

Cartan's idea was then to translate $dg(e)$ in order to obtain a 1-form on $G$ which would be $G$-invariant *by construction*. Let $\Lambda_G : TG \to \mathfrak{g}$ be this 1-form. It is called the *Maurer–Cartan form* of $G$ and can be interpreted geometrically very easily. By definition, $\Lambda_G(g) = (T_g L_{g^{-1}})^* dg(e)$. [27] If $\varsigma \in T_g G$ is a tangent vector to $G$ at $g$, $\Lambda_G(g)(\varsigma) = T_g L_{g^{-1}}(\varsigma)$ and $\Lambda_G$ transports $\varsigma$ in $\mathfrak{g}$ by means of the global trivialization provided by the left translations $L_g$. In our case, we verify promptly that we have: $\Lambda_G(g) = \begin{pmatrix} 0 & 0 \\ e^{-i\theta}dp & id\theta \end{pmatrix}$. Traditionally, $\Lambda_G$ is written

$$\Lambda_G = g^{-1} dg, \quad (104)$$

where $g^{-1}$ symbolizes $(T_g L_{g^{-1}})^*$.

The Lie algebra structure of $\mathfrak{g}$ can be easily recovered from the Maurer–Cartan form $\Lambda_G$ through a *universal formula*. The idea is to compute the exterior derivative $d\Lambda_G$ of $\Lambda_G$, which is a $G$-invariant $\mathfrak{g}$-valued 2-form on $G$. One shows that

$$d\Lambda_G = -\frac{1}{2}[\Lambda_G, \Lambda_G]. \quad (105)$$

These universal Maurer–Cartan equations encode the geometry of every Lie group (see the example of Section 4.2.5).

## 8.4. Variational problems on Lie groups according to Bryant and Griffiths

The Euclidean group of displacements $E(2) = G$, with its principal bundle structure, its adjoint and co-adjoint representations, and its Maurer–Cartan form is universal and geometrizes a neurally implemented functional architecture. It is therefore natural to formulate our variational models of modal illusory contours in this deeper framework.

In an exciting paper "Reduction for constrained variational problems and $\int \frac{\kappa^2}{2} ds$", Bryant and Griffiths [8] have developed this new approach of variational problems and applied it to David Mumford's theory of elastica.

Let us return to the curves $\gamma$ in the base plane $M$. If, when the point $p = (x, y)$ wanders along $\gamma$, we track the

---

[26] One must not confuse the translation $L_g$ in $G$ and the component $p$ of $g = (p, r_\theta)$ which is a translation in $M$.

[27] We have $T_g L_{g^{-1}} = (T_e L_g)^{-1} : T_g G \to T_e G$ and, by duality, $(T_g L_{g^{-1}})^* : T_e^* G = \mathfrak{g}^* \to T_g^* G$.

Frénet moving frame $R = (p, e_1, e_2) = (p, r_\theta)$ (where $(e_1, e_2)$ are the tangent and normal unitary vectors of $\gamma$ at $p$), we get a curve $\tilde{\gamma}$ which *lifts* $\gamma$ in $G$ and is called its *Frénet lifting*.

As $G$ is a principal bundle to which the contact bundle of $M$ is associated, we can interpret in this new framework the Legendrian lifting of $\gamma$ in $J^1M$ studied in Section 4.2.3. The problem is the same: we must characterize, but this time *in $G$*, the skew curves $\Gamma$ which are Frénet liftings $\tilde{\gamma}$ of curves $\gamma$ *in $M$*. The idea is to express an infinitesimal displacement $dR = (dp, de_1, de_2)$ of the moving frame $R = (p, e_1, e_2)$ in two different ways: the first will be general and universal, and associated directly to $G$, the second will be more particular and associated to $\gamma$. Comparing the two, we can then show that $\Gamma$ must satisfy an *integrability condition* which, as the Frobenius integrability condition of Section 4.2.3, is expressed by the vanishing of a system of differential 1-forms on an appropriate space $X$ derived from $G$, what is called a *Pfaff system*.

