



From receptive profiles to a metric model of V1

Noemi Montobbio^{1,2} · Giovanna Citti¹ · Alessandro Sarti²

Received: 26 March 2018 / Revised: 23 February 2019 / Accepted: 28 March 2019 / Published online: 12 April 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

In this work we show how to construct connectivity kernels induced by the receptive profiles of simple cells of the primary visual cortex (V1). These kernels are directly defined by the shape of such profiles: this provides a metric model for the functional architecture of V1, whose global geometry is determined by the reciprocal interactions between local elements. Our construction adapts to any bank of filters chosen to represent a set of receptive profiles, since it does not require any structure on the parameterization of the family. The connectivity kernel that we define carries a geometrical structure consistent with the well-known properties of long-range horizontal connections in V1, and it is compatible with the perceptual rules synthesized by the concept of association field. These characteristics are still present when the kernel is constructed from a bank of filters arising from an unsupervised learning algorithm.

Keywords Visual cortex · Connectivity kernel · Functional architecture · Metric space · Association field · Neurogeometry

1 Introduction

The primary visual cortex (V1) implements the first stage of cortical processing of the visual information, and it is the most studied and the best understood among the visual areas in the brain. An essential concept related to the structure of V1 is that of *receptive profile* (RP), i.e. the function modeling the impulse response of a neuron. Indeed, the primary step of computation taking place in V1 involves a class of neurons, called *simple cells*, which act approximately as linear filters on the retinal image. In other words, the action of such a neuron on a visual stimulus can be expressed, to a first approximation, as the convolution between the image and the RP of the cell.

A classical model for the set of receptive profiles of V1 simple cells is provided by a bank of bidimensional Gabor filters (Jones and Palmer 1987; Daugman 1985; Lee 1996). Such a family can be indexed by a set of parameters showing a group structure: this fact is at the basis of a number of models, which identify the set of simple cells with such space of parameters and define on it a differential structure, designed to be compatible with the above-mentioned group law (Petitot and Tondut 1999; Citti and Sarti 2006; Sarti et al. 2008). This idea can in principle be replicated as long as the bank of filters modeling the RPs can be parameterized by a group. Yet, this condition is not verified if one considers for example a bank of filters learned through automatic learning procedures (see e.g. Olshausen and Field (1996) and Anselmi and Poggio (2010)). It is nonetheless crucial for a cortex model to describe the *functional architecture* ruling the geometry of intra-cortical connections: these are indeed thought to be responsible for the integration of local features into contours, shapes, objects. Specifically, the contextual influences modulating V1 neurons' activity have been described through the concept of *association field* (Field et al. 1993), which specifies the geometrical properties of the influences between neurons even with markedly separated RPs, based on their reciprocal position and orientation. The long-range, orientation specific horizontal connections which are known to take place in V1 are believed to be at the basis of such mechanisms of perceptual grouping (Bosking et al. 1997).

Action Editor: B. A. Olshausen

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10827-019-00716-6>) contains supplementary material, which is available to authorized users.

✉ Noemi Montobbio
noemi.montobbio2@unibo.it

¹ Department of Mathematics, Università di Bologna, Bologna, Italy

² CAMS Center of Mathematical Analysis (CNRS - EHESS), Paris, France

The aim of this paper is to show that a bank of receptive profiles can induce such a functional architecture, with or without the presence of a group structure.

We propose to express the connectivity strength through a kernel defined onto the family of RPs and generated by the RPs themselves. Specifically, we consider a bank of filters $\{\psi_p\}_{p \in \mathcal{G}} \subseteq L^2(\mathbb{R}^2)$, where \mathcal{G} is any set of parameters indexing the family. As in Citti and Sarti (2006), we identify each filter ψ_p with the corresponding index $p \in \mathcal{G}$. We shall refer to \mathcal{G} as the *feature space*: intuitively, one may think that each element $p \in \mathcal{G}$ encodes the features extracted by the corresponding filter ψ_p when it is applied to some image. We then define, for each couple of points $p, p_0 \in \mathcal{G}$, the *generating kernel*

$$K(p, p_0) := \operatorname{Re} \left(\int_{\mathbb{R}^2} \psi_p(x, y) \overline{\psi_{p_0}(x, y)} dx dy \right),$$

i.e. $K(p, p_0)$ is defined to be the real part of the L^2 scalar product between the filters ψ_p and ψ_{p_0} . This kernel allows to define a natural metric structure on the feature space. We then construct, through an iterative mechanism of repeated integrations against K , a wider connectivity kernel from which we are also able to recover association-field-like patterns. On the other hand, in differential cortex models the spreading of neural activity across V1 is typically described by means of diffusion equations expressed through second order operators associated to the metric. In our work (Montobbio et al. 2017), we consider a diffusion process defined in terms of the distance, and we study its relation with the iterative procedure described in this paper. The technique introduced in Montobbio et al. (2017) allows to extend the diffusion-based approach used in differential models to the present metric setting, including possibly non-differential metrics, while still incorporating the sub-Riemannian model introduced in Citti and Sarti (2006) as a limit case.

In the second part of the paper, we present some results and numerical simulations to demonstrate the applicability of our model to both the classical framework of a family of Gabor filters (including a spatiotemporal example), and the situation where the RPs of simple cells are represented by a bank of filters obtained through an unsupervised learning algorithm. The application of this construction to *endstopped* RPs also allowed to recover the property of curvature selectivity observed (Dobbins et al. 1987) in these neurons. Our results suggest that *the geometry of the horizontal connections in V1 can indeed be obtained from the shape of the receptive profiles*.

The promising outcomes obtained even in the case of a non-structured bank of learned filters led us to the idea of applying our construction in the context of convolutional neural networks (CNNs) (LeCun et al. 2015). We thus propose, as a future development, the application

of our model to insert lateral connections induced by the feedforward filters in the convolutional layers of a CNN.

2 The functional architecture of V1

In this section we first recall the basic notions of receptive field and receptive profile of a visual neuron, with a particular focus on V1. We then review the link between the mechanisms of perceptual grouping and the neural circuitry of horizontal connections in V1. Lastly, we briefly outline the main concepts underlying some existing cortex models, mainly based on differential structures.

2.1 Receptive fields and receptive profiles

Every subcortical or cortical neuron is linked, through the connectivity of retino-geniculo-cortical pathways, to a local domain D of the retina, called the *receptive field* (RF) of the cell. The *receptive profile* (RP) of a visual cell is a function $\psi(x, y)$ defined on D , measuring quantitatively the impulse response of the neuron.

Certain classes of cells act approximately as *linear filters* on the optic signal, meaning that their receptive profile provides, up to a first approximation, the necessary information to measure their reaction to a general stimulus (expressed as a function $I(x, y)$ of the retinal coordinates): this is indeed given by a *convolution*

$$\int_D I(x, y) \psi(x - x_0, y - y_0) dx dy.$$

Here, the RP is expressed in local coordinates around (x_0, y_0) , the retinal location around which it is concentrated: that is, the response of the cell to a punctual stimulus at (x_0, y_0) is given by $\psi(0, 0)$.

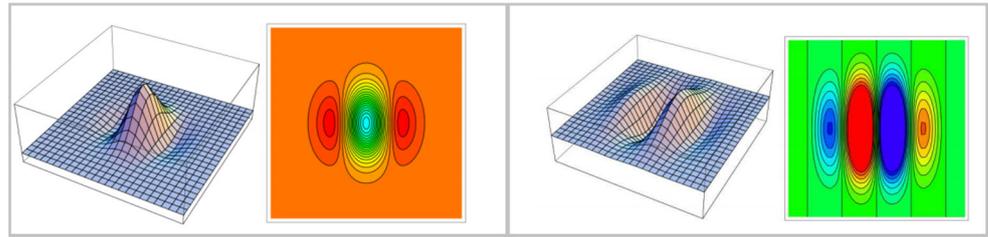
2.2 The structure of V1

Let us now focus on the structure of the first of the visual cortical areas. V1 neurons can be divided into two main classes, as first discovered by Hubel and Wiesel in 1962 (Hubel and Wiesel 1962): these are called *simple* and *complex* cells.

Simple cells are the first neurons in the visual pathway which are sensitive to orientation.

The basic structure of their RPs consists of two to five tapering rows of alternating excitatory and inhibitory subregions (Daugman 1985). Two examples of such profiles are shown in Fig. 1. As already mentioned, the responses of simple cells display orientation and positional selectivity: each of them fires at an optimal orientation, giving progressively weaker responses as the orientation of the stimulus shifts sub-optimally. This behavior is classically

Fig. 1 Two examples of RPs of simple cells of V1. Source: Petitot (2008)



described by modeling the RPs of this class of cells through a family of *Gabor filters*, defined as complex exponential functions modulated by bidimensional Gaussian functions (Jones and Palmer 1987; Daugman 1985; Lee 1996). The filters vary in retinal position, orientation and scale in order to describe the whole set of profiles. Starting from the basic filter

$$\psi_{0,0,0,1}(u, v) = \exp\left(\frac{2\pi i u}{\lambda}\right) \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right), \quad (1)$$

one can obtain the whole bank by translations, rotations and dilations. These are expressed by the following parameters: $(x, y) \in \mathbb{R}^2$ represents the position at which the filter is translated, $\theta \in S^1$ is the angle of the rotation and $\sigma > 0$ is the factor of dilation. Each combination of these transformations determines a filter $\psi_{x,y,\theta,\sigma}$ indexed by such parameters.

A different kind of processing is carried out by complex cells: unlike simple cells, their response cannot be explained through the linear action of a receptive profile on an image. These cells still display orientation sensitivity, but are *phase invariant*, i.e. they respond equally to optimally oriented bars independent of where they are placed inside the RF. A number of findings (see Martinez and Alonso (2003) for a review) led many authors to model the response of a complex cell as a square sum of simple cells with similar orientation and scale but with phases that differed in 90 degrees. Such a couple of cells is commonly referred to as a *quadrature pair*. This characterization of the development of phase invariance in complex cells is often called *energy model*.

However, the interactions between V1 neurons are not limited to the classical hierarchy of simple - to - complex cells.

Indeed, there exist two main types of intracortical connectivity affecting the responses of V1 neurons. Short-range connections, linking neurons whose RPs have the same retinal position but different orientations, act by selecting the orientation causing the maximal response to a stimulus and suppressing the others. On the other hand, lateral (or “horizontal”) connections associate neurons sharing the same preferred orientation, but linked to different positions on the retina. The qualitative and quantitative properties of the latter have been investigated in

several neurophysiological experiments, the best known of which are the ones performed by W. Bosking et al. in 1997 (Bosking et al. 1997) (see Fig. 2): these connections show a marked orientation specificity and spread for millimeters horizontally along the cortex, linking cells whose RF do not overlap. Such attributes make them a likely candidate to account for the processes of perceptual grouping carried out by our visual system.

2.3 Contextual influences and association fields

The human visual system is highly efficient in detecting and segregating objects. One of the main aspects in the perception of shapes is the identification of contours: this cannot be explained through the classical concept of RF (Grossberg and Mingolla 1985; Neumann and Mingolla 2001; Angelucci and Bullier 2003), which refers to the extraction of *local* image features from the visual scene. In other words, the activity of cells is not only influenced by inputs within their RF, but rather it strongly depends on the context.

As far as V1 is concerned, it has been proposed (Field, Hayes and Hess, 1993 (Field et al. 1993)) that the interaction of local edge elements in this area can be described through an *association field*, consisting of a region around each V1 cell where the activation of other neurons characterized

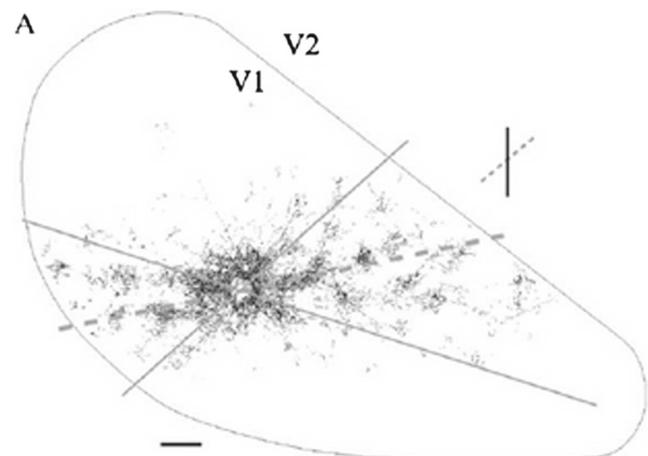


Fig. 2 The horizontal connections departing from a V1 neuron extend along an axis according to the preferential orientation of the cell. Source: Bosking et al. (1997)

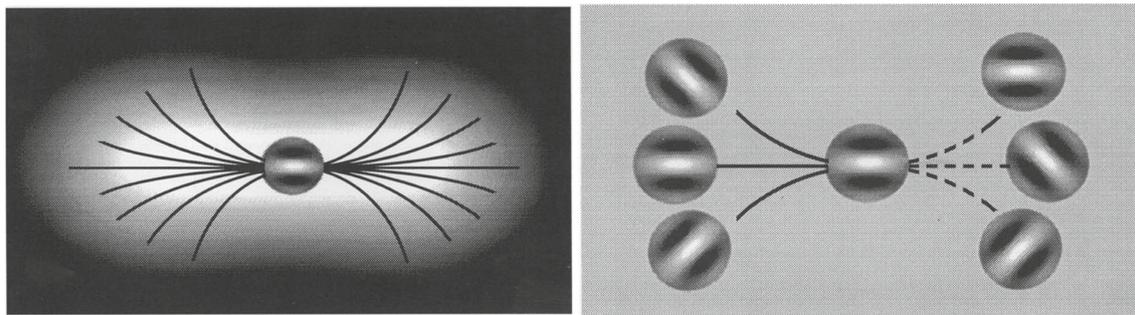


Fig. 3 A schematic representation of the concept of association field introduced in Field et al. (1993). Left: the “area of influence” of a cell and some of the possible paths departing from it. Right: examples of neurons for which grouping with the central cell occurs (solid curves) or does not occur (dashed curves)

by certain preferred orientations tends to be enhanced by the excitation of the central cell. See Fig. 3. A number of psychophysical experiments, described in Field et al. (1993), suggested that the area covered by this activation region around a neuron is considerably wider than the RF of the cell.

