



The  
University  
Of  
Sheffield.

**The shape of the PCA trajectories and the  
population neural coding of movement initiation  
in the basal ganglia**

by:

*Alejandro Jimenez Rodriguez*

A thesis submitted in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

The University of Sheffield  
Faculty of Science  
Department of Psychology  
June 14, 2019

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>The basal ganglia and neural codes</b>	<b>8</b>
2.1	Basal ganglia, population codes and neural space analyses . . . . .	9
2.1.1	The basal ganglia . . . . .	9
2.1.2	Architecture . . . . .	10
2.2	Neural codes and ensembles . . . . .	13
2.2.1	Heterogeneity sparseness . . . . .	14
2.2.2	Neural code and computations in the basal ganglia . . . . .	14
2.3	The neural trajectories . . . . .	15
2.4	Related work . . . . .	17
<b>3</b>	<b>Mathematical preliminaries</b>	<b>19</b>
3.1	Algebra and geometry basics . . . . .	20
3.1.1	Vector spaces . . . . .	21
3.1.2	Groups, maps and actions . . . . .	22
3.2	Singular Value Decomposition and Principal Components Analysis .	23
3.2.1	Characterization and interpretations of PCA . . . . .	24
3.2.2	PCA of random matrices . . . . .	28
3.2.3	Functional PCA . . . . .	29
3.3	Differential geometry . . . . .	30
3.3.1	Parametrization . . . . .	30
3.3.2	Frenet frame . . . . .	30
3.3.3	Contact of curves: Vertices and flat regions . . . . .	31
3.4	Semiorders . . . . .	32
3.5	Heat equation and diffusion . . . . .	32
3.5.1	The laplacian . . . . .	33
3.5.2	Relationship with the heat kernel . . . . .	35
3.6	Further reading . . . . .	35

<b>4</b>	<b>Description of the physiological data</b>	<b>36</b>
4.1	Experimental setup and methods . . . . .	37
4.1.1	Data collection . . . . .	37
4.1.2	Task structure . . . . .	37
4.1.3	General methods . . . . .	37
4.2	Overview of the PCA information for the basal ganglia around movement initiation . . . . .	39
<b>5</b>	<b>The geometry of neural trajectories</b>	<b>41</b>
5.1	Chapter overview . . . . .	41
5.2	Preliminary aspects of the shape of the principal trajectories . . . .	42
5.2.1	The geometric nature of the neural trajectory . . . . .	42
5.2.2	The low dimensional PCA trajectories can be irregular . . . .	45
5.2.3	Geometric pre-processing: reparametrization and smoothing . . . .	47
5.3	The curvature . . . . .	53
5.3.1	Simulation of specific patterns of curvature . . . . .	56
5.4	The shape of the neural trajectories . . . . .	65
5.4.1	Low dimensional invariance . . . . .	65
5.4.2	The form of the trajectory: salient features. . . . .	69
5.5	Null shapes . . . . .	75
5.5.1	PCA of gaussian SDFs . . . . .	75
5.5.2	Two models of autocorrelation . . . . .	78
5.5.3	The basal ganglia trajectories behave like laplacian eigen- functions . . . . .	85
<b>6</b>	<b>Population interpretation and neural ensembles</b>	<b>90</b>
6.1	Chapter overview . . . . .	91
6.2	Neural ensembles are important for the understanding of curvature . . . .	93
6.3	Preliminary insights in the hierarchical nature of PCA . . . . .	93
6.4	The induced system of hyperplanes . . . . .	95
6.4.1	A change in perspective: neural ensembles as faces of the hyperplane arrangement . . . . .	97
6.5	Catalan objects: overlap trees and the neural semiorder . . . . .	100
6.5.1	The difference tree . . . . .	100
6.5.2	The compound tree . . . . .	104
6.5.3	Coding tree . . . . .	105
6.6	The semiorder gives information about the embedding of neural low dimensional subspaces in the high dimensional space . . . . .	107
6.7	Mean subtraction and the center surround interpretation . . . . .	108

<b>7</b>	<b>Neural ensembles and trajectory shape in the basal ganglia</b>	<b>112</b>
7.1	Chapter overview . . . . .	113
7.2	The basal ganglia output behaves as a low dimensional gate for movement . . . . .	113
7.2.1	The anatomy of the principal trajectories suggests characteristic time resolved ensemble structures for each region . .	114
7.2.2	The observed low dimensionality comes from separated but homogeneous clusters . . . . .	117
7.2.3	Heterogeneity and sparseness measures support the dimensionality reduction view of the basal ganglia . . . . .	118
7.3	Movement initiation occurs through parallel channels with specific functional roles at the basal ganglia output . . . . .	123
7.3.1	The singularities are associated with orthogonal subspaces of the population trajectory . . . . .	123
7.3.2	The population dynamics in the SNr is driven by minimally overlapping ensembles . . . . .	123
7.3.3	The third PC can be used to separate the population into different channels with specific computational roles . . . . .	128
7.3.4	The STN is mostly excitation with low dimensional patterns that reflect mixed selectivity . . . . .	132
7.4	Other regions and features . . . . .	134
7.5	Appendix: The growing of the trees . . . . .	135
<b>8</b>	<b>Discussion</b>	<b>144</b>
<b>9</b>	<b>Code</b>	<b>148</b>
9.1	Population api . . . . .	148
9.2	Curve api . . . . .	148
9.3	TameD library . . . . .	148
9.4	Analyses . . . . .	148
	<b>Bibliography</b>	<b>149</b>

# Chapter 1

## Introduction

Neural trajectory analysis has become widespread in neuroscience as a means to investigate population phenomena that transcend the single neuron [34, 81]. The interpretation of the results, however, remains controversial [31]. Given the complexity and magnitude of the data, it is not clear what aspects of the trajectory are side effects of the method of choice and which are informative of the underlying mechanisms of the studied behaviour.