We can then select certain curves $\gamma$ in $M$ by means of a *variational principle applied to their Frénet lifting*. For this purpose we have to introduce a Lagrangian $L$ and look at the 1-form on $X$: $\varphi = L \, dt$. If $\Gamma : I \to X$ ($I = [0, 1]$) is a curve in $X$, we associate to it the "energy" $\Phi$

$$\Phi(\Gamma) = \int_I \Gamma^* \varphi \qquad (106)$$

(where the 1-form $\Gamma^* \varphi$ on $I$ is the inverse image by $\Gamma$ of the 1-form $\varphi$ on $X$), and we look for curves $\Gamma$ minimizing $\Phi$. For that, we have to solve anew constrained E–L equations. In our case, the Lagrangian 1-form $\varphi$ is

$$\varphi = (dp^2 + (\tan^2 \theta)' d\theta^2)^{\frac{1}{2}} \qquad (107)$$

The computation (rather complex) returns the results of Section 6.6.4.

We investigate that way 3 different levels of structure linking the universal geometry of $G$ to particular families of curves in $M$.

1. The Lie group $G$ and its associated structures: $\mathfrak{g}$, $\mathfrak{g}^*$, $\Lambda_G$, as well as the adjoint and coadjoint representations. This first level concerns a universal geometric framework implemented in a functional architecture.
2. Paths $\tilde{\gamma}$ in $G$ which are Frénet liftings of curves $\gamma$ in the base $M$. This second level concerns the coding of particular stimuli in this universal framework.
3. Finally, for the modal illusory contours, those curves $\Gamma$ which are solutions of a variational problem. This third level concerns the variational interpretation of the completion problem.

## 9. Conclusion

We have shown how experimental results on pinwheels and lateral cortico-cortical connections of V1 allow to develop a detailed neurogeometrical model of their functional architecture as a discrete approximation of the contact structure of the fibration $\pi : R \times P \to P$. We have also shown how this contact structure allows to understand typical Gestalt phenomena such as good continuation (association field) and modal illusory contours.

This neurogeometrical model is as elementary as possible. It would have to be considerably complexified in order to take into account other retinotopic and non-retinotopic areas with their feedback top-down projections on V1. But, as elementary as it may be, it already makes use of non-trivial mathematical concepts such as blowing-up, contact structures, Frobenius integrability condition, sub-Riemannian geometry, Carnot-Carathéodory metrics, Euler–Lagrange equations, Lie groups, Maurer–Cartan form, etc. It highlights a fundamental fact. The geometrical formatting of the optical signal requires an integration of local data into global structures. But how such an integration can be performed via fields of neural point processors processing in parallel only point measures (of course at a certain scale of resolution)? The solution which was selected by evolution and theoretically conceived of by the main post-Riemannian geometers is to introduce supplementary local degrees of freedom (orientations $p$ for V1) with the constraint that they have to be interpreted as differential entities ($p = a'$ for V1).

The existence of such a functional architecture explains how a neural calculus is able to solve non-trivial integration problems. A synchronized wave of activity propagating in the functional architecture is equivalent to the integration of a specific differential equation. It is in that sense that computation is the hardware.

We see that an idea emerges, which brings neurosciences close to physical sciences. It is not a new idea since there exists already the case of the Hopfield neural networks. Spin glass models coming from statistical physics of magnetic media are relevant for the connectionist perspective since they express in a general way the dynamical macroproperties (global attractors, bifurcations, etc.) of interacting systems of elementary micro-units coupled via positive and negative couplings. We meet here the same situation. Modern fundamental physics rests on formalisms of the type we have used here. A physical field is a section of a fiber bundle with the space–time as base space and as fiber an algebraic type (scalar, vector and spinor) on which acts the group $G$ of internal symmetries. In gauge theories, interactions between particles are described by connections on these fiber bundles. The origin of the necessity of such formalisms is the same as in our case: to understand the

physical origin of space as a global framework for phenomena, in other words, to explain the material genesis of the ideality of space.

## References

[1] D.M. Alexander, P. Sheridan, P.D. Bourke, O. Konstandatos, J.J. Wright, Global and local symmetry of the primary visual cortex: derivation of orientation preference, Available from <http://www.mhri.edu.au/~dma/>.

[2] A. Bellaïche, J.-J. Risler (Eds.), Sub-Riemannian Geometry, Progress in Mathematics, 144, Birkhäuser, Basel, 1996.

[3] A. Berthoz, Le sens du mouvement, Odile Jacob, Paris, 1997.

[4] T. Bonhöffer, A. Grinvald, Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns, Nature 353 (1991) 429–431.