In V1, the interactions mediated by the association field show facilitatory influences for neurons whose RPs have a similar preferred orientation; however, there may exist an analogue entity in other cortical areas (Gilbert and Wu 2013), which links neurons sharing different and more abstract features. In other terms, each area could be endowed with an *intrinsic* correlation kernel, closely linked to the type of analysis performed in it, yielding corresponding perceptual patterns.

As just mentioned, the association field observed in V1 is believed to be intrinsic to this area: a possible anatomical framework is represented by the long-range horizontal connections (Petitot and Tondut 1999; Gilbert et al. 1996), as they fittingly link neurons with widely separated RPs whose reciprocal orientations are collinear or co-circular.

2.4 Previous models

In this section, we give an overview on some previous mathematical models describing the connectivity of the visual cortex and the laws of perceptual organization. Neuromathematical models of V1 often adopt a differential geometry approach to modeling the functional architecture of this area. A breakthrough idea has been that of representing V1 as a fiber bundle with the space of retinal locations as a basis. This approach was originally proposed by Petitot and Tondut (1999), with the definition of a left-invariant sub-Riemannian structure on the Heisenberg group. In 2006, Citti and Sarti wrote the model in the Lie group $SE(2) = \mathbb{R}^2 \times S^1$ by requiring the invariance under roto-translations (Citti and Sarti 2006). Such sub-Riemannian structure has been shown to be naturally

induced by the action of a bank of Gabor filters on a visual stimulus. In particular, consider a Gabor function

$$\Psi(x, y, \theta) = \exp\left(-\frac{\xi^2 + \nu^2}{\sigma^2}\right) \exp\left(i \frac{\nu}{\sigma^2}\right),$$

where

$$\begin{aligned} \xi &= x \cos \theta + y \sin \theta, \\ \nu &= -x \sin \theta + y \cos \theta. \end{aligned}$$

The outcome of filtering an image I with Ψ can be locally approximated as

$$O(x, y, \theta) = X_3 \exp\left(-\frac{\xi^2 + \nu^2}{\sigma^2}\right) * I =: X_3 I_\sigma, \tag{2}$$

where

$$X_3 = X_3(\theta) = -\sin \theta \partial_x + \cos \theta \partial_y. \tag{3}$$

I_σ is a smoothed version of I , obtained by convolving it with a Gaussian kernel.

This procedure *lifts* the 2D image domain to the 3D space $\mathbb{R}^2 \times S^1$: each point $(x, y) \in \mathbb{R}^2$ is sent to a point $(x, y, \bar{\theta}) \in \mathbb{R}^2 \times S^1$ such that $\bar{\theta}$ is a local maximum point of $\theta \mapsto O(x, y, \theta)$, so that the whole image domain is lifted to the set

$$\{(x, y, \bar{\theta}) : O(x, y, \bar{\theta}) = \max_{\theta} O(x, y, \theta)\} \subseteq \mathbb{R}^2 \times S^1.$$

This “non-maximal suppression” principle is based on experimental evidence on the sharp orientation tuning of V1 neurons (see also Bressloff and Cowan (2003) and Bressloff et al. (2002)). The selection of a direction X_3 defines at each point a subspace of dimension 2 of the Euclidean tangent space to $\mathbb{R}^2 \times S^1$, generated by

$$\begin{aligned} X_1 &= \cos \theta \partial_x + \sin \theta \partial_y, \\ X_2 &= \partial_\theta. \end{aligned}$$

The planes generated by X_1 and X_2 at each point are called *horizontal planes* and determine a bracket-generating

distribution on $\mathbb{R}^2 \times S^1$, thus defining a sub-Riemannian structure on it. In particular one has

$$[X_1, X_2] = -X_3, \tag{4}$$

so that the Lie algebra generated by X_1 and X_2 is the whole three-dimensional tangent space at every point. This property implies, thanks to the Chow Theorem (see e.g. Montgomery (2002)), that any two points (x_0, y_0, θ_0) and (x, y, θ) of $\mathbb{R}^2 \times S^1$ can be connected through an integral curve of $X_1 + kX_2$, where k varies in \mathbb{R} . Such curves are called *horizontal curves*.

Just as the anatomical mechanism of horizontal connections is proposed to implement the perceptual association fields, the connectivity defined by this sub-Riemannian structure allows to recover such a pattern: the family of horizontal curves starting at a point (x_0, y_0, θ_0) can be interpreted as the local association field spreading from (x_0, y_0, θ_0) , see Fig. 4.

The propagation of neural activity in this sub-Riemannian space can be modeled by means of a diffusion equation expressed through second order operators defined in terms of the vector fields X_1 and X_2 . A solid reference for these topics is the book (Bonfiglioli et al. 2007).

In particular, the heat equation associated to the sub-Riemannian Laplacian $\Delta u = X_1^2 u + X_2^2 u$ is considered. Figure 5 displays a level set of the fundamental solution to this operator.

The local invariance of the architecture of V1 w.r.t. $SE(2)$ has been exploited in a number of other works. In Barbieri et al. (2014), cortical orientation maps are modeled as minimizers of an uncertainty principle stated in this setting. An extension of Citti and Sarti (2006) in the context of spatio-temporal analysis of visual stimuli was developed in Barbieri et al. (2014). On the other hand, a semi-discrete variant of Citti and Sarti (2006) was proposed in Boscain et al. (2014), where only a finite number of

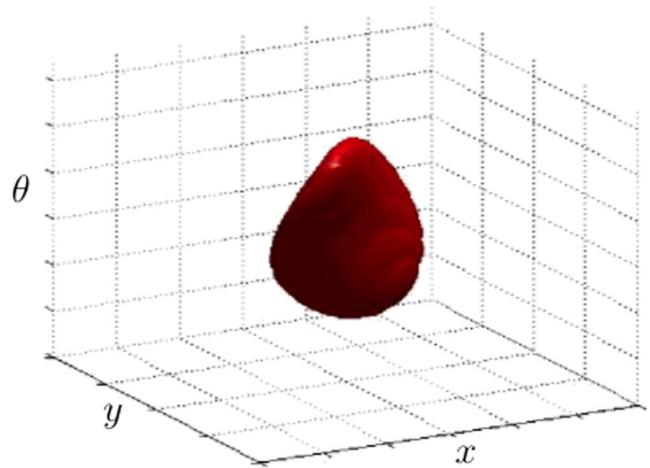


Fig. 5 A level set of the fundamental solution of the sub-Riemannian Laplacian

angles is taken into consideration. Moreover, a further study on the relation between association field curves and sub-Riemannian geodesics in $\mathbb{R}^2 \times S^1$ was carried out in Duits et al. (2014).

In Sanguinetti et al. (2010), a Fokker-Planck equation was considered instead of a sub-Riemannian heat equation, resulting from a study of statistical kernels of edge co-occurrence in natural images (see also Mumford (1993) and Sarti and Citti (2015)). This corresponds to assuming that the propagation has a deterministic component along X_1 and a stochastic component in direction X_2 .

Linear as well as non-linear diffusion equations in $\mathbb{R}^2 \times S^1$ were also employed in (Duits and Franken 2010a, b) for medical image processing; here, the lifting process of the image is based on a so-called orientation score, essentially providing a local orientation representation of an image through a generalized unitary wavelet transform (Duits et al. 2007). In Abbasi-Sureshjani et al. (2016), such a lifting was

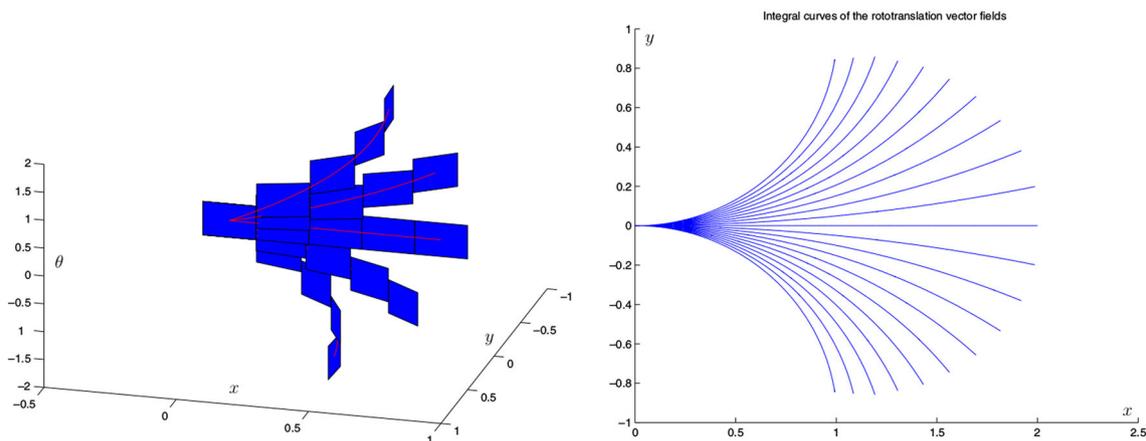


Fig. 4 Integral curves of $X_1 + kX_2$ for varying values of k . Left: the curves visualized in $\mathbb{R}^2 \times S^1$, with contact planes displayed. Right: their projection onto the image plane. Source: Citti and Sarti (2006)

complemented with the definition of a sub- Riemannian structure in a 5-dimensional space obtained by introducing the features of curvature and intensity, in order to overcome limitations given by crossing and bifurcations in a blood vessel tracking task. In Duits et al. (2013), the approach of evolution equations was extended to the range of Gabor transforms, again with applications to the processing of images.

Another model of V1 horizontal connections based on differential geometry has been proposed in Ben-Shahar and Zucker (2004), where a crucial role is again given to curvatures, providing a directional rate of change of orientation. In this work, the relationship between nearby tangents is analyzed for curves as well as textures. An analysis of other possible relationships between the lateral connectivity and visual function beyond contour integration had been carried out in Ben-Shahar et al. (2003) as well.

Moreover, a number of models have been proposed where perceptual grouping and texture segmentation tasks were addressed through a combination of long-range interactions and recurrent processing, involving the so-called bipole model (Grossberg and Mingolla 1985; Neumann and Mingolla 2001; Hansen and Neumann 2008).

In the next sections, we will propose a new model of the functional architecture of V1 that does not need the differential structures of most previous models. It is based on a metric structure on a suitable *feature space* indexing the RPs of simple cells. Such a geometry is induced by the RPs themselves. Interestingly, when the family of RPs is represented by a bank of Gabor filters indexed by $\mathbb{R}^2 \times S^1$, the induced distance is locally equivalent (Montobbio et al. 2017) to a Riemannian approximation to the sub-Riemannian structure described in Citti and Sarti (2006). Still, our model can be applied to a wide range of filter banks, with no need for any group structure, and is able to recover the perceptual rules of association fields even when applied to a randomly ordered family of learned orientation-selective filters.

3 A functional architecture generated by RPs

3.1 The generating kernel

We fix a bank of real- or complex-valued linear filters on the plane, modeling a set of RPs of simple cells of V1. Specifically, we consider a family $\{\psi_p\}_{p \in \mathcal{G}} \subseteq L^2(\mathbb{R}^2)$, indexed by a *feature space* \mathcal{G} . These can be taken to be Gabor filters as well as any other bank of filters, for instance obtained through automatic learning procedures or to fit the shape of a set of experimentally measured RPs. As we will show later, endstopped simple cells may be included as well by adding a suitable sub-family of filters. Typically, the

feature space has the form $\mathcal{G} = \mathbb{R}^2 \times \mathcal{F}$, where $(x, y) \in \mathbb{R}^2$ represents the retinal position at which the filter is centered and $f \in \mathcal{F}$ encodes the other features detected by the filters. Let us define, for every $p, p_0 \in \mathcal{G}$, the *generating kernel*

$$K(p, p_0) := \operatorname{Re} \langle \psi_p, \psi_{p_0} \rangle_{L^2} = \operatorname{Re} \left(\int_{\mathbb{R}^2} \psi_p(x, y) \overline{\psi_{p_0}(x, y)} dx dy \right).$$

This kernel is the reproducing kernel (Aronszajn 1950) on the closure of the linear span of $\{\psi_p\}_p$ in $L^2(\mathbb{R}^2)$ (see also Duits (2005)); it extends to a general family of filters the familiar idea of the reproducing kernel associated to a family of wavelets (Antoine and Murenzi 1996; Deng et al. 2010). In the special case where \mathcal{G} has a group structure and the filters are given by $\psi_p = \mathcal{U}_p \psi$ for a fixed *mother filter* ψ through a unitary group representation \mathcal{U} , one has

$$K(p, q) = K(p^{-1}q, \mathbf{1}). \tag{5}$$

This is indeed the case for the example of a bank of Gabor filters that will be examined later.