The ideas behind neural trajectories are closely related to that of dimensionality reduction [24]. They have been useful, for example, in situations in which a few surrogate variables seem to describe the *dynamics* of the population [13]. The underlying concepts in this case come from dynamical systems theory; the space is the phase space of some system whose (partial) behaviour is observed. In this sense, a dynamical approach is reasonable because, given observations of the dynamical system (i.e. the recorded neuron's activity), it is possible to reconstruct its evolution law by means of linear algebra techniques [68].

This approach, even though has been instrumental in gaining insights about neural population dynamics in some other instances [21], is unsatisfactory or, at least, incomplete for 2 reasons. Firstly, it assumes some sort of asymptotic behavior, like an attractor, towards which the system tends. This is made patent in some of the methods in which a proper dynamical system is fit to the low dimensional trajectories and used as a surrogate of the original population [21]. This however is not the case in most experimental situations in which only transients in response to an external stimulus are observed [58].

Secondly, and more important, the low dimensional trajectories in the case of PCA, which are the focus of this work, do not meet the assumptions that most dynamical system approaches require. On one hand, they display pathological features like irregularities (the curve stops or backtracks) and they can be non-simple (i.e. have self-crossings). On the other hand, the trajectories' shape has natural constraints that come from the PCA computation itself. In this work we spend a good deal of effort in studying the normal and pathological aspects of the shape and elucidating the mentioned restrictions.

The study of trajectories faces then major difficulties in the interpretation that go along with wrong intuitions, for example, about the relationship between the individual firing rates and the geometry. When seeing a sharp turn in the trajectory, most people will be inclined to think that they come from an equally sharp pattern in the firing rate. This is not true in most of the cases as we explain in this work. We show that the curvature and other aspects of the trajectory depend critically on the nature of the overlap between the firing patterns.

Along with the study of the trajectories, other related mathematical concepts have come to the attention of the neuroscience community. That is the case of subspaces. They have become effectively the unit of computation proposed for

some mechanisms like the preparation of movement [48]. To this end, important questions to be asked are whether the underlying subspaces are orthogonal or not [81]. In this respect, orthogonal subspaces enjoy the property of minimizing the errors when decoding different messages [23].

Hence, an effort is also made in understanding the high dimensional aspects of the trajectory and the underlying subspaces in the basal ganglia. It is widely known that phenomena in high dimensions can be counter intuitive and paradoxical at best [28]. However, it has been recently recognized that despite the curses of dimensionality, a dominant feature of high dimensional phenomena, that of clustering, can make them remarkably simple [65]. The clustering endows the high dimensional space with a tree structure that, in the case of PCA, is instrumental in the understanding of the individual principal components and their relations.

Instead of applying the dynamical system approach directly, we use then the more pragmatic view of neural ensembles [5, 55]. Neuronal assemblies or ensembles turn out to be very convenient in relation to PCA as linear dependence of the inputs influence the underlying PCA subspaces. We combine our study with the development of a framework as rigorous as possible from the mathematical point of view. Such framework leverages state of the art techniques drawn from differential geometry, combinatorics, analysis and algebra that are widely available to mathematicians but might be unknown for other scientist.

The focus of this study is on the neural populations of the basal ganglia for which no current description of the trajectories is known. We study the population activity around movement initiation and use the insights developed about the shape and ensemble (clustering) structure to understand aspects of the code and computation performed mainly at the output stage. We propose a computational interpretation based on parallel channels with minimal cross talk for a well learned stereotypical behavior. This makes sense as the process of learning might be thought as a process of dimensionality reduction that finds orthogonal subspaces for the movement involved minimizing the errors in the execution [7].

An outline of this work is as follows. In chapter 2 we introduce the main aspects related to the basal ganglia and neural codes that will be important in the understanding of this work. Chapter 3 develops the main mathematical themes recurring in this thesis. The three following chapters are the main results chapters. Chapter 5 develops an in depth study of the shape of the trajectories from different aspects and theoretical frameworks. In chapter 6 we occupy ourselves in the development of a major connection between PCA and some combinatorial objects called Catalan objects. Finally, chapter 7 provides a complete analysis of the population dynamics using the framework and tools just developed.

## Chapter 2. Neural codes and Basal Ganglia

## Chapter 2

### The basal ganglia and neural codes

## 2.1 Basal ganglia, population codes and neural space analyses

### 2.1.1 The basal ganglia

The basal ganglia comprise a highly preserved collection of subcortical nuclei involved in a variety of motor and non motor functions. In this section we focus on general structural and functional motor related aspects mostly in the rodent and macaque instances of such regions.

#### Decision making

The basal ganglia have a clear role in decision making that has been evidenced by several studies[60]. The action selection hypothesis of the basal ganglia[76] leverages the structural properties of the inhibitory networks in the basal ganglia to suggest a winner take all mechanism for selection of the desired action outcome. This is accompanied by a suppression of unwanted options in a center surround-like organization similar to the observed in the visual system[60].

Basal ganglia inspired models of decision making can be readily implemented suggesting that the proposed mechanisms are, to viable from a functional perspective [54, 39]. This view has been challenged, or at least complemented, by evidence of complementary roles of the basal ganglia with respect to the motor/frontal cortices. One account suggests that the output nuclei provide a vigor[92] signal that controls the movement gain.