[5] W. Bosking, Y. Zhang, B. Schofield, D. Fitzpatrick, Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex, J. Neurosci. 17 (6) (1997) 2112–2127.

[6] M. Brady, W.E.L. Grimson, D. Langridge, Shape encoding and subjective contours, Proc. AAAI, Stanford Univer. (1980) 15–17.

[7] P. Bressloff, J. Cowan, M. Golubitsky, P. Thomas, M. Wiener, Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex, Philos. Trans. R. Soc. London B 356 (2001) 299–330.

[8] R. Bryant, P. Griffiths, Reduction for constrained variational problems and $\int \frac{\kappa^2}{2} \, \mathrm{d}s$, Am. J. Math. 108 (1986) 525–570.

[9] P. Buser, M. Imbert, Vision, Hermann, Paris, 1987.

[10] M.C. Crair, E.S. Ruthazer, D.C. Gillepsie, M.P. Stryker, Ocular dominance peaks at pinwheels center singularities of the orientation map in cat visual cortex, J. Neurophysiol. 77 (1997) 3381–3385.

[11] A. Das, C.D. Gilbert, Long range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex, Nature 375 (1995) 780–784.

[12] G.C. De Angelis, I. Ozhawa, R.D. Freeman, Receptive-field dynamics in the central visual pathways, Trends Neurosci. 18 (10) (1995) 451–458.

[13] G.C. De Angelis, G.M. Ghose, I. Ohzawa, R.D. Freeman, Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons, J. Neurosci. 19 (9) (1999) 4046–4064.

[14] B. Dresp, C. Bonnet, Subthreshold summation with illusory contours, Vision Res. 35 (1995) 1071–1078.

[15] G.B. Ermentrout, J.D. Cowan, A mathematical theory of visual hallucinations, Kybernetic 34 (1979) 137–150.

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

[17] L.M.J. Florack, The syntactical structure of scalar images, Ph.D., University of Utrecht, 1993.

[18] L.M.J. Florack, B.M. Ter Haar Romeny, J.J. Koenderink, M.A. ViergeVer, Scale and the differential structure of images, Image Vision Comput. 10 (6) (1992) 376–388.

[19] Y. Frégnac, V. Bringuier, F. Chavane, L. Glaeser, J. Lorenceau, An intracellular study of space and time representation in primary visual cortical receptive fields, J. Physiol. 90 (1996) 189–197.

[20] Y. Frégnac, D. Shulz, Activity-dependent regulation of receptive field properties of cat area 17 by supervised Hebbian learning, J. Neurobiol. 41 (1) (1999) 69–82.

[21] S. Georges, P. Seriès, Y. Frégnac, J. Lorenceau, Orientation-dependent modulation of apparent speed: psychophysical evidence, Vision Res. 42 (2002) 2557–2572.

[22] C.D. Gilbert, A. Das, M. Ito, M. Kapadia, G. Westheimer, Spatial integration and cortical dynamics, Proc. Natl. Acad. Sci. USA 93 (1996) 615–622.

[23] M. Gromov, Carnot-Carathéodory spaces seen from within, in: A. Bellaïche, J.-J. Risler (Eds.), Sub-Riemannian Geometry, Progress in Mathematics, 144, Birkhäuser, Basel, 1996, pp. 79–323.

[24] M. Gromov, In: Metric Structures for Riemannian and Non-Riemannian Spaces, Progress in Mathematics, 152, Birkhäuser, Boston, 1999.

[25] S. Grossberg, E. Mingolla, Neural dynamics of form perception: boundary completion, illusory figures and neon color spreading, Psychol. Rev. 92 (1985) 173–211.

[26] H. Hamy, Méthodes géométriques multi-échelle en vision computationnelle, Ph.D., École Polytechnique, Paris, 1997.

[27] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, O. Kübler, Simulation of neural contour mechanisms: from simple to end-stopped cells, Vision Res. 32 (5) (1992) 963–981.

[28] F. Heitger, R. von der Heydt, A computational model of neural contour processing: figure-ground segregation and illusory contours, in: Proceedings of the 4th International Conference on Computer Vision, IEEE, New York, 1993, pp. 32–40.

[29] W.C. Hoffman, Some reasons why algebraic topology is important in neuropsychology: perceptual and cognitive systems as fibrations, Int. J. Man-Machine Studies 22 (1985) 613–650.