The kernel K expresses the strength of *correlation* between the RPs of two cells, represented by the points p and p_0 of the feature space. Now, it is not restrictive to assume that the filters are normalized to have squared L^2 -norm equal to some $\eta > 0$. Under this assumption, we define the *kernel distance*

$$d^2(p, p_0) = 2(\eta - K(p, p_0)). \tag{6}$$

In fact, d^2 coincides with the squared L^2 distance between the functions ψ_p and ψ_{p_0} . Indeed, $\|\psi_p - \psi_{p_0}\|_{L^2}^2$ can be rewritten as

$$\|\psi_p\|_{L^2}^2 + \|\psi_{p_0}\|_{L^2}^2 - 2\operatorname{Re} \langle \psi_p, \psi_{p_0} \rangle_{L^2}.$$

The generating kernel K and the distance d have a local sense. Indeed, the filters typically have a very localized support and it does not make much sense to compute L^2 scalar products between them if they are not sufficiently close together in the space domain. Moreover, the shape of the specific filters taken into consideration may bring one to introduce some restriction on which filters can directly be influenced by one another. This corresponds to defining around each point $p_0 \in \mathcal{G}$ a local distance function defined on a *patch*, which we shall call $\mathcal{P}(p_0)$, and to gluing all these local distances together to obtain a global distance function

$$\tilde{d}(p, p_0) = \inf \left\{ \sum_{j=1}^N d(q_{j-1}, q_j) : N \in \mathbb{N}, q_0 = p_0, q_N = p, q_j \in \mathcal{P}(q_{j-1}) \forall j \right\}.$$

Note that the length $N + 1$ of the sequence $\{q_j\}$ is let vary: the inf is taken over all $N \in \mathbb{N}$ and over all sequences

compatible with the patches. Under a non-degeneracy condition on the definition of such patches, this is still well-defined as a distance function (see Montobbio et al. (2017) for more details).

Although we restrict our analysis to the pairwise relations between RPs, some studies have emphasized the significance of higher order components. In Lawlor and Zucker (2013), third order statistics of natural scenes are examined: in terms of reciprocal influences between cells, this would mean characterizing not only the probability of each cell to be activated given the activation of a “central” neuron, but also the probability of *co-activation* of two neurons given the activation of the central one. In geometrical terms, such a model allows for curvature dependencies to emerge. See also Ben-Shahar and Zucker (2004) and Abbasi-Sureshjani et al. (2016). As a matter of fact, we will show that some information on the curvature tuning of some classes of neurons, namely endstopped simple cells, emerges in our model as well.

Example: Gabor filters. Let us now consider, as a first basic example, the case of a bank of Gabor filters of fixed scale σ . Its general element reads:

$$\psi_{x,y,\theta}(u, v) = \exp\left(\frac{2\pi i X}{\lambda}\right) \exp\left(-\frac{X^2 + Y^2}{2\sigma^2}\right), \quad (7)$$

with

$$\begin{cases} X = (u - x) \cos \theta + (v - y) \sin \theta \\ Y = -(u - x) \sin \theta + (v - y) \cos \theta. \end{cases}$$

Denote $p = (x, y, \theta)$. A direct calculation yields:

$$\begin{aligned} K(p, 0) = & \sigma^2 \pi \exp\left(-\frac{x^2}{4\sigma^2} - \frac{y^2}{4\sigma^2} - \frac{2\sigma^2 \pi^2 (1 - \cos \theta)}{\lambda^2}\right) \\ & \cdot \cos\left(\pi \frac{x(1 + \cos \theta) + y \sin \theta}{\lambda}\right). \end{aligned}$$

The distance induced by K is given by:

$$d^2(p, 0) = 2\sigma^2 \pi - 2 K(p, 0). \quad (8)$$

Recall that the feature space in this case is $\mathcal{G} = \mathbb{R}^2 \times S^1 = SE(2)$: as pointed out before, the explicit expression for $K(p, p_0)$ for $p_0 \neq 0$ can then be obtained from the one for $K(p, 0)$ through the group operation. We have:

$$\begin{aligned} & K((x, y, \theta), (x_0, y_0, \theta_0)) \\ & = K((R_{\theta_0} T_{(x_0, y_0)}(x, y), \theta - \theta_0), (0, 0, 0)). \end{aligned}$$

In Montobbio et al. (2017) we show that, in the Gabor case, d is estimated by a Riemannian distance, and that such a metric turns out to be a Riemannian approximation to the sub-Riemannian structure defined in Citti and Sarti (2006) on $\mathbb{R}^2 \times S^1$. Specifically, such convergence takes place when σ^2 is proportional to the wavelength λ and $\lambda \rightarrow 0$. This

means that the number of oscillations of the filters under the Gaussian bell goes to infinity; on the other hand, their (squared) norm $\eta = \sigma^2 \pi$ shrinks.

3.2 Propagation by repeated applications of K

A cortex model must be able to explain not only the action of RPs but also contextual influences and the emergence of association fields. Several existing differential models (reviewed in Section 2.4) define on the feature space a Lie group structure, on which the horizontal connectivity is described by means of a diffusion equation expressed through second order operators associated to an invariant sub-Riemannian metric. By contrast, in the present model the local geometry of the cortical space is determined by the receptive profiles of the neurons themselves, through a generating kernel which expresses the reciprocal influences between simple cells with overlapping RFs. The long-range connectivity is then described by iterating the action of this kernel.

It is important to remark that the horizontal connections take place in V1 layer 2/3, which is composed of phase-invariant complex cells. Note that

$$\begin{aligned} K(p, p_0) &= \text{Re}\langle \psi_p, \psi_{p_0} \rangle_{L^2} \\ &= \langle \text{Re}\psi_p, \text{Re}\psi_{p_0} \rangle_{L^2} + \langle \text{Im}\psi_p, \text{Im}\psi_{p_0} \rangle_{L^2}, \end{aligned}$$

which for $p = p_0$ reads

$$K(p, p) = \|\text{Re}\psi_p\|_{L^2}^2 + \|\text{Im}\psi_p\|_{L^2}^2.$$

This recalls the classical energy model for V1 complex cells. In the case of Gabor filters, this is indeed the square sum of a quadrature pair of filters, since the real and imaginary parts of $\psi_{x,y,\theta,\sigma}$ show the same orientation and scale but are shifted in phase of 90° .

Now, how does the activity spread to reach neurons with widely separated RFs? First, consider the operator

$$S[K](p, q) := \frac{K^{(1)}(p, q)}{\int_{\mathcal{G}} K^{(1)}(p', q) d\mu(p')} \quad (9)$$

with

$$K^{(1)}(p, q) = \frac{h(K(p, q))}{\int_{\mathcal{G}} h(K(p', q)) d\mu(p') \cdot \int_{\mathcal{G}} h(K(p, q')) d\mu(q')},$$

where h is a nonlinear function. Note that the kernel is first normalized w.r.t. both its variables, followed by a further *anisotropic* normalization, as done in Coifman and Lafon (2006), where a kernel expressing the local geometry of a given dataset is defined. Then consider the new kernel computed around a point p_0 , i.e.

$$K_1^{p_0}(p) := S[K](p, p_0). \quad (10)$$

The simplest choice for the function h is $h(z) := \max(z - \tau, 0)$. Setting $\tau = 0$ implies discarding the inhibitory

effects (represented by negative values), while $\tau > 0$ would mean only keeping the strongest excitatory responses, thus essentially implementing the principle of non-maxima suppression at every diffusion step (as in Bressloff and Cowan (2003) and Bressloff et al. (2002) or Citti and Sarti (2006)). Another common choice of h is the *logistic sigmoidal function*

$$h(z) = \frac{1}{1 + e^{-z}}.$$

We now iteratively define, for each $n > 0$,

$$K_n^{p_0}(p) := \int_{\mathcal{G}} S[K](p, q) K_{n-1}^{p_0}(q) d\mu(q). \quad (11)$$

Note that $K_n^{p_0}$ integrates to 1 for every n . Alternatively, one may apply the same iteration rule to the original kernel, i.e. take $K_1^{p_0}(p) := K(p, p_0)$. This would mean keeping the initial inhibitory information, while preserving the mean of the original kernel throughout the propagation thanks to the normalization of the kernels in (11). In both cases, one has

$$\int_{\mathcal{G}} |K_n^{p_0}(p)| d\mu(p) = \int_{\mathcal{G}} |K_1^{p_0}(p)| d\mu(p) \quad \forall n. \quad (12)$$

Also note that, in the presence of a group structure, Eq. (5) holds and the integral in (11) becomes a group convolution.

Here, μ is the spherical Hausdorff measure (Hausdorff 1918) associated to the distance d on \mathcal{G} . This choice is motivated by the fact that, in the general case, the feature space \mathcal{G} is just a set of parameters with no geometric structure whatsoever, apart from the one given by the distance: since we are integrating a function of d , it is natural to do so with respect to a measure induced by the distance itself, which “compensates” for its in-homogeneities. In a simpler case, suppose that \mathcal{G} is a subset of some \mathbb{R}^n , which is in fact equipped with a measure: the actual space might be much lower-dimensional, and integrating with respect to the n -dimensional Euclidean measure would make no sense in this case. On the contrary, the spherical Hausdorff measure captures the “true” dimension of the space.

Note that the spatial extent of the kernel widens at each step, and this width depends on the size of the RPs, i.e. on the diameter of their RFs (RPs are local objects, and they are always modeled through compactly supported or rapidly decaying functions).

There is neurophysiological evidence (Angelucci et al. 2002) that the extension of horizontal connections departing from a V1 cell matches the size of its so-called *low-contrast summation field*; this is identified as the area measured by presenting low-contrast bars or gratings of increasing sizes at the RF center, and keeping the size of peak response. The low-contrast summation field has in turn been shown (Kapadia et al. 1999; Angelucci et al. 2002) to be on average 2.2-fold greater than the “classical” RF. Therefore, since

a relationship between the RF size and the spatial extent of horizontal connections seems to hold from a biological point of view, we believe that the optimal number of iterations may not depend on the RP size. As for the choice of this number, a first stopping criterion is this average ratio between the size of low-contrast summation fields and classical RFs. Another possible quantitative framework for fitting the kernel to the neurophysiological data was proposed in Favali et al. (2017) to fit a Fokker-Planck kernel to the data from Bosking et al. (1997) and Angelucci et al. (2002): the kernel was evaluated on a pinwheel map and compared with the measured distribution of the tracer by comparing their densities onto the rectangles of a grid (whose sampling size was chosen to match the pinwheel scale).

Note that this idea of repeated convolutions can be applied to model the response of the layers of V1 to a visual stimulus I applied to the retina, taking into account the contextual influences. In this case, the activation computed by lifting the image to the cortex is a function of the cortical coordinates:

$$I_0(p) := h \left(\int I(x, y) \psi_p(x, y) dx dy \right).$$

The action of the kernel on this activation at the n -th step may then be expressed by

$$I_n(p) := \int_{\mathcal{G}} S[K](p, q) I_{n-1}(q) d\mu(q). \quad (13)$$

In our work (Montobbio et al. 2017) we constructed a diffusion process on the feature space, equipped with the measure μ , to describe the propagation along the connectivity. Subject to some requirements (Sturm 1995, 1998; Coifman and Lafon 2006), this process can be approximated by updating the initial datum through repeated integrations against a proper normalization of K : indeed, under certain conditions such a kernel provides an estimate for the heat kernel associated to the diffusion process (Coifman and Lafon 2006). See Montobbio et al. (2017) for further details. This remark highlights a link between our kernel-based model and the diffusion-based differential models cited before, showing that these two approaches can actually coexist.

4 Numerical scheme and results

In this section we show, through some numerical simulations, the geometrical properties of the connectivity kernels obtained by applying our model to some different banks of filters. The first distinction to be made in the treatment of these examples from the numerical point of view is that between the *continuous* case and the *discrete* one. This can refer either to the feature space \mathcal{G} or to the spatial domain

of the single filters. For instance, in the case of a Gabor system, both the feature space $\mathcal{G} = \mathbb{R}^2 \times S^1$ and the spatial variable $(u, v) \in \mathbb{R}^2$ need to be sampled. On the contrary, when dealing with a bank of learned filters, the family of indices \mathcal{G} and the filter domains are finite sets. Of course, one could also consider “hybrid” cases, e.g. a bank of filters defined on \mathbb{R}^2 but indexed by a discrete set.

For what concerns the computation of the initial kernel, the numerical setup is very basic. The filters are either defined on \mathbb{R}^2 or numerically known onto a set of the type $\{(\frac{i}{n}, \frac{j}{n})\}_{i,j=-n,\dots,n} \subseteq \mathbb{R}^2$; computing the generating kernel $K(p, q)$ simply involves taking the integral of the function $\psi_p \cdot \overline{\psi_q} \in L^1(\mathbb{R}^2)$, which is either compactly supported or exponentially decaying. Thus, the filter domains can safely be sampled (when needed) through a bounded square grid

$$\{(x_i, y_j)\}_{i,j=0,\dots,N} \subseteq [-W, W] \times [-H, H],$$

for some $N, W, H > 0$, with

$$x_{i+1} - x_i = y_{j+1} - y_j =: \delta > 0 \quad \forall i, j.$$

The kernel $K(p, q)$ is then standardly calculated as

$$\frac{1}{\delta^2} \sum_{i,j} \psi_p(x_i, y_j) \overline{\psi_q(x_i, y_j)}.$$

In the Gabor case, as shown in Section 3.1, this integral can even be computed analytically, thus avoiding this approximation.