From a phenomenological point of view, it is clear that neurons in the basal ganglia show phasic and tonic activity that is modulated by movement and by the behavioral outcome. Specific cell subtypes show a variety of responses to movement [33] some of them directly related to the decision target. Activity in those neurons has been related to other movement related aspects like urgency [91] and action suppression [83, 82]

#### Other functions

The basal ganglia are heavily involved in habit formation [99] and, in close relationship, in repetitive behaviors associated with tics [61] and higher order dysfunctions like Obsessive Compulsive Disorder [63] or addiction. The functional aspects that intervene in those behaviors are central in the ability of sequence learning, encoding and, potentially to a lesser extend, execution [42]. The sequence learning aspects of the basal ganglia is a rich area that is beyond the scope of this work.

Dopaminergic, serotonergic and collinergic signals modulate learning and motivation in the corresponding behavioral output. There is evidence of value and reward related information encoding in some of the nuclei (see next sections). Dopaminergic input is instrumental not only for the learning aspects of the task

but also for the smooth execution of the movements; indeed, dopamine dysregulation is the universally recognized substrate of Parkinsonian symptoms in patients [77].

Finally, the functionality associated with these regions seems to be recruited abstractly in higher order non motor functions like language production [10], executive functions, working memory among others. For example, higher level of abstraction in action execution seem to be achieved by action chunking or representing full sequences as a single entity at the same level of the single movement [36].

### 2.1.2 Architecture

The basal ganglia have appealing architectural features that have attracted all sort of computational hypotheses. A manifest aspect of their connectivity is that it is largely inhibitory and its ultimate effect in downstream regions is that of a disinhibition [19].

The basic architectural plan is shown in figure 2.1. The inputs to the basal ganglia come mainly from the cortex and some nuclei in the thalamus, apart from the modulatory input from the dopaminergic centers in the Substantia nigra pars compacta and the ventral tegmental area. In the striatum, two pathways are recognized. The direct pathway sends inhibitory projections to the output stages, internal globus pallidus (GPi) and the Substantia nigra pars reticulata (SNr). The indirect pathway projects to the external globus pallidus which, in turn projects to the output stages through the STN.

Note that the indirect pathway has as a ultimate result, when activated (i.e. neurons fire), is a disinhibition of the SNr while the direct pathway's inhibits the output stages. Also notable about the indirect and direct pathways is the existence of differences in the kind of receptor expressed by their projection neurons (D2, D1 subtypes respectively[85]); they also have other neurochemical differences, notably, enkephalin vs. substance p neuropeptide expression. Finally, a hyperdirect pathway provides fast excitation to the different nuclei by means of the subthalamic nucleus (STN) [66].

Another architectural property of the basal ganglia is its organization in parallel nested loops [75]. For example, the outputs of the basal ganglia that target the thalamus are partially fed back to the striatum by structures like the parafascicular nucleus, which we study in this work. Also, in a broader loop, the thalamo-cortical pathways feed back to the striatum through the mentioned cortical input. The parallel aspect of the loops is, as shown in figure 2.2, the segregation of the functions of the basal ganglia in different cognitive domains related with different functions.

Finally, a conspicuous feature of the inter nuclei connectivity is that of funneling [64]. The afferents from the cortex are much more numerous than the targets in the

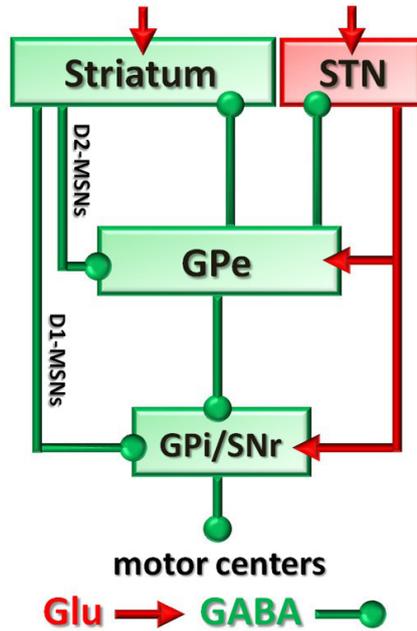


Figure 2.1: Gross architecture of the basal ganglia. The input stages Striatum and STN receive excitatory input from the cortex. Inhibitory pathways are shown in green, excitatory ones in red. See text for details (Taken from [26])

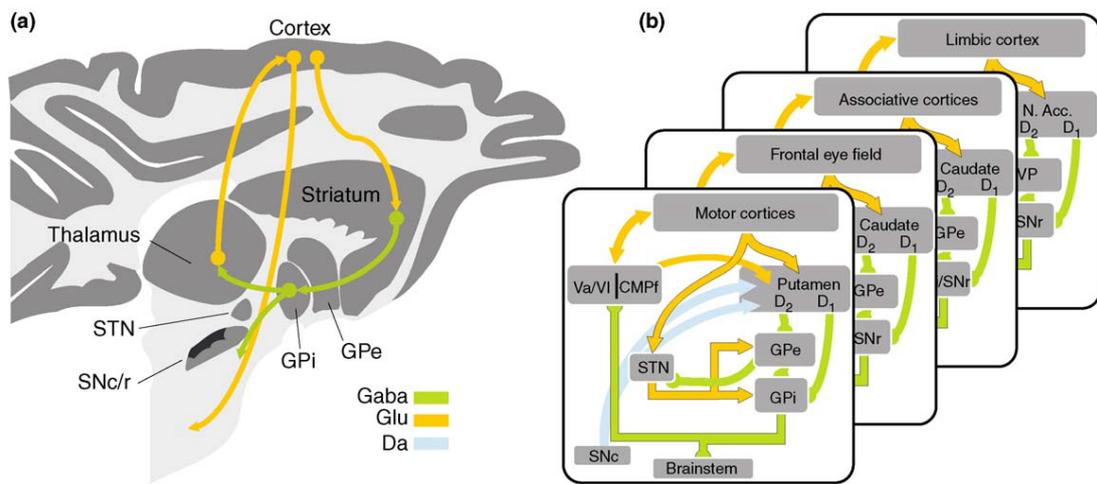


Figure 2.2: Parallel and nested loops in the basal ganglia. To the left, the basic network architecture is shown. To the right, different domains have loops with similar architecture but different functional and behavioral outcomes. (Taken from [92])