[30] W.C. Hoffman, The visual cortex is a contact bundle, Appl. Math. Computat. 32 (1989) 137–167.

[31] B.K.P. Horn, The curves of least energy, ACM Trans. Math. Software 9 (4) (1983) 441–460.

[32] D.H. Hubel, Eye, Brain and Vision, Scientific American Library, 1988.

[33] M. Hübener, D. Shoham, A. Grinvald, T. Bonhöffer, Spatial relationships among three columnar systems in cat area 17, J. Neurosci. 17 (1997) 9270–9284.

[34] G. Kanizsa, Organization in Vision: Essays on Visual Perception, Praeger, New York, 1979.

[35] M.K. Kapadia, M. Ito, C.D. Gilbert, G. Westheimer, Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys, Neuron 15 (1995) 843–856.

[36] M.K. Kapadia, G. Westheimer, C.D. Gilbert, Dynamics of spatial summation in primary visual cortex of alert monkey, Proc. Natl. Acad. Sci. 96 (21) (1999) 12073–12078.

[37] P.J. Kellman, T.F. Shipley, A theory of visual interpolation in object perception, Cognitive Psychol. 23 (1991) 141–221.

[38] J.J. Koenderink, Simultaneous order in nervous nets from a functional standpoint, Biol. Cybernet. 50 (1984) 35–41.

[39] J.J. Koenderink, A.J. van Doorn, Representation of local geometry in the visual system, Biol. Cybernet. 55 (1987) 367–375.

[40] J.J. Koenderink, Operational significance of receptive field assemblies, Biol. Cybernet. 58 (1988) 163–171.

[41] V.A.F. Lamme, H. Super, H. Speckreijse, Feedforward, horizontal and feedback processing in the visual cortex, Curr. Opin. Neurobiol. 8 (1998) 529–535.

[42] T.S. Lee, D. Mumford, R. Romero, V.A.F. Lamme, The role of primary visual cortex in higher level vision, Vision Res. 38 (1998) 2429–2454.

[43] T.S. Lee, M. Nguyen, Dynamics of subjective contour formation in the early visual cortex, Proc. Natl. Acad. Sci. 98 (4) (2001) 1907–1911.

[44] G. Longo, Mathematical intelligence, infinity and machines: beyond the Gödelitis, J. Consciousness Studies (Special issue on Cognition) 6 (11–12) (1999).

[45] S. MacEvoy, W. Kim, M. Paradiso, Integration of surface information in primary visual cortex, Nature Neurosci. 1 (1998) 616–620.

[46] M. Maffei, A. Fiorentini, The unresponsive regions of visual cortical receptive fields, Vision Res. 16 (1976) 1131–1139.

[47] P.E. Maldonado, I. Gödecke, C.M. Gray, T. Bonhöffer, Orientation selectivity in pinwheel centers in cat striate cortex, Science 276 (1997) 1551–1555.

[48] S.G. Mallat, Multifrequency channel decompositions of images and wavelet models, IEEE Trans. Acoust. Speech Signal Process. 37 (12) (1989) 2091–2110.

[49] D. Marr, Vision, Freeman, New York, 1982.

[50] D. Mumford, Elastica and computer vision, in: C. Bajaj (Ed.), Algebraic Geometry and Applications, Springer Verlag, Heidelberg, 1992.

[51] D. Mumford, Pattern theory: a unifying perspective, in: D.C. Knill, W. Richards (Eds.), Perception as Bayesian Inference, Cambridge University Press, 1996, pp. 25–62.

[52] J. Ninio, L'Empreinte des Sens, Odile Jacob, Paris, 1996.

[53] M. Nitzberg, D. Mumford, T. Shiota, In: Filtering Segmentation and Depth, Lecture Notes in Computer Science, 662, Springer Verlag, Berlin, 1993.

[54] P. Pansu, Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un, Ann. Math. 129 (1) (1989) 1–60.

[55] P. Parent, S.W. Zucker, Trace inference, curvature consistency, and curve detection, IEEE Trans. Pattern Anal. Mach. Intell. II 8 (1989) 823–839.

[56] S. Pauls, Minimal surfaces in the Heisenberg group (forthcoming).