As for the propagation, first note that, in the case of a finite feature space $\mathcal{G} = \{p_1, \dots, p_M\}$, the Hausdorff measure reduces to the counting measure: integrals over \mathcal{G} are finite unweighted sums over its elements. Although seemingly trivial, this indeed captures the distribution of the features throughout the data thanks to the inhomogeneity of the distance d . Loosely speaking, given a certain set of filters, some features may be “more densely” represented than others w.r.t. d . As a basic example, suppose that the filter ψ_{p_1} is highly correlated with 8 other filters, i.e. for a fixed ε

$$|\{i \in \{2, \dots, M\} : d(p_i, p_1) < \varepsilon\}| = 8.$$

On the other hand, suppose that ψ_{p_2} has only 2 filters highly correlated with it. As a consequence, $\mu(B_\varepsilon^d(p_1)) = 9$ and $\mu(B_\varepsilon^d(p_2)) = 3$, i.e. the balls of d of the same radius ε centered at p_1 and p_2 have different measures. For a non-finite metric space, the spherical Hausdorff measure extends this concept, although in general it is not explicitly computed; however, the distance d in the Gabor case turns out to be estimated by a Riemannian distance (Montobbio et al. 2017), and the spherical Hausdorff measure on a Riemannian manifold coincides up to a constant factor with the Riemannian volume form (Federer 1969). We will go into more detail on the discretization of the Gabor feature space in Section 4.1.

We stress that this work focuses on showing that our construction allows to recover a connectivity pattern starting from non-structured banks of filters. A great interest is thus addressed to numerically known (therefore finite) sets of filters, obtained for instance through learning algorithms; one may even apply it to neurophysiologically measured RPs. We refer to Montobbio et al. (2017) for some theoretical issues and more examples related to the spherical Hausdorff measure in that context.

A last point to be mentioned is the role of the normalization performed at each kernel iteration. Namely, as shown in Eq. (12), the L^1 norm of $K_n^{p_0}$ is kept the same throughout the propagation, which prevents the values of $K_n^{p_0}$ from exploding as $n \rightarrow \infty$.

In the following, we first resume the classical example of Gabor filters presented in Section 3.1 and display the association fields generated by the connectivity in this case. We then show the curvature selectivity of the association fields emerging from endstopped simple cells. We also take into account the spatiotemporal behavior of the RPs by considering a family of three-dimensional Gabor filters including a time parameter. Finally, we show that it is still possible to recover a coherent “bow-tie” pattern starting from a family of RPs obtained through a learning algorithm, whose parameterization carries no a priori geometric information.

4.1 Gabor filters

We now go back to the bank of Gabor filters $\{\psi_{x,y,\theta}\}$ defined by Eq. (7). Since the generating kernel in this case is known analytically, no discretization of the filter domain \mathbb{R}^2 is needed for its computation. As for the feature space $\mathbb{R}^2 \times S^1$, in the following we consider a sampling of the type

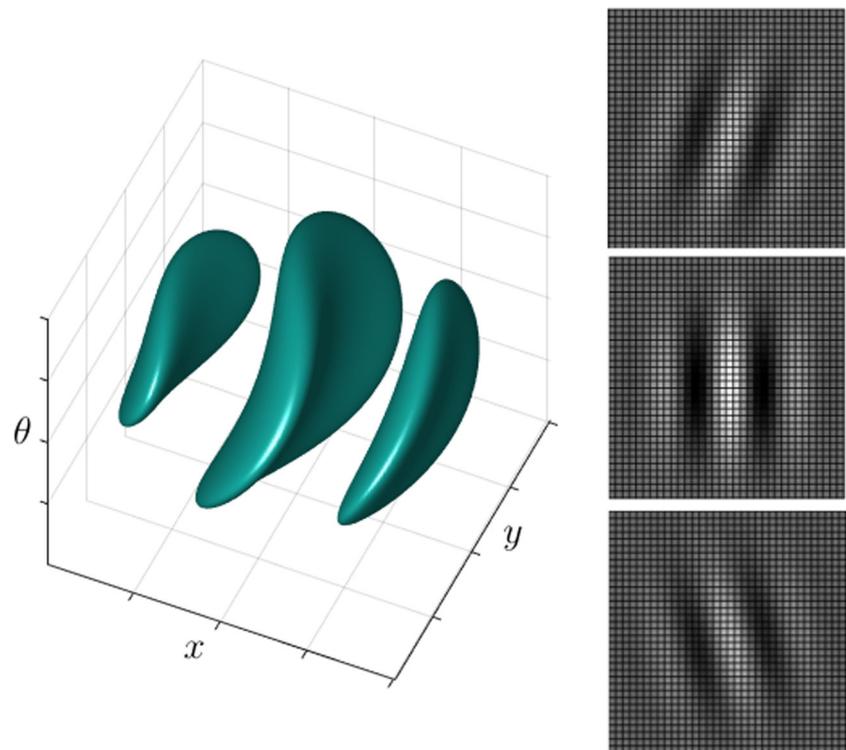
$$\{-\bar{x}, \dots, \bar{x}\} \times \{-\bar{y}, \dots, \bar{y}\} \times \{-\bar{\theta}, \dots, \bar{\theta}\}$$

where $\bar{x} = \bar{y} = 1$ with discretization step .01 and $\bar{\theta} = 1.5$ with step .015 for the visualizations of the generating kernel; $\bar{x} = 1.5$ and $\bar{y} = 3$ with step .1, and $\bar{\theta} = 1.5$ with step .15 for the propagation.

In the case of Gabor filters, the level sets of K (or equivalently, of the distance d) around a fixed point $p_0 = (x_0, y_0, \theta_0)$ of the feature space are in general not connected (see Fig. 6). This is due to the oscillations of the periodic factor.

In terms of the interactions between RPs, the central lobe of the level sets corresponds to the *co-axial* effect, i.e. the one along the axis of the preferred orientation of the starting RP ψ_{p_0} : along this axis, the strongest interactions are the ones with filters collinear with ψ_{p_0} (see Fig. 7b), and the value of K decreases as the orientation of the filters varies from θ_0 . On the other hand, the smaller lobes developing along the orthogonal axis result from the

Fig. 6 On the left, a level set of $K((x, y, \theta), (0, 0, 0))$. On the right, three slices for $\theta = -\frac{\pi}{6}, 0, \frac{\pi}{6}$



“periodic” correlation of the central RP with filters parallel to ψ_{p_0} (i.e. filters with orientation θ_0 but centered on points of the orthogonal axis): the kernel takes positive values when areas of the two RPs with the same sign overlap (as in Fig. 7d), and negative values when areas of opposite

signs overlap (as in Fig. 7c). In a first stage we thought it best to restrict ourselves to the effect along the axis of the preferred orientation, i.e. to consider only the central connected component of the level sets. This only takes a straightforward analytical operation in the Gabor case, and it

Fig. 7 A schematic representation of the interactions between odd RPs sharing the same orientation, depending on their reciprocal position. **a** An example odd filter, with white ON areas and gray OFF areas. The axis of its preferred orientation is displayed. The next pictures show different reciprocal configurations of two such filters. Overlapping areas where the sign of the product is positive are displayed in red and marked by a plus symbol, while zones where the product is negative are blue and marked by a minus symbol. **b** Collinear filters bring to high values of K along their common axis. **c** Parallel filters with overlapping regions of opposite signs yield negative values of K . **d** Parallel filters with overlapping regions of the same sign yield positive values of K

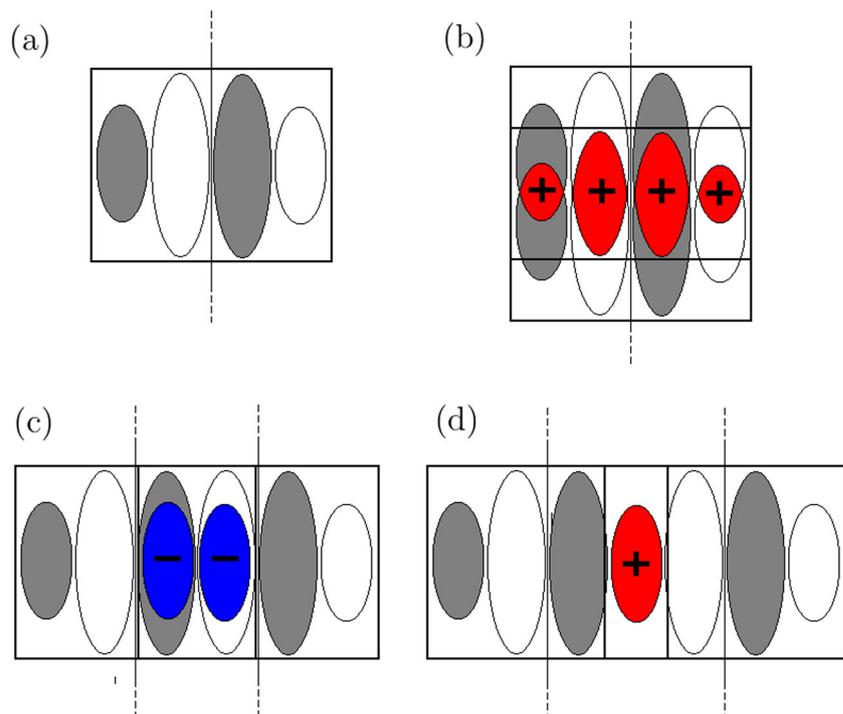
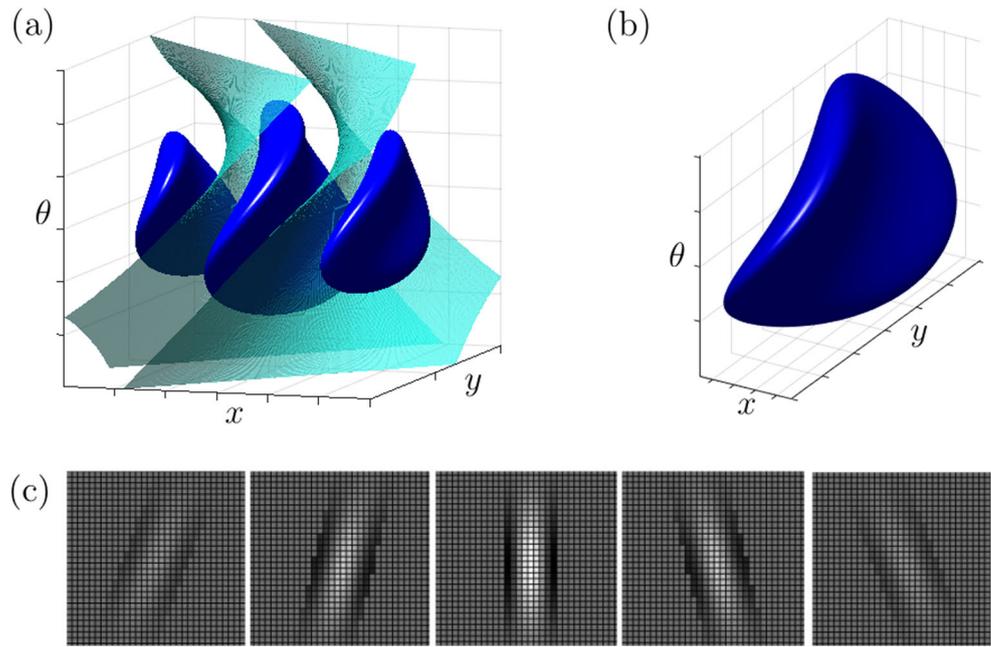


Fig. 8 **a** In dark blue, a level set of $K((x, y, \theta), (0, 0, 0))$. The patch $\mathcal{P}(0, 0, 0)$ is the area between the two surfaces displayed in light blue. **b** The same level set, after truncating. Here, $\lambda = 1$. **c** Horizontal slices of the truncated kernel for $\theta = -0.75, -0.45, 0, 0.45, 0.75$. The white regions represent positive values, the black ones represent negative values. Note that the kernel has been truncated at its minimum



makes it easier to compare our model with the ones obtained through real-valued diffusion equations. We shall therefore clip the kernel as anticipated in the general description above, according to local patches defined around each point $p_0 = (x_0, y_0, \theta_0)$ of $\mathbb{R}^2 \times S^1$:

$$\mathcal{P}(p_0) := \{(x, y, \theta) : |a(1 + \cos \delta) + b \sin \delta| < \lambda\},$$

with $(a, b, \delta) = (R_{\theta_0} T_{(x_0, y_0)}(x, y), \theta - \theta_0)$.

This provides an example of adjustment that can be introduced on the original kernel in order to define some restriction on which filters are influenced by one another. At this point we can truncate the distance function by defining the kernel K only on $\mathcal{P}(p_0)$ around each point p_0 (see Fig. 8), and then glue the resulting distance functions together as in Eq. (8) to obtain a new global structure.

Also note that, since the reciprocal influence of collinear filters is greater than the one between parallel filters, all the values of $K(\cdot, p_0)$ above a sufficiently high threshold are confined inside $\mathcal{P}(p_0)$: this means that small balls of the distance d are not altered by the clipping. In other words, the local structure is preserved.