striatum. In turn, the connections to the following structures in the feedforward neuron display reductions that are thought to be instrumental in the compression and decorrelation of the input patterns [97]

### Input stage: The striatum

The input stage to the basal ganglia is the striatum. The main cell type in this region are the Medium spiny neurons (MSNs) which show sparse patterns of firing, broad arborization, collateral but sparse inhibition; these are the projection neurons. The second group of cells are inhibitory interneurons called Fast Spiking Interneurons (FSIs) which are thought to directly drive the MSNs by their fast firing rates. Finally there are other types of neurons (like TANs) whose study goes beyond the scope of this work.

The striatum has different domains which go from motor (dorsal) to limbic (ventral) and shows somatotopy to some extent [66], and has spatial organization into domains called striosomes with and external matrix called matrix [85]. With respect to movement initiation, principal neurons in the dorsolateral striatum differentiate between contra and ipsilateral movement by a fast increase in firing rate [83], moreover, the FSIs show diverse individualized patterns of activity which differ according to the nature of the decision (ipsi or contralateral)[33]. Some of the striatal movement related neurons show later modulations in the firing rate due to potential stop or pause signals coming from the GPe [82].

### The Globus pallidus and the Subthalamic Nucleus

The subthalamic-pallidal network, globus pallidus external segment (GPe) and subthalamic nucleus (STN), is characterized by high firing rates in its principal neurons ( $\approx 40$ spikes/second in the GPe and  $\approx 20$  spikes/second in the STN [85]). This network is usually thought as an oscillator that generates the  $\beta$  rhythms and their response to movement initiation is better thought as a synchronization/de-synchronization of the oscillation in this frequency band [53, 62].

The GPe nucleus is inhibitory with inhibitory feedforward, convergent connections to the STN and the output nuclei, inhibitory collaterals with high influence on proximal neurons and some backward projections to the striatum[85]. Neurons in the STN have fewer collaterals and their influence is excitatory with divergent connectivity onto the GPe. Two different cell types stand out in the GPe neurons. Prototypical neurons project to the STN and SNr/GPi in the patterns described before. Arky-pallidal neurons are the source have dense connectivity with striatal interneurons and are thought to cancel actions in three-way model of movement cancellation: hyperdirect stop signals provide a pause mechanism that allows the

continuation of the started action in some occasions while the akrypalliday pathway is the source of a stronger, definitive cancellation of movements [82].

**Output:** The substantia nigra/ GPi

The influence of the basal ganglia is exerted through the output nuclei: the substantia nigra pars reticulata and the globus pallidus internal segment; both differ in the output targets (i.e. different thalamic nuclei, superior colliculus among others for the SNr; lateral habenula, other subcortical nuclei, etc. for GPi) but have similar neurophysiological responses [85]. The firing rate response to cortical excitation is characterized by an initial excitation, a pause and a subsequent excitation [85]. Firing rate pauses are the hallmark of the disinhibition model of basal ganglia with the intermediate pause dominated by the striatal direct pathway [19].

**The parafascicular nucleus**

The thalamus is the target but also the source of many topographically organized afferents to the striatal network. Inputs coming from the central lateral nucleus target the spines of MSNs, have high firing rates and exhibit bursts, while inputs coming from the parafascicular nucleus target mainly the dendritic shaft of interneurons in the striatum along with spines of MSNs, have lower firing rates and are less bursty[85].

## 2.2 Neural codes and ensembles

The idea that neurons form highly connected groups is central to the neuron doctrine [8]. They were first introduced in the seminal work of donald hebb [40] and, since then, evidence of them as a functional and organizational principle of the neural code have been found in different brain regions [16]

Neurons in the brain encode information at the single neuron level, for example in their firing rate (rate code) or in the spike timing (temporal code). When the information is encoded by a group of neurons, the code is called a population code. A famous example of such code is the *population vector* in which, it was shown, linear combinations of neuron activities in the motor cortex could be used to decode the direction of movement in monkeys [35]; this class of readouts are called linear decoders.

The responsiveness of a neuron to a feature of a stimulus or an event[71], like direction or orientation, is called *selectivity*. Sensory codes are usually studied creating what are called *tuning curves*[15], however, such procedure is not possible or well defined for other brain regions like the striatum. Neurons display, additionally, a feature called mixed selectivity [78], which implies that a neuron may

respond to different features or in different aspects of the task. Mixed selectivity justifies the novel analysis developed in this work as neurons participating in an ensemble at some point in time might be regrouped for later aspects of the neural processing.

### 2.2.1 Heterogeneity sparseness

Two main features, apart from selectivity, of a neural code are heterogeneity and sparseness. Heterogeneity is related to the fact that brain regions are diverse. Firstly, there are a variety of cell types with differing chemical and morphological features. Secondly, even in morphologically homogeneous networks like the cortex, firing patterns are diverse and shared with neurons within ensembles [45]. This heterogeneity has implications in the population code: it limits the usefulness of linear readouts [84].