[57] E. Peterhans, R. von der Heydt, Subjective contours: bridging the gap between psychophysics and psychology, Trends Neurosci. 14 (3) (1991) 112–119.

[58] J. Petitot, in collaboration with Y. Tondut, Vers une Neurogéométrie. Fibrations corticales, structures de contact et contours subjectifs modaux, Mathématiques, Informatique et Sciences Humaines, 145, EHESS, CAMS, Paris, 1999, pp. 5–101.

[59] J. Petitot, Morphological eidetics for phenomenology of perception, in: J. Petitot, F.J. Varela, J.-M. Roy, B. Pachoud (Eds.), Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science, Stanford University Press, Stanford, pp. 330–371.

[60] U. Polat, D. Sagi, Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiment, Vision Res. 33 (7) (1993) 993–999.

[61] A. Rossi, M. Paradiso, Neural correlates of brightness in the responses of neurons in the retina, LGN, and primary visual cortex, J. Neurosci. 19 (1999) 6145–6156.

[62] B.R. Sheth, J. Sharma, C. Rao, M. Sur, Orientation maps of subjective contours in visual cortex, Nature 274 (1996) 2110–2115.

[63] H.P. Snippe, J.J. Koenderink, Discrimination thresholds for channel-coded systems, Biol. Cybernet. 66 (1992) 543–551.

[64] L. Spillman, B. Dresp, Phenomena of illusory form: can we bridge the gap between levels of explanation?, Perception 24 (1995) 1333–1364.

[65] M. Spivak, A Comprehensive Treatise on Differential Geometry, Publish or Perish, Boston, MA, 1974.

[66] N.V. Swindale, The development of topography in the visual cortex: a review of models, Network: Comput. Neur. Syst. 7 (2) (1996) 161–247.

[67] A. Toet, J. Blom, J.J. Koenderink, The construction of a simultaneous functional order in nervous systems, Biol. Cybernet. 57 (1987) 115–125.

[68] Y. Tondut, J. Petitot, Géométrie de contact et champ d'association dans le cortex visuel, CREA reports # 9725, Ecole Polytechnique, Paris, 1997.

[69] D. Ts'o, C.D. Gilbert, T.N. Wiesel, Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis, J. Neurosc. 6 (4) (1986) 1160–1170.

[70] S. Ullman, Filling in the gaps: the shape of subjective contours and a model for their generation, Biol. Cybernet. 25 (1976) 1–6.

[71] R. Van Rullen, S.J. Thorpe, The time course of visual processing: from early perception to decision making, J. Cognitive Neurosc. 13 (4) (2001) 454–461.

[72] R. von der Heydt, E. Peterhans, Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity, J. Neurosci. 9 (5) (1989) 1731–1748.

[73] R. von der Heydt, E. Peterhans, Cortical contour mechanisms and geometrical illusions, in: D.M. Lam, C.D. Gilbert (Eds.), Neural Mechanisms of Visual Perception, Portfolio Publishing Company, 1989, pp. 158–170.

[74] C. von der Malsburg, Self-organization of orientation-sensitive cells in the striate cortex, Kybernetik 14 (1973) 85–100.

[75] J.A. Webb, E. Pervin, The shape of subjective contours, Proc. IAAA (1984) 340–343.

[76] M. Weliki, W. Bosking, D. Fitzpatrick, A systematic map of direction preference in primary visual cortex, Nature 379 (1996) 725–728.

[77] L.R. Williams, D.W. Jacobs, Stochastic completion fields: a neural model of illusory contour shape and salience, Proc. Int. Conf. Comp. Vis. 5 (1995) 408–415.

[78] F. Wolf, T. Geisel, Spontaneous pinwheel annihilation during visual development, Nature 395 (1998) 73–78.

[79] S. Zeki, A Vision of the Brain, Blackwell Scientific Publications, Oxford, 1993.

[80] J. Zhang, S. Wu, Structure of visual perception, Proc. Natl. Acad. Sci. USA 87 (1990) 7819–7823.

[81] S.W. Zucker, C. David, A. Dobbins, L. Iverson, The organization of curve detection: coarse tangent fields and fine spline covering, in: Proceedings of the Second International Conference on Computer Vision, IEEE, New York, 1988.

[82] P.C. Bressloff, J.D. Cowan, The functional geometry of local and horizontal connections in a model of V1, Journal of Physiology - Paris 97 (2003) 221–236.