Nonetheless, we believe that the whole kernel is interesting for a further analysis, since its oscillatory behavior in the orthogonal direction seems to account for the *trans-axial* “ladder” effect which has indeed been observed in both neurophysiological and psychophysical studies (Mitchison and Crick 1982; Field et al. 1993; Yen and Finkel 1998) (see also Fig. 9), thus naturally including in the model the Gestalt perceptual principle of parallelism.

The generating kernel K computed around a point $(x_0, y_0, \theta_0) \in \mathcal{G}$ is a three-dimensional function

$$(x, y, \theta) \mapsto K((x, y, \theta), (x_0, y_0, \theta_0)).$$

Since the feature space for this family is of the form *position* \times *orientation*, it is possible to display a projection of it on the retinal plane.

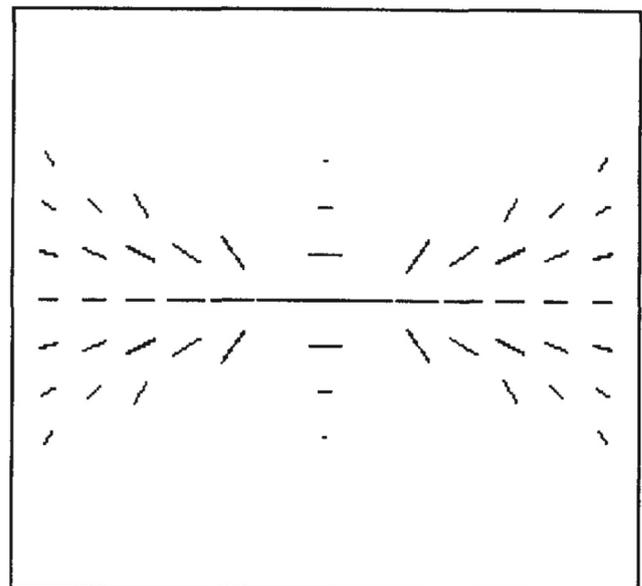


Fig. 9 The *connectivity pattern* found in the model of Yen and Finkel (1998) includes both co-axial and trans-axial contributors. The length of the lines indicates connection strength

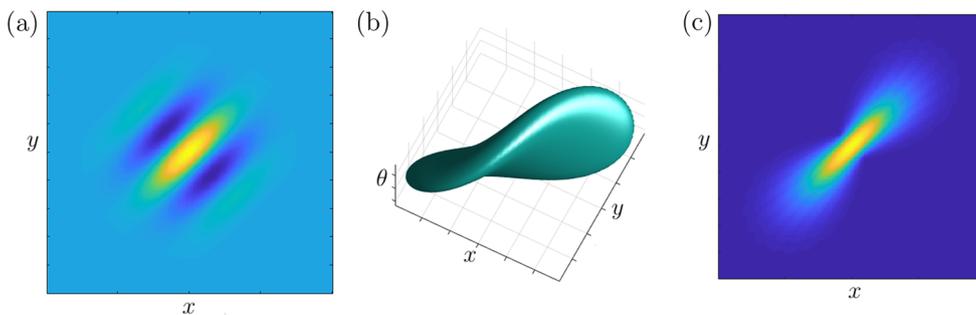


Fig. 10 The behavior of $(x, y, \theta) \mapsto K((x, y, \theta), (0, 0, \frac{\pi}{4}))$ in the case of Gabor filters. The kernel has been truncated as explained above. In particular, the images show: **a** the receptive profile ψ_{p_0}

corresponding to $p_0 = (0, 0, \frac{\pi}{4})$; **b** a level set in $\mathbb{R}^2 \times S^1$; **c** the projection on the (x, y) plane obtained taking the maximum in θ

Specifically, by taking the maximum in the variable θ and projecting onto the (x, y) plane, we obtain a 2D function concentrated around (x_0, y_0) , as displayed in Fig. 10c.

We now display some stages of the iterative process described in Section 3.2, for the truncated kernel K .

Figure 11b displays the projection onto the retinal plane of the kernel around $(0, 0, 0)$, at four steps of the iterative process. For $n = 1, 2, 3, 4$ the function

$$(x, y, \theta) \mapsto K_n^{(0,0,0)}(x, y, \theta) \tag{14}$$

was projected onto the (x, y) plane by taking the maximum in the variable θ . The projection of the connectivity kernel onto the retinal plane highlights the orientation selectivity of the propagation, with the points of maximum intensity concentrating along the axis of the preferred orientation of the central filter. As specified earlier, the horizontal connections of V1 are considered as a likely anatomical background to the perceptual rules synthesized by association fields. To develop this aspect, we added a further visualization displaying the orientation of the filter maximizing the intensity at each location (where the intensity exceeds a threshold). Specifically, Fig. 11c shows at each such location (x, y) a tiny Gabor filter whose

orientation value $\bar{\theta}(x, y)$ is the one that maximizes the value of the final 3D function, i.e.

$$\bar{\theta}(x, y) := \arg \max_{\theta} K_4^{(0,0,0)}(x, y, \theta).$$

Similar results have been obtained in several works concerning second-order statistics on the distribution of edges in natural images. In August and Zucker (2000), the statistical measurement of the organization of local edge elements into curves led to the introduction of an oriented filtering operation, expressed by a kernel on $\mathbb{R}^2 \times S^1$ (see Figure 15.3 in August and Zucker (2000)); its orientation specificity and the negative lateral lobes recall some characteristics of our connectivity kernel, especially after clipping it at its minimum as explained above (cf. the slices in Fig. 8c). In Geisler et al. (2001), association field-like structures comparable to our Fig. 11c emerge from both a statistical analysis of edge co-occurrence in natural images, and a Bayesian study analyzing the probability of two edge elements belonging to the same extended contour. A similar analysis is carried out in Elder and Goldberg (2002), except that here contours are described as *ordered*

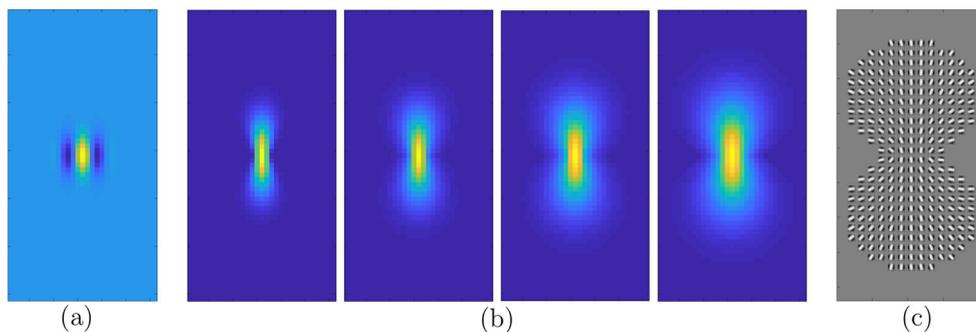


Fig. 11 Development of a connectivity kernel around $(x_0, y_0, \theta_0) = (0, 0, 0)$, obtained through repeated integrations against the generator K . Four iterations were performed. The kernel was truncated as explained in Section 3.1, and the operator S defined in Section 3.2 was applied. The resulting function $K_4^0(x, y, \theta)$ was then projected on

the image plane by taking the maximum in the variable θ . The pictures show: **a** the Gabor filter $\psi_{0,0,0}$ (real part); **b** the projection on the (x, y) -plane of the 3D function obtained at each of the three steps; **c** for each (x, y) , the value $\bar{\theta} = \arg \max_{\theta} K_4^0(x, y, \theta)$ represented through a Gabor filter with orientation $\bar{\theta}$

sequences of tangent elements; in this work a stronger role of the parallelism cue emerges, i.e. there is a marked correlation between parallel edge elements. See also Kruger (1998) and Sigman et al. (2001) as further references on the emergence of the collinearity, cocircularity and parallelism cues from statistics on natural images.

These studies are not to be regarded as a separate kind of analysis leading to results consistent with ours. Indeed, it is believed that the cortical processing is influenced by environmental statistics and adapts to best process those signals that are most likely to occur.

Propagation on the pinwheel surface Consider now the subset $\{\tilde{\psi}_{x,y}\} = \{\psi_{x,y,\theta(x,y)}\}$ of the above family determined by a fixed orientation map $\theta(x, y)$. As pointed out in Petitot (2008), orientation map-like structures can be reproduced as a superposition of plane waves with

random phases. The generating kernel $\tilde{K}((x, y), (x_0, y_0))$ associated to this family is given by

$$K((x, y, \theta(x, y)), (x_0, y_0, \theta(x_0, y_0))).$$

Starting from this 2D generator, we computed the connectivity kernel

$$(x, y) \mapsto \tilde{K}_n^{(0,0)}(x, y) \tag{15}$$

onto the pinwheel surface θ parameterized by (x, y) , around its point $(0, 0)$. We then displayed it (for $n = 6$), as well as the orientation map itself, onto the (x, y) plane, see Fig. 12a. Note that, along with the orientation specificity already observed in the preceding experiment, a patchy nature of the propagation emerges as well, according to the regions of the map with orientation values close to $\theta(0, 0)$. This behavior matches the one observed in Bosking et al. (1997), where small biocytin injections in a site of tree shew V1 were

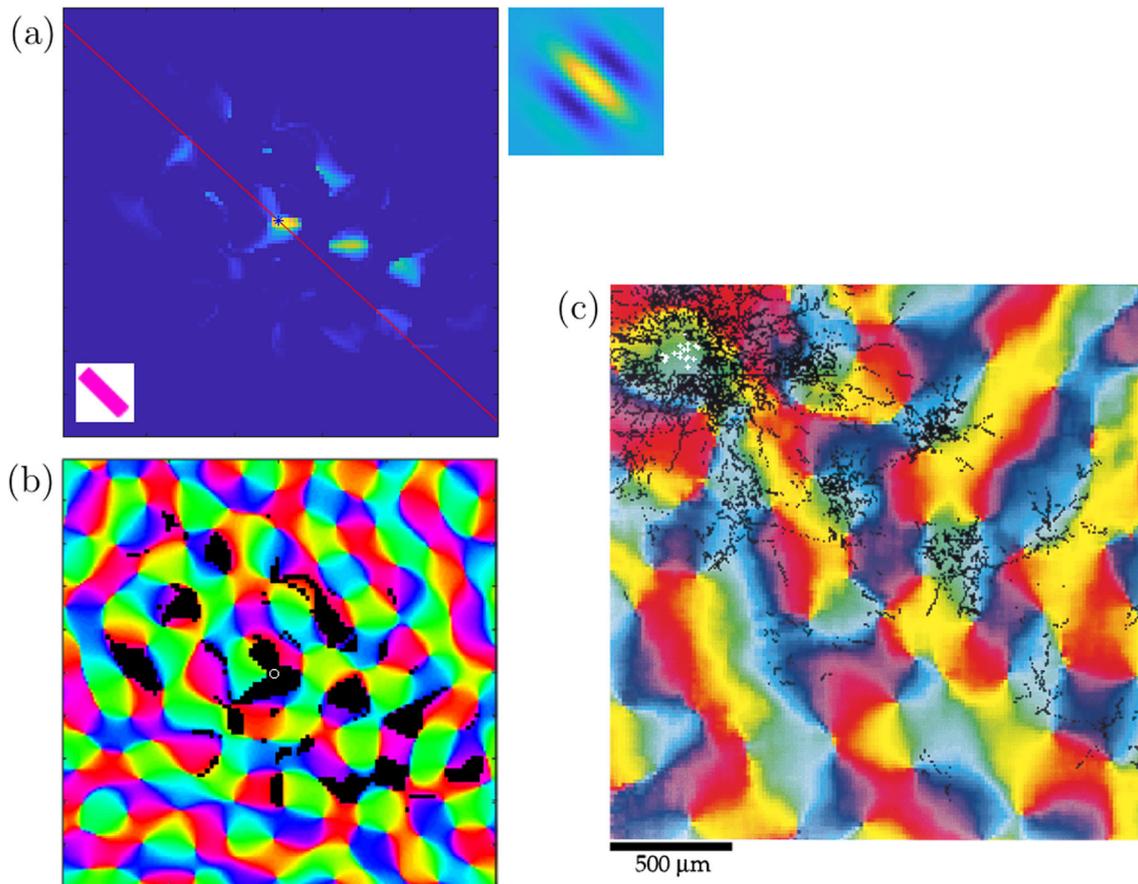


Fig. 12 Left side: propagation onto a pinwheel map $\theta(x, y)$ generated through superposition of plane waves, as explained in Petitot (2008). In particular, image (a) shows the function $\tilde{K}_6^{(0,0)}(x, y)$ resulting from six steps of evolution through \tilde{K} , with initial point $(0, 0)$. In the lower left corner, a bar with orientation $\theta(0, 0)$ is displayed, whose color corresponds to $\theta(0, 0)$ in the orientation map. The starting filter $\tilde{\psi}_{0,0}$ (real part) is displayed as well (top right). Image (b)

shows the values of u exceeding a threshold, displayed in black onto the orientation map. The initial point $(0, 0)$ is marked with a white circle. Right side: distribution of boutons formed after a biocytin injection into a site of tree shew V1, displayed over the topographic map of orientation preference. The cells that took up the biocytin are displayed in white. (Image (c), modified from Bosking et al. (1997))

made to analyze the distribution of the resulting horizontal connections with respect to the orientation preference map. Figure 12c shows the distribution of boutons obtained from one such injection. For the sake of comparison, we displayed onto the orientation map $\theta(x, y)$ the points (in black) where (15) exceeds a threshold (Fig. 12b). In both images, the starting point is displayed in white. We provide Matlab code to generate simulations on pinwheel maps in Online Resource 1.