Some measures of heterogeneity or diversity have been developed in the context of graph theory and population dynamics, many of them, like the Simpson’s D index [69] use the Shannon entropy in their formulation. A neuroscience example is found in [89] where a similar measure like the one proposed in this work is used to characterize the population activity. It turns out that the main feature of a diversity measure is convexity [74] and its measurement is related to the dimensionality of the population in question.

Sparseness or sparsity, on the other hand, is related to the number of neurons being active or participating in the encoding of a certain feature. In studies of the receptive fields in the visual system, it was found that the emerging patterns of activity were sparse [67]; indeed many other regions are naturally sparse (like the MSNs in the striatum) while others are characterized by a more dense firing.

Some measures of sparseness have been developed before. An ideal measure of sparseness will maximize the contrast in the pattern of firing, for example by using the  $L_1$  norm of the population vector. A widely used measure is the *treves-roll* sparseness [96] defined as

$$sp(X) = 1 - \frac{1}{n} \frac{[\sum_i r_i]^2}{\sum_i r_i^2}, \quad (2.1)$$

where  $r_i = \|x_i\|_1 = \sum_k |x_{ik}|$  is the 1-norm of the neuron’s  $i$  z-score. This index has  $[0, 1]$  as its range. Another alternative is to use entropy again and the fact that it is a measure of the *width* of a distribution.

### 2.2.2 Neural code and computations in the basal ganglia

One of the main computations ascribed to the basal ganglia is that of dimensionality reduction [7], more precisely, reinforcement driven dimensionality reduction.

In this framework, the funneling architecture presented before has as an ultimate consequence the emergence of patterns of activity in the output regions that closely resemble PCA. Indeed, neural networks with similar Converging (Ctx  $\rightarrow$  SNr) - diverging (SNr  $\rightarrow$  Ctx) architecture have been shown to *span* the same low dimensional subspace than the principal components[12].

This dimensionality reduction aspect of computation is closely related to the compression or “chunking of action repertoires” [36] that is observed in this regions. Indeed, ensembles of MSNs in the striatum have been shown to encode specific actions or behaviors [49]. Ensemble or assembly formation is one of the observed features of striatal spatio temporal dynamics [72] and it is natural to assume that the topology of locally compact bumps of activity is preserved in some way downstream in the basal ganglia hierarchy.

Based upon the timing and polarity (either increases or decreases) of the responses at the different stages in the basal ganglia, a model of the computations performed in the selection of an action is that of a center-surround processing in 3 steps[66]. It has been known that the inhibitory networks are ideal for competition and selection in a winner-take-all fashion, however, this model provides specific roles for each for the three main pathways in these nuclei.

Firstly, a fast, broad excitation to the output regions (SNr/GPi), through the hyperdirect pathway, shuts all the potential actions down by inhibiting their targets in the thalamus. Secondly, a focused inhibition through the direct pathway “frees” the targets corresponding to the desired action while holding the undesired ones; finally an excitation coming through the indirect pathway restores the inhibitory hold on the thalamus. Although most likely oversimplified, this model provides a physiologically realistic framework for understanding the observed firing patterns (figure 2.3).

## 2.3 The neural trajectories

The neural trajectory formalism has been widely used in neuroscience to understand neural activity and population coding from a dynamical systems perspective. In the neural space each coordinate axis corresponds to the firing rate of one neuron and the population activity at certain time corresponds to a point in that  $n$  dimensional space. The activity in a given interval of time traces a curve associated with the dynamics of the population in relation to a given internal or external event[34, 24].

The neural trajectories appear, in general, to be constrained to a low dimensional manifold of the high dimensional space, therefore, they can be studied by constructing appropriate surrogate variables obtained either by constructing linear combinations (i.e. projecting them in a low dimensional space), finding hidden

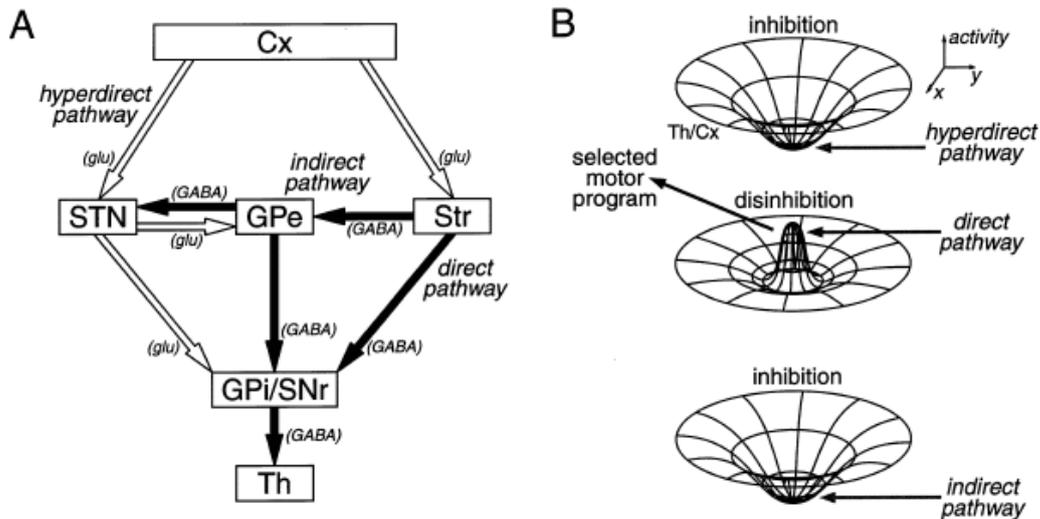


Figure 2.3: Center surround model of action selection in the basal ganglia. Taken from [66]. A. The three pathways and B. The different roles played based upon the timing of their responses

variables or reconstructing the said manifold by local approximations (i.e. LLE) [24].