In fact, in the experiment of Bosking et al. (1997), the recorded connections departing from the injection site originate not from a single cell, but rather from a few cells, the ones which took up the biocytin. However, all of these cells are confined in a localized area of V1: thanks to the continuity (outside pinwheels) of the orientation map, they happen to have very close preferred orientations. We believe that, if the biocytin injection could address a single neuron, the effect (although weaker) would be substantially the same in terms of orientation specificity. On the other hand, the multiple neurons situation may be reproduced in the context of our model by considering a set of RPs, close together in the feature space, and by combining the kernels generated by each of them (e.g. by summation).

4.2 Endstopped simple cells

We will now focus on the connectivity pattern emerging from a particular class of neurons, namely *endstopped simple* (ES) cells. Endstopped cells (also referred to as *hypercomplex*) are orientation-selective neurons characterized by their sensitivity to bars of specific lengths: their RPs are characterized by inhibitory “end zones” suppressing the response to stimuli longer than the central excitatory area. Endstopped cells have been shown Hubel and Wiesel (1962) to react to curved stimuli; a mathematical description of the relation between endstopping and curvature has been developed in Dobbins et al. (1987).

ES cells have *simple* RPs equipped with end zones along the preferred orientation. They can be modeled by linear combination of two similarly positioned and oriented simple RPs of different sizes, as proposed in Dobbins et al. (1987). Specifically, one such RP ψ can be written as the (weighted) sum of the positive contribute of a smaller simple RP ψ^S and the negative contribute of a larger simple RP ψ^L , as follows:

$$\psi := c_S \psi^S - c_L \psi^L, \quad (16)$$

where $c_S > c_L > 0$.

Figure 13 shows an example of even and odd RPs generated as in (16), where the simple RPs ψ^S and ψ^L are modeled by Gabor filters. To be more precise, the model in Dobbins et al. (1987) proposes to represent the response R of ψ to some contrast pattern as

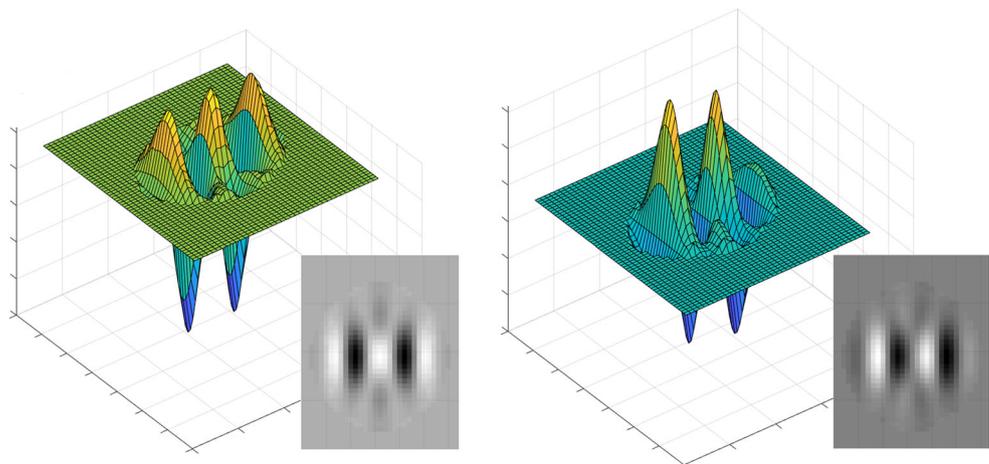
$$R = h(c_S h(R_S) - c_L h(R_L)), \quad (17)$$

where R_S and R_L are the responses of ψ^S and ψ^L to such pattern, and h is the rectifying function $h(z) = \max(0, z)$. This increases the stimulus specificity compared to directly computing the response from (16).

Nonetheless, the two models coincide when the responses of both ψ^S and ψ^L to a visual stimulus are non-negative, i.e. for those stimuli in the “preferred range” of orientation and curvature of the ES cell.

We computed the kernels associated to some families of ES profiles of different lengths and displayed the resulting association fields, obtained as before by selecting the orientation that maximizes the intensity at each location. This experiment shows that our model allows to recover the link between the length of ES cells and the curvature of the corresponding association fields. The left column of Fig. 14 shows the connectivity patterns obtained for a set of non-endstopped simple cells (top image), corresponding to zero curvature, and for three families of ES cells of decreasing

Fig. 13 ES profiles obtained as in Eq. (16)



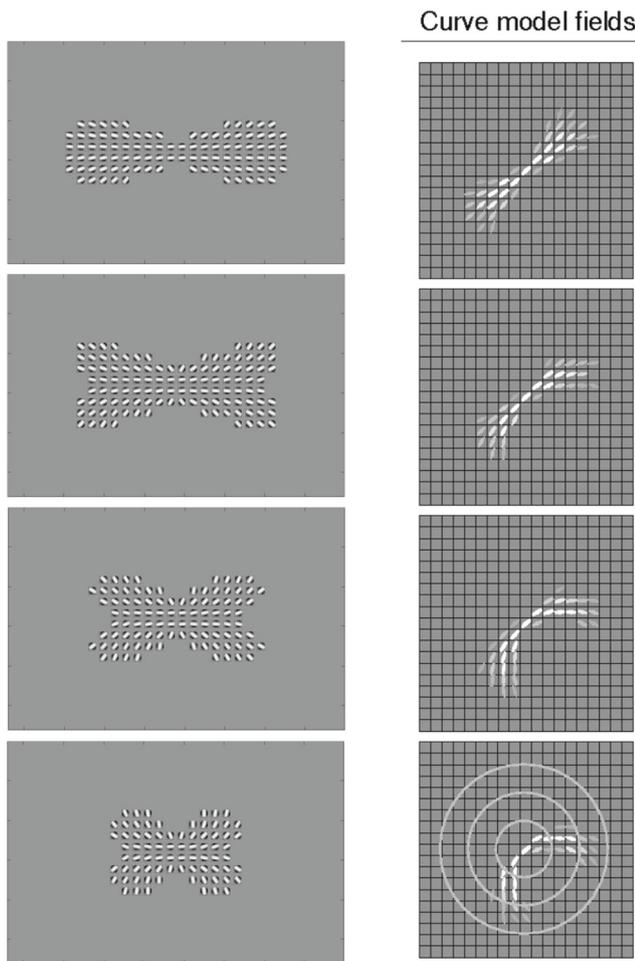


Fig. 14 On the left, association fields emerging from the connectivity kernels associated to a simple cell (top) and to three ES profiles of varying length. On the right, the connection fields of curves obtained in Ben-Shahar and Zucker (2004) relative to curvature $\kappa = 0$ (top) and three increasing positive values of κ respectively

lengths, corresponding to increasing absolute values of curvature. In each of the cases displayed, we considered 21 RPs with orientation values equally spaced in $[-\pi, \pi]$. We then shifted them spatially to generate the whole filter bank. Our results are compared to the curvature-based connection fields obtained in Ben-Shahar and Zucker (2004) through

differential tools, displayed in the right column of Fig. 14. Since the sign of the curvature is taken into consideration in Ben-Shahar and Zucker (2004), each of our association fields corresponds to a couple of theirs (with curvatures κ and $-\kappa$).

4.3 Spatiotemporal Gabor filters

Let us consider a family of filters still of Gabor type, but taking into account the movement of the stimulus. Cocci et al. (2012) fitted the RPs of a set of V1 neurons showing velocity-selective behaviours with a three dimensional Gabor model. That is we have, in addition to the two spatial dimensions, a third temporal dimension in the domain of these filters, which in fact form a subset of $L^2(\mathbb{R}^3)$.

A convenient visualization of such filters is as a temporal sequence of spatial maps (see Fig. 15, taken from Cocci et al. (2012)). In Barbieri et al. (2014), the authors also develop a differential model of functional architecture based on such 3D Gabor family.

Now, in order to maintain the notations as similar as possible to the one that we used for 2D Gabor filters, we shall write the general *inseparable* element of the family of filters as

$$\psi_{x,y,\theta,t,\alpha}(u, v, s) = \exp\left(-2\pi i \left(\frac{X}{\lambda} + \alpha(s - t)\right)\right) \cdot \exp\left(-\frac{X^2 + Y^2}{2\sigma^2} - \frac{(s - t)^2}{2\beta^2}\right),$$

where

$$\begin{cases} X = (u - x) \cos \theta + (v - y) \sin \theta \\ Y = -(u - x) \sin \theta + (v - y) \cos \theta. \end{cases}$$

The set of filters is indexed by the position $(x, y) \in \mathbb{R}^2$, the orientation $\theta \in S^1$, the time $t \in \mathbb{R}^+$ at which the response of the filter is maximum, and the velocity $\alpha \in \mathbb{R}$. Parameters λ and β are fixed.

Let us compute the generating kernel K in this setting. The feature space in this case is $\mathbb{R}^2 \times S^1 \times \mathbb{R} \times \mathbb{R}$. We compute $K((x, y, \theta, t, \alpha), (x_0, y_0, \theta_0, t_0, \alpha_0))$ with $t = t_0$: this means considering two cells whose activation peaks

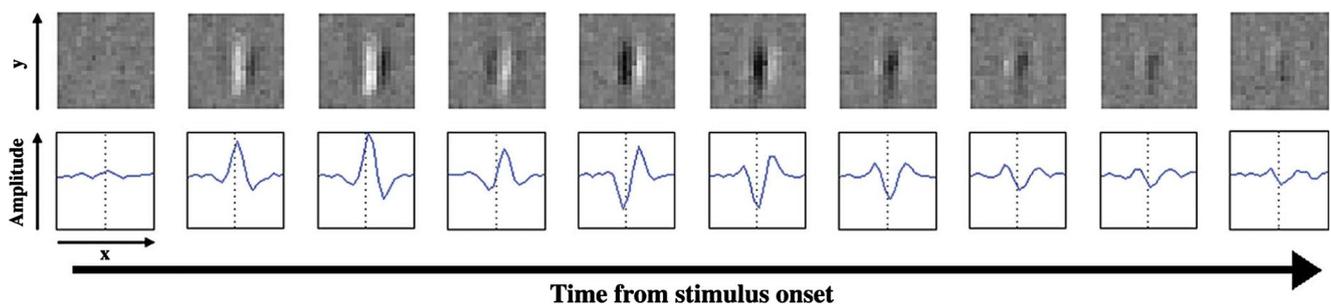


Fig. 15 The time course of the recording of a simple cell’s RP. Source: Cocci et al. (2012)

at the same time. In other words, we fix t and we take $\mathcal{G} = \mathbb{R}^2 \times S^1 \times \mathbb{R}$ as our feature space. We restrict to this case in order to be able to better interpret the results in terms of the orientation and velocity parameters. Denote $p = (x, y, \theta, \alpha)$ and $p_0 = (x_0, y_0, t_0, \alpha_0)$. We obtain the following expression.

$$K(p, p_0) = K^{\text{spatial}}((x, y, \theta), (x_0, y_0, \theta_0)) \cdot \beta^2 \sqrt{\pi} \exp\left(-\frac{\beta^2(\alpha - \alpha_0)^2}{4}\right),$$

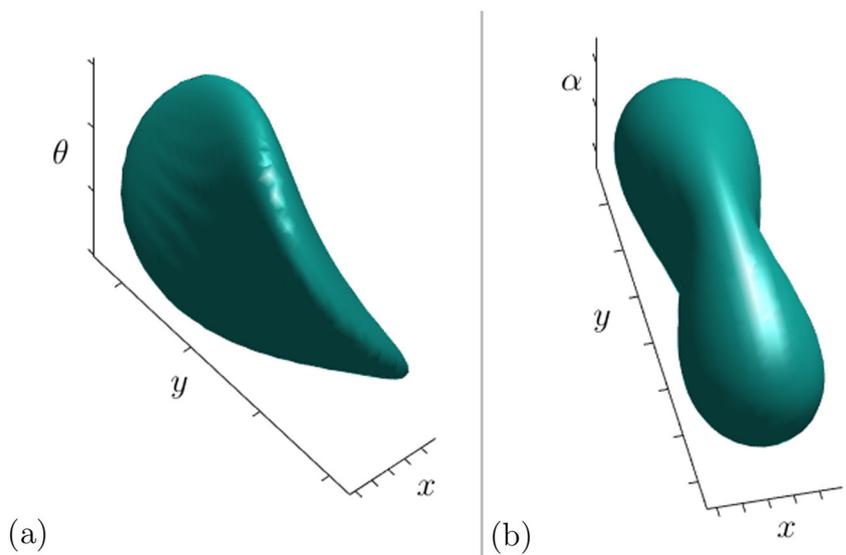
where K^{spatial} denotes the generator obtained in the time-independent case discussed above. Through repeated integrations against K , we obtain the connectivity kernels $K_n^{(x_0, y_0, \theta_0, \alpha_0)}$ associated to this family of filters.

Since we are considering a 4-dimensional feature space, we visualize the generating kernel by projecting it onto the 3-dimensional spaces $\mathbb{R}^2 \times S^1$ (position and orientation) and $\mathbb{R}^2 \times \mathbb{R}$ (position and velocity). Figure 16 displays the kernel around $p_0 = (0, 0, 0, 0)$ projected (a) onto $\mathbb{R}^2 \times S^1$ by taking the maximum in the variable α , and (b) onto $\mathbb{R}^2 \times \mathbb{R}$ by taking the maximum in the variable θ .

Note that the filters $\psi_{x,y,\theta,t,\alpha}$ represent cells that respond maximally to only one direction of movement (depending on the sign of α): such profiles are called *inseparable*. However, there exist also cells which are equally sensitive to both directions, whose profiles are called *separable* (since they can be obtained by the product of two *real* functions of space and time respectively), as well as cells sensitive to both directions but to a different extent. The family of all these profiles can be obtained through weighted sums of inseparable profiles (Cocci et al. 2012):

$$\psi_{x,y,\theta,t,\alpha}^C(u, v, s) = C\psi_{x,y,\theta,t,\alpha}(u, v, s) + (1 - C)\psi_{x,y,\theta,t,-\alpha}(u, v, s),$$

Fig. 16 Visualizations of the spatiotemporal generating kernel computed around the point $(x_0, y_0, \theta_0, \alpha_0) = (0, 0, 0, 0)$. **a** A level set of its projection onto the (x, y, θ) space, obtained by taking the maximum in α . **b** A level set of its projection onto the (x, y, α) space, obtained by taking the maximum in θ



where $C \in [0, 1]$ is the *separability index*, weighing the contribute of sensitivity to the two opposite directions of movement, expressed by the velocity parameters α and $-\alpha$. Note that, if we introduce C as a parameter, then α can be taken to be nonnegative as an index.

The generator \tilde{K} for the complete family

$$\{\psi_{x,y,\theta,t,\alpha}^C\}_{x,y,\theta,t,\alpha,C}$$

is easily obtained from the generator K for the inseparable family above. Denote

$$q = (x, y, \theta, t, \alpha, C),$$

$$q_0 = (x_0, y_0, \theta_0, t_0, \alpha_0, C_0).$$

We have:

$$\begin{aligned} \tilde{K}(q, q_0) = & CC_0K(p^+, p_0^+) \\ & + C(1 - C_0)K(p^+, p_0^-) \\ & + (1 - C)C_0K(p^-, p_0^+) \\ & + (1 - C)(1 - C_0)K(p^-, p_0^-), \end{aligned}$$

where $p^+ = (x, y, \theta, t, \alpha)$, $p^- = (x, y, \theta, t, -\alpha)$ and p_0^+, p_0^- are defined similarly.

4.4 A family of learned filters

We now test our model in the case of a family of numerically-known filters indexed by a discrete subset of $\mathbb{R}^2 \times \mathcal{F}$, where the *fiber* \mathcal{F} is just a set of indices with no a priori geometric structures. The aim of this example is to show that the metric structure defined on this feature space by the kernel K still generates “bow-tie” patterns onto the retinal plane.

Specifically, we chose a bank of filters obtained through an unsupervised learning algorithm which maximizes sparseness; this procedure was first proposed by Olshausen

and Field in 1996 (Olshausen and Field 1996) in order to find efficient linear codes for natural scenes, as an attempt to understand the response properties of visual neurons in terms of the statistical structure of natural images.

Their algorithm generates a family of localized, oriented, bandpass RPs, as the 128 shown in Fig. 17 (which were generated using a later version of the algorithm, provided by Lee et al. in 2007 (Lee et al. 2007)).

We then “centered” the support of each filter ψ by identifying the spatial location around which ψ is concentrated and cropping its domain (16×16 pixels originally) symmetrically around this point to obtain a 11×11 pixel support. The central location was simply chosen to be the point where the function ψ reaches its maximum. To manage the cases in which the maximum point was near the border, we added a 5-pixel padding of zeros around the initial filters before cropping.

At this point, we have a set $\{\psi_f\}_{f=1,\dots,128}$ of functions centered at zero: by shifting them spatially, we obtain a family of filters $\psi_{x,y,f}$ centered at (x, y) , where $x, y \in \{-W, \dots, W\}$. Therefore, the feature space in this case can be written as

$$\mathcal{G} = \{-W, \dots, W\}^2 \times \mathcal{F},$$

where $\mathcal{F} = \{1, \dots, 128\}$. The width W depends on the size of the images one considers onto the retinal plane. For example, in Olshausen and Field (1996), 512×512 -pixel images were considered to generate the bank of filters: in this case, one may take $W = 256$.

We are now able to compute K . Note that, since the filters have an 11×11 support, the support of the generating kernel K projected onto the (x, y) -plane will be of size 21×21 and the supports of the kernels obtained by repeated integrations will widen progressively. Figure 18 displays the projection onto the retinal plane of the kernel around a filter, for four

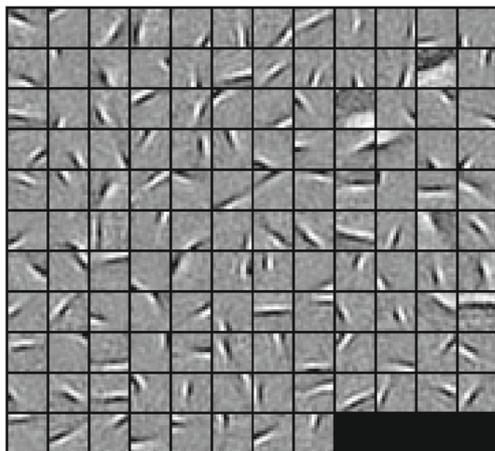


Fig. 17 A bank of 128 filters obtained by training a set of basis functions on natural images (see Olshausen and Field (1996)). The algorithm used is described in Lee et al. (2007)

different filters of the family. As in the Gabor example, the first projection (central column) shows at each (x, y) the maximum value (color-coded) of the kernel over all $f \in \mathcal{F}$; the second projection (right column) displays, at the points (x, y) where this value is above a threshold, a tiny version of the filter $\psi_{x,y,f}$ where the maximum is achieved. Even in this non-structured case, a “bow-tie” pattern along the preferred axis of the starting filter develops, with the orientations of the maximizing filters organizing across space in a way consistent with the association fields of Field et al. (1993). However, in this example the “side lobes” emerging from the interaction of parallel RPs were not discarded. A clipping similar to the one performed for Gabor filters may be reproduced e.g. by adopting techniques of connected component extraction; nonetheless, since the pattern obtained turns out to be consistent with the “ladder” effect described in Yen and Finkel (1998), truncating the kernel might mean ignoring some relevant information.

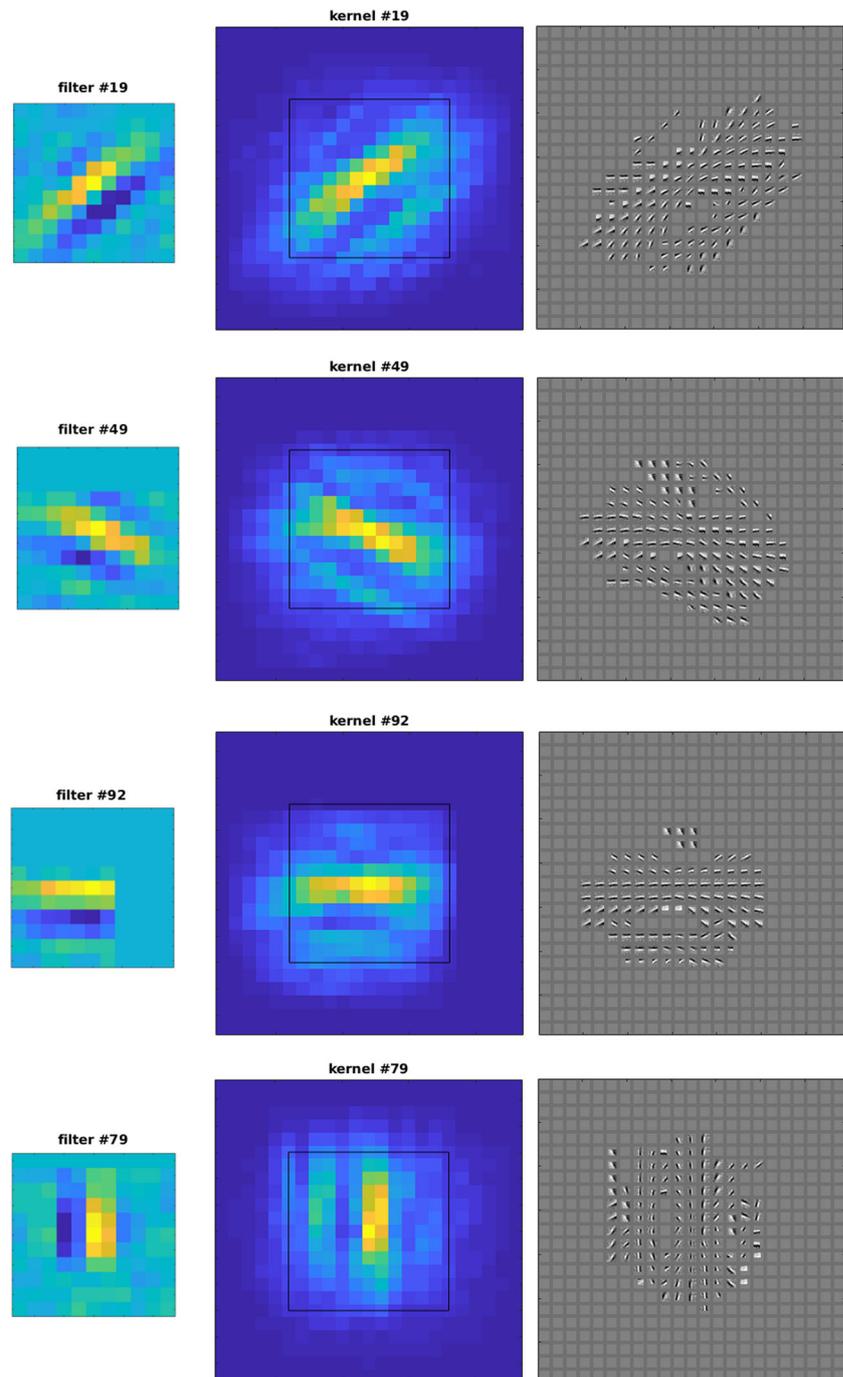
Note that in this case we only displayed bidimensional representations of the connectivity kernel. This is still possible since we constructed a parameterization of the family of filters in order to have a feature space of the form $position \times features$, so it makes sense to project the results on the (x, y) -plane. However, a visualization of the kernels in the space $\mathcal{G} = \{-W, \dots, W\}^2 \times \mathcal{F}$ would no longer be meaningful, since the set of indices $\mathcal{F} = \{1, \dots, 128\}$ is not ordered with respect to the distance.

A simple Matlab code to compute and display the connectivity kernel for a translation-invariant family of RPs is provided in Online Resource 2. The code is intended to provide visualizations of the kernel associated to such a bank of filters, as the ones contained in this paper. However, for other purposes requiring a greater amount of computation (as is the case for the future application sketched in the next paragraph) it is advisable to use tools that provide efficient computation of convolutions through optimized CPU and GPU implementations, written e.g. in C++ and CUDA (Karas and Svoboda 2013; Vedaldi and Lenc 2015; Paszke et al. 2017).

Future application to deep learning The kernel we defined specifies a connectivity pattern associated to a bank of filters according to their shape, with no need for any analytical formulation. In the case of Gabor-like RPs, the pattern we obtain is indeed compatible with the perceptual rules of grouping and the neurophysiological results on lateral connectivity in V1. Therefore, we believe that such a construction could be used to introduce some geometric prior knowledge in convolutional neural networks (CNNs) (LeCun et al. 2015), by way of *intra-layer* connections.

A number of recent works addressed the issue of inserting prior knowledge in CNNs in the form of group invariances. In a layer of a CNN, starting from a set

Fig. 18 Visualization of the kernel K computed around four filters $\psi_{0,0,f_0}$ ($f_0 = 19, 49, 92, 79$). In each of the four cases the image displays, from left to right: the original filter (already cropped around its center); the projection of the kernel on the (x, y) -plane obtained by taking the maximum in the variable f ; the same projection, where instead of the intensity value we displayed at each pixel the filter $\psi_{x,y,f}$ maximizing $K((x, y, f), (0, 0, f_0))$



of (trainable) filters $\{\psi_k\}_{k=1,\dots,n_f}$, invariance by spatial translations is assumed, so that from those n_f filters one gets a family $\{\psi_{x,y,k}\}$ where $\psi_{x,y,k}$ is the (x, y) -translation of $\psi_{0,0,k} \equiv \psi_k$. In Bekkers et al. (2018), this idea is extended to *rototranslations*: the family is thus augmented to $\{\psi_{x,y,\theta,k}\}$. This allows to obtain good results when testing on rotated inputs even when such examples were not included in the training set. In other words, “kernel augmentation” replaces data augmentation. The subsequent

layers consist then of group convolutions with learned kernels that live in the “lifted” space, and the output is finally projected onto \mathbb{R}^2 by taking the maximum in the orientation variable. For some more related work, see Sifre and Mallat (2013), Cohen and Welling (2016), and Worrall et al. (2017).