The low dimensional activity, at first, seems to be in direct contradiction of the mixed selectivity - state dependent computation framework [78], in there, the neural activity must live in a high dimensional space in order to be able to perform rich computation tasks like discrimination. Such contradiction however disappears when noticing that such manifold is embedded in a very high dimensional space and its low dimensionality is a global property (although, locally, the manifold behaves as the high dimensional space itself); it comes from constraints imposed on the dynamics by connectivity, computation and the existence of neural ensembles[57].

Functional substrates of those manifolds have been found in a variety of brain regions and systems. For example, what are called rotational patterns have been associated to the coding of movement parameters like direction or preparation in the motor cortices [21]. The trajectory of the neural population has been associated to decision making tasks in which the trajectories diverge at the moment the decision is made [13]. Classical work also associate the visited points of the trajectory with the identity of distinct odors in the locust brain[58].

This last study gave rise to a computational interpretation of the trajectory in terms of *saddle points* instead of attractors. In that framework, the trajectory visits a sequence of states that are saddle points in the evolution of the dynamical system but it does not stop or “tends” to a specific point in the limit. This set of

states are the starting point of our view of the trajectories in terms of ensembles or faces of the high dimensional hyperplane arrangement. Indeed, an approach that has been fruitful in the past is to keep track of the activated ensembles along the trajectories, for example, to decode sleep activity in the *c. elegans* [47].

Churchland, Cunningham, Null space [48]

## 2.4 Related work

Principal Components Analysis (PCA, see next section) has been widely used in the study of neural trajectories as was elaborated in the previous section [13, 24]; the extraction of low dimensional trajectories by other means is reviewed elsewhere. The traditional analysis can be done either by averaging over all trials for a given condition as is done in this work, extending the method to higher order structures like tensors [95] or partially averaging over different conditions to de-mix condition specific patterns [50]. We refer the reader to the original paper for details about these more specialized variations of PCA.

The approach taken in this work is that of neural ensembles. PCA has been related to clustering algorithms like k-means before [27] although their approach is not as rigorous and the relation is at best restricted to specific cases. The identification of ensembles based on the components of the eigenvectors has thoroughly studied in [17, 55]; they concluded that, although it is possible to use them for that purpose, it is not clear how the ensemble information is mixed in subsequent PCs, a question for which our work proposes an answer.

The key of our answer is an hierarchical approach to disentangling the multiple interactions between ensembles. Similar observations have led to the development of algorithms to identify highly correlated groups which in other fields are called communities [56]. Those algorithms are based in a sequence of thresholdings of the covariance matrices in the process of the construction of the nested groups in a similar guise to ours. Detailed studies of the commonalities in high dimensional data has led researchers to the conclusion that high dimensional spaces tend to have a natural hierarchical structure that can be exploited in favor of better methods of analysis [52].

As a final remark, it is important to note that the field of neuroscience is not the only one that has realized the usefulness of PCA in the construction of low dimensional approximations. Notable mentions of such methods are the *essential dynamics* of proteins [4] and the *proper orthogonal decomposition* in the study of turbulence [18]. Both methods use PCA or its probabilistic counterpart, the Karhunen-Loeve transformation to identify an essential subspace that simplifies the otherwise complex dynamics of either the chemical reactions or the coherent structures in the fluid.

## Chapter 3. Mathematical preliminaries

## Chapter 3

### Mathematical preliminaries

### 3.1 Algebra and geometry basics

In this section we develop basic terminology and concepts that will be used in the treatment of PCA from the algebraic point of view. The most important concept in this section is that of eigen-decomposition as the principal components are defined in terms of the eigenvectors and eigenvalues of the *sample covariance matrix*[43].

A  $n$  dimensional vector is a  $n$  tuple  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  of numbers called scalars, with  $v_i \in \mathbb{R}$  or  $\mathbb{C}$ . The space of all real vectors is denoted  $\mathbb{R}^n$ . A matrix is an arrangement  $\mathbf{A}$  of scalars in  $n$  rows and  $m$  columns, where the element in the row  $i$  and the column  $j$  is denoted  $\mathbf{A}_{ij}$ . The space of all matrices is denoted by  $\mathbb{R}^{n \times m}$ . The matrix  $\mathbf{B}$  such that each  $\mathbf{B}_{ij} = \mathbf{A}_{ji}$  is called the transpose of  $\mathbf{A}$  and is denoted by  $\mathbf{A}^T$ . The columns of  $\mathbf{A}$  are denoted by  $\mathbf{A}_{*j}$  and the rows are  $\mathbf{A}_{i*}$ .

A linear combination of the vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  is defined as the sum  $\mathbf{u} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \dots + \alpha_m \mathbf{v}_m$ , with  $\alpha_i \in \mathbb{R}$ . The vectors are called linearly independent if  $\mathbf{u} = 0$  if and only if <sup>1</sup>  $\alpha_i = 0$  for all  $i$ . The *span* of a set of vectors is the set of all their linear combinations. A line is the span of a single vector and a ray in the direction of a vector  $\mathbf{v}$  is the set of all  $\mathbf{u} = \alpha \mathbf{v}$  with  $\alpha \geq 0$ .