The idea that we propose can be regarded as a different modification of a CNN. Specifically, we propose that the activation of each convolutional layer be “convolved”

(possibly multiple times) with the kernel *associated to the filters of that layer*, cf. Eq. (13). This is actually a convolution only in the spatial variable. K plays the role of a “transition kernel” in the feature space, slightly modifying the output of the convolutional layer according to the *correlation* between its filters: the activation of a filter encourages the activation of other filters highly correlated with it. This operation should not be regarded as a deeper layer, but rather as an *intra-layer* phenomenon, in analogy with:

- (i) lateral connections in the layers of the visual cortex (and particularly the connectivity of V1, which indeed we propose to model through the kernel associated to simple cells);
- (ii) recurrent convolutional nets (RCNNs), where the lateral connections are modeled by recurrent weights (i.e. they are learned, and independent of the feedforward ones) (Liang and Hu 2015; Spoerer et al. 2017).

Since the application of K does not change the dimensionality of the space the outputs live in, one may simply start from a CNN and add these lateral connections after each convolutional layer, with no further changes in structure – differently from the group theoretical models described above. Moreover, unlike for RCNNs, the introduction of such intra-layer connections would not increase the number of parameters of the original CNN.

We believe that our construction might be useful for improving the performance of CNNs in tasks related to perceptual grouping, and we will address this challenge in our future work.

5 Conclusions and future prospects

In this paper, we proposed a description of the functional architecture of V1 based on a structure defined directly by the RPs of simple cells. Our construction is very general, since it does not take into account the parameterization of the bank of filters chosen to model the RPs: this makes it possible to define a cortical connectivity even starting from a randomly ordered set of filters resulting e.g. from a regression algorithm, and even in such a case it allows to recover association-field-like patterns.

As was said in Section 2.3, it has been hypothesized (Gilbert and Wu 2013) that each cortical area might be endowed with an intrinsic connectivity kernel defining a notion analogue to that of V1’s association fields. Our model might provide a way to express this kind of interactions in terms of the shape of the RPs of cells different from V1 simple cells, possibly in higher cortical regions.

With a view to future applications, we regard this technique as a possible way to introduce some geometric information on horizontal connections in the context of deep neural networks for artificial vision.

Acknowledgements The authors have been supported by Horizon 2020 Project ref. 777822: GHAI, PRIN 2015 “Variational and perturbative aspects of nonlinear differential problems”, and European Union’s Seventh Framework Programme ref. 607643: MAnET.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

- Abbasi-Sureshjani, S., Favali, M., Citti, G., Sarti, A., ter Haar Romeny, B.M. (2016). Curvature integration in a 5D kernel for extracting vessel connections in retinal images. *IEEE Transactions on Image Processing*, 2018(27), 606–621.
- Angelucci, A., Levitt, J.B., Walton, E., Hup, J.M., Bullier, J., Lund, J.S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22, 8633–8646.
- Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: Horizontal or feedback axons?. *Journal of Physiology Paris*, 97, 141–154.
- Anselmi, F., & Poggio, T. (2010). Representation learning in sensory cortex: a theory. CBMM memo n. 26.
- Antoine, J.-P., & Murenzi, R. (1996). Two-dimensional directional wavelets and the scale-angle representation. *Signal Processing*, 52(3), 241–272.
- Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 66, 937–404.
- August, J., & Zucker, S.W. (2000). The curve indicator random field: Curve organization via edge correlation. In Boyer, K., & Sarkar, S. (Eds.) *Perceptual organization for artificial vision systems*. Boston: Kluwer Academic.
- Barbieri, D., Citti, G., Sanguinetti, G., Sarti, A. (2014). An uncertainty principle underlying the functional architecture of V1. *Journal of Physiology Paris*, 106(5-6), 183–193.
- Barbieri, D., Cocci, G., Citti, G., Sarti, A. (2014). A cortical-inspired geometry for contour perception and motion integration. *J. Math. Imaging Vis.*, 49(3), 511–529.
- Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A.J., Pluim, J.P.W., Duits, R. (2018). Roto-translation covariant convolutional networks for medical image analysis. In Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G., Frangi, A.F. (Eds.) *Medical image computing and computer assisted intervention - MICCAI 2018 - 21st International Conference, 2018, Proceedings (pp. 440-448)*. (Lecture Notes in Computer Science; Vol. 11070).
- Ben-Shahar, O., Huggins, P., Izo, T., Zucker, S.W. (2003). Cortical connections and early visual function: intra- and inter-columnar processing. *J. Physiol. Paris*, 97(2-3), 191–208.
- Ben-Shahar, O., & Zucker, S. (2004). Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. *Neural Computation*, 16(3), 445–476.
- Bonfiglioli, A., Lanconelli, E., Uguzzoni, F. (2007). *Stratified Lie groups and potential theory for their sub-Laplacians*, Springer Monographs in Math. Berlin: Springer.

- Boscain, U., Chertovskih, R., Gauthier, J.P., Remizov, A. (2014). Hypoelliptic diffusion and human vision: a semi-discrete new twist. *SIAM Journal on Imaging Sciences*, 7(2), 669–695.
- Bosking, W., Zhang, Y., Schoenfield, B., Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6), 2112–2127.
- Bressloff, P.C., & Cowan, J.D. (2003). The functional geometry of local and long-range connections in a model of V1. *J. Physiol. Paris*, 97(2-3), 221–236.
- Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Wiener, M.C. (2002). What Geometric Visual Hallucinations Tell Us about the Visual Cortex. *Neural Computation*, 14, 473–491.
- Citti, G., & Sarti, A. (2006). A Cortical Based Model of Perceptual Completion in the Roto-Translation Space. *Journal of Mathematical Imaging and Vision archive*, 24(3), 307–326.
- Cohen, T., & Welling, M. (2016). Group equivariant convolutional networks. Int. Conf. on Machine Learning, 2990–2999.
- Cocci, G., Barbieri, D., Sarti, A. (2012). Spatiotemporal receptive fields of cells in V1 are optimally shaped for stimulus velocity estimation. *Journal of the Optical Society of America. A*, vol. 29, no. 1.
- Coifman, R.R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21, 5–30.
- Daugman, J.G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, A2, 1160–1169.
- Deng, C.-X., Li, S., Fu, Z.-X. (2010). The reproducing kernel Hilbert space based on wavelet transform, Proceedings of the 2010 international conference on wavelet analysis and pattern recognition. Qingdao, 370–374.
- Dobbins, A., Zucker, S., Cynader, M. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature*, 329(6138), 438–441.
- Duits, R. (2005). *Perceptual organization in image analysis: A mathematical approach based on scale orientation and curvature*. Phd thesis: Eindhoven University of Technology.
- Duits, R., Felsberg, M., Granlund, G., ter Haar Romeny, B.M. (2007). Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the euclidean motion group. *International Journal of Computer Vision*, 72(1), 79–102.
- Duits, R., & Franken, E.M. (2010a). Left-invariant parabolic evolutions on SE(2) and contour enhancement via invertible orientation scores, Part I: Linear Left-Invariant Diffusion Equations on SE(2). *Quarterly of Applied Mathematics*, 68, 293–331.
- Duits, R., & Franken, E.M. (2010b). Left-invariant parabolic evolutions on SE(2) and contour enhancement via invertible orientation scores, Part II: Nonlinear left-invariant diffusions on invertible orientation scores. *Quarterly of Applied Mathematics*, 68, 255–292.
- Duits, R., Führ, H., Janssen, B., Bruurmijn, M., Florack, L., van Assen, H. (2013). Evolution Equations on Gabor Transforms and their Applications. *Applied and Computational Harmonic Analysis*, 35(3), 483–526.
- Duits, R., Boscain, U., Rossi, F., Sachkov, Y. (2014). Association Fields via Cuspless Sub-Riemannian Geodesics in SE(2). *Journal of Mathematical Imaging and Vision*, 49(2), 384–417.
- Elder, J.H., & Goldberg, R.M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*, 2(4), 5,324–353.
- Favali, M., Citti, G., Sarti, A. (2017). Local and Global Gestalt Laws: A Neurally Based Spectral Approach. *Neural Computation*, 29(2), 394–422.
- Federer, H. (1969). *Geometric measure theory*. Berlin: Springer-Verlag.
- Field, D.J., Hayes, A., Hess, R.F. (1993). Contour integration by the human visual system: evidence for a local association field. *Vision Res*, 33, 173–193.
- Geisler, W.S., Perry, J.S., Super, B.J., Gallogly, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41, 711–724.
- Gilbert, C.D., Das, A., Ito, M., Kapadia, M., Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences USA*, 93, 615–622.
- Gilbert, C.D., & Wu, L. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14, 350–363.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38(2), 141–171.
- Hansen, T., & Neumann, H. (2008). A recurrent model of contour integration in primary visual cortex. *J. of Vision*, 8(8), 1–25.
- Hausdorff, F. (1918). Dimension und äusseres Mass. *Mathematische Annalen*, 79(1-2), 157–179.
- Hubel, D.H., & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J. Physiol. (London)*, 160, 106–154.
- Jones, J.P., & Palmer, L.A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1233–1258.
- Kapadia, M.K., Westheimer, G., Gilbert, C.D. (1999). Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proc Natl Acad Sci USA*, 96, 12073–12078.
- Karas, P., & Svoboda, D. (2013). Algorithms for efficient computation of convolution, in Design and Architectures for Digital Signal Processing, InTech.
- Kruger, N. (1998). Collinearity and parallelism are statistically significant second order relations of complex cell responses. *Neural Processing Letters*, 8, 117–129.
- Lawlor, M., & Zucker, S.W. (2013). Third-order edge statistics: contour continuation, curvature, cortical connections. In Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (Eds.) *Advances in neural information processing systems 26*, (Vol. 26 pp. 1763–1771). Red Hook: Curran Associates, Inc.
- LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- Lee, H., Battle, A., Raina, R., Ng, A.Y. (2007). Efficient sparse coding algorithms. In *Proceedings of the 19th annual conference on neural information processing systems* (pp. 801–808). Cambridge: MIT Press.
- Lee, T.S. (1996). Image Representation Using 2D Gabor Wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10.
- Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition, CVPR.
- Martinez, L.M., & Alonso, J.-M. (2003). Complex receptive fields in primary visual cortex. *The Neuroscientist*, 9(5), 317–331.
- Mitchison, G., & Crick, F. (1982). Long axons within the striate cortex: their distribution, orientation, and patterns of connection. *Proceedings of National Academy of Sciences USA*, 79, 3661–3665.
- Montobbio, N., Sarti, A., Citti, G. (2017). A metric model for the functional architecture of the visual cortex (submitted).
- Montgomery, R. (2002). A tour of subriemannian geometries, their geodesics and applications, *Mathematical surveys and monographs*, Vol. 91 American mathematical society, Providence, RI.
- Mumford, D. (1993). *Elastica and computer vision*. In Bajaj, C. (Ed.) *Algebraic geometry and its applications* (pp. 507–518). Berlin: Springer-Verlag.

- Neumann, H., & Mingolla, E. (2001). Computational neural models of spatial integration in perceptual grouping. In Shipley, T.F., & Kellman, P.J. (Eds.) *Advances in psychology, 130, From fragments to objects: Segmentation and grouping in vision*, 353–400.
- Olshausen, B.A., & Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Lerer, A.A. (2017). Automatic differentiation in pytorch, in NIPS-W.
- Petitot, J. (2008). Neurogéométrie de la vision - Modèles mathématiques et physiques des architectures fonctionnelles, Éditions de l'École Polytechnique.
- Petitot, J., & Tondut, Y. (1999). Vers une neuro-géométrie. Fibrations corticales, structures de contact et contours subjectifs modaux, Mathématiques, Informatique et Sciences Humaines, vol. 145, 5–101. EHESS, Paris.
- Sanguinetti, G., Citti, G., Sarti, A. (2010). A model of natural image edge co-occurrence in the roto-translation group. *Journal of Vision* 10(14).
- Sarti, A., & Citti, G. (2015). The constitution of visual perceptual units in the functional architecture of V1. *Journal of Computational Neuroscience*, *38*(2), 285–300.
- Sarti, A., Citti, G., Petitot, J. (2008). The symplectic structure of the visual cortex. *Biological Cybernetics*, *98*(1), 33–48.
- Sifre, L., & Mallat, S. (2013). Rotation, scaling and deformation invariant scattering for texture discrimination, CVPR, IEEE 1233–1240.
- Sigman, M., Cecchi, G.A., Gilbert, C.D., Magnasco, M.O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, *98*(4), 1935–1940.
- Spoerer, C.J., McClure, P., Kriegeskorte, N. (2017). Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in psychology*, *8*, 1551.
- Sturm, K.-T. (1995). On the geometry defined by Dirichlet forms. In Bolthausen, E. et al. (Eds.) *Seminar on stochastic analysis, random fields and applications* (pp. 231–242). Boston: Birkhäuser.
- Sturm, K.-T. (1998). Diffusion processes and heat kernels on metric spaces. *Annals of Probability*, *26*(1), 1–55.
- Vedaldi, A., & Lenc, K. (2015). MatConvNet - convolutional neural networks for MATLAB, Proc. of the ACM Int. Conf. on Multimedia.
- Worrall, D.E., Garbin, S.J., Turmukhambetov, D., Brostow, G.J. (2017). Harmonic networks: Deep translation and rotation equivariance. CVPR 5028–5037.
- Yen, S.C., & Finkel, L.H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Res*, *38*(5), 719–741.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.