The multiplication of a matrix and a vector can be seen as a linear combination of its columns:

$$\mathbf{A}\mathbf{v} = v_1 \mathbf{A}_{*1} + v_2 \mathbf{A}_{*2} + \dots + v_m \mathbf{A}_{*m}$$

And matrix multiplication  $\mathbf{C} = \mathbf{A}\mathbf{B}$  can be seen as the result of multiplying  $\mathbf{A}$  by each of the columns of  $\mathbf{B}$

$$\mathbf{A}\mathbf{B} = (\mathbf{A}\mathbf{B}_{*1}, \mathbf{A}\mathbf{B}_{*2}, \dots, \mathbf{A}\mathbf{B}_{*m})$$

Matrix multiplication is not commutative in general. A matrix is square if  $n = m$ . A square matrix is invertible or nonsingular if there exists a matrix  $\mathbf{A}^{-1}$  such that  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$ . A matrix is symmetric if  $\mathbf{A}^T = \mathbf{A}$ , a matrix is unitary if  $\mathbf{A}^{-1} = \mathbf{A}$ ; a matrix is idempotent if  $\mathbf{A}^2 = \mathbf{A}$ . The rank of a matrix is the minimum number of columns that are linearly independent.

For two vectors  $\mathbf{u}$  and  $\mathbf{v}$ , the product  $\mathbf{u}^T \mathbf{v}$  is called the *inner product*. The two vectors are called orthogonal if  $\mathbf{u}^T \mathbf{v} = 0$ . A set of orthogonal vectors  $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  is called a basis of  $\mathbb{R}^n$  if its span is the whole  $\mathbb{R}^n$ . The function  $\|u\| = \sqrt{\mathbf{u}^T \mathbf{u}}$  is called the *norm* of  $\mathbf{u}$ . A vector is a unit vector if  $\|u\| = 1$ . For a unit vector  $\mathbf{u}$  and an arbitrary vector  $\mathbf{v}$ , the vector  $\mathbf{p} = (\mathbf{v}^T \mathbf{u}) \mathbf{u}$  can be seen also as the projection or shadow of  $\mathbf{u}$  onto  $\mathbf{v}$ . The matrix  $\mathbf{P} = \mathbf{u}\mathbf{u}^T$  is called a projector.

---

<sup>1</sup>abbreviated as “iff” from now on

## Eigendecomposition

Given a square matrix  $\mathbf{A}$ , the pairs  $(\lambda, \mathbf{v})$  for which  $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$  are called the eigenvalues and eigenvectors of the matrix  $\mathbf{A}$ . The eigenvalues are real if the matrix is symmetric. If the matrix is invertible  $\mathbf{A}^{-1}\mathbf{v} = \lambda^{-1}\mathbf{v}$  for the same eigenpair  $(\lambda, \mathbf{v})$ . The set of eigenvalues of a given matrix  $\mathbf{A}$  is called *the spectrum* and is denoted by  $\sigma(\mathbf{A})$ . Two matrices  $\mathbf{A}$  and  $\mathbf{D}$  are similar if there exist a matrix  $\mathbf{P}$  such that  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ , in particular, if the matrix is similar to a diagonal matrix it is called *Diagonalizable*.

Finally, we close the section by stating the *spectral theorem* which guaranties the existence of the principal components:

**Theorem 1** (Spectral theorem). *Every symmetric matrix  $\mathbf{A}$  with spectrum  $\sigma(\mathbf{A}) = (\lambda_1, \lambda_2, \dots, \lambda_n)$  and associated eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , can be decomposed as*

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{G}_i, \quad (3.1)$$

where  $\mathbf{G}_i = \mathbf{v}_i \mathbf{v}_i^T$  are the spectral projectors.

### 3.1.1 Vector spaces

The concepts introduced before apply to a wider range of objects that behave like vectors (for example, functions), in that general case, matrices are replaced by *linear transformations*. This generalization is necessary because the objects we are dealing with when analyzing firing rates of neurons are functions of time; this has important consequences that constraint the geometry and the spectrum, ultimately affecting the shape as we will see in the next section.

A vector space  $V$  is a 4-tuple  $(X, F, +, \cdot)$ .  $X$  is a set whose elements are called vectors,  $F$  is a field or a set of numbers called scalars, and,  $+$  and  $\cdot$  are operations called vector addition and multiplication by a scalar that satisfy a set of axioms that can be found in any linear algebra textbook [59]. During this work we assume  $F = \mathbb{R}$  unless otherwise stated. The spaces  $\mathbb{R}^n$  and  $\mathbb{R}^{n \times m}$  introduced before are vector spaces.

A subspace is a subset  $S \subset X$  that is a vector space in itself. The dimension of a vector space is the minimum number of basis vectors that span it. A vector space is finite dimensional if  $\dim(V) = n < \infty$ , otherwise it is infinite dimensional. It is called a inner product space if there is an inner product  $\langle u, v \rangle$  defined on it, the more important inner product spaces are *Hilbert spaces*: spaces that are complete all the limits of sequences of elements inside the space converge to an element of the space. An inner product induces a norm as we saw in the previous section. The distance between two objects in a normed space is given by  $d(u, v) = \|u - v\|$ .

Apart from the euclidean space  $\mathbb{R}^n$ , other spaces used in this text are

- $C^k(\mathbb{R})$ , The space of continuous functions on  $\mathbb{R}$  with  $k$  continuous derivatives. When  $k = \infty$ , its elements are called *smooth functions*
- $L^p(\mathbb{R})$ , The space of integrable functions with norm

$$\|f\| = \left( \int_{\mathbb{R}} f^p(s) ds \right)^{1/p}.$$

Given two elements  $f$  and  $g$  in a linear vector space  $X$ , the line between them, given in parametric form, is

$$L(t) = tf + (1 - t)g, \tag{3.2}$$

for  $t \in [0, 1]$ . A subset  $S$  is called convex if for every  $f, g \in S$ , the line  $L(t) \in S$  for all  $t$ . It is called *open* if it does not contain its boundary. For example, the set  $\{(x, y) \in \mathbb{R}^2 \text{ s.t. } x > 0 \text{ and } y > 0\}$  is open because it does not contain all the points in the coordinate axes. A more rigorous definition can be found in any book of topology. If a set is not open, then it is closed.

### 3.1.2 Groups, maps and actions

Now we turn to some basic concepts about groups and group actions. Groups are used in chapter 5 to encode the sort of transformations a shape is invariant to. For example, a rotated triangle is still a triangle, so we say that the *group action of the group of rotations* on the triangle leaves the shape invariant[86].

The Cartesian product of two sets  $A$  and  $B$  is denoted by  $A \times B$ . A map or function is relation, (i.e. a subset of  $A \times B$ ) with no repeated elements from  $A$ . A map is injective or into if no two elements from  $A$  are mapped to the same element in  $B$ . It is onto or surjective if all the elements of  $B$  have a *preimage* on  $A$ . A map is a bijection or is one to one if it is into and onto. We use bijections to prove the equivalence of two seemingly unrelated mathematical objects like the eigenvector systems and the Catalan trees (see chapter 6).

An equivalence relation  $\sim$  is a relation that is reflexive ( $a \sim a$ ), symmetric  $a \sim b = b \sim a$  and transitive  $a \sim b$  and  $b \sim c$  implies  $a \sim c$ . An equivalence relation partitions a set into classes. If the element  $a \in A$  is a representative of a class, the set of all elements equivalent to it is denoted by  $[a]$ . For example, the shape of an object is the equivalence class of all the objects with similar shapes.

A group is a set  $G$  along with an operation  $\cdot$  that satisfies the following axioms

- Closure: For all  $x, y \in G$ ,  $x \cdot y \in G$

- Identity: There is an identity element  $e$  such that  $e \cdot x = x$  for all  $x \in X$
- Associativity: For all  $x, y, z \in G$ ,  $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
- Inverse: For each  $x \in G$  there is an element  $x^{-1} \in G$  such that  $x \cdot x^{-1} = e$

When the group satisfies the commutative axiom ( $x \cdot y = y \cdot x$ ) it is called *Abelian*. Given a set  $X$ , the *action* of  $G$  on  $X$  is given by a map  $g \cdot x$ ,  $g \in G$  and  $x \in X$ , that satisfies i.  $e \cdot x = x$ , where  $e \in G$  is the identity element of the group, and ii.  $g \cdot (h \cdot x) = (g \cdot h) \cdot x$  for  $g, h \in G$ . The set  $\mathcal{O}(x) = \{g \cdot x \text{ for } g \in G\}$  is called the *orbit* of  $x$ .

Examples of groups are the group of invertible matrices  $GL$ , the set of rotation matrices in  $n$  dimensions  $SO(n)$  and the symmetric group  $S_n$  or the group of permutation of  $n$  elements.

## 3.2 Singular Value Decomposition and Principal Components Analysis

PCA will be used to mean Principal Component Analysis, taking analysis in its dialectic sense, that is, as a decomposition of our data matrix in individual constituents and not just the method of reducing dimension itself which is only a consequence of it.

We will work with a  $n \times m$  matrix  $\mathbf{X}$  called the *data matrix*. We assume, as this is our case, that each row of the matrix corresponds to an unit or neuron and each column to a sampling point of the estimated firing rate for that neuron. The symmetric matrix  $\mathbf{C} = \frac{1}{m} \mathbf{X} \mathbf{X}^T$  will be called the (sample) covariance matrix or simply covariance matrix. The dual matrix  $\mathbf{G} = \frac{1}{n} \mathbf{X}^T \mathbf{X}$  is called the time-covariance matrix.

**Definition 1** (Principal Component Analysis). *The principal component analysis of a data matrix  $\mathbf{X}$  is the triplet  $(\mathbf{V}, \mathbf{D}, \mathbf{Z})$ , where  $\mathbf{V}$  is the matrix whose columns are the eigenvectors of the (sample) covariance matrix,  $\mathbf{D}$  is a diagonal matrix with the corresponding eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$ , sorted such in a descending way, and  $\mathbf{Z}$  is the matrix with the projections of the the population on the basis given by  $\mathbf{V}$  as rows, that is:*

$$\mathbf{Z} = \mathbf{V}^T \mathbf{X}. \quad (3.3)$$

$\mathbf{Z}$  is called the matrix of principal components (PC).

There is a dual decomposition based upon the time covariance matrix. Let  $(\mathbf{U}, \mathbf{D}', \mathbf{W})$  be the eigenvectors, eigenvalues and projections with respect to  $\mathbf{G}$  ( $\mathbf{W} = \mathbf{X} \mathbf{U}$ ), then the following relations are held:

$$\mathbf{D} = \mathbf{D}' \quad (3.4)$$

$$\mathbf{U} = \mathbf{Z}^T \mathbf{D}^{-1/2} \quad (3.5)$$

$$\mathbf{W} = \mathbf{D}^{-1/2} \mathbf{V}^T \quad (3.6)$$

The matrix  $\mathbf{V}$  and  $\mathbf{U}$  are unitary, they, along with the eigenvalues also form the *Singular Value Decomposition* of  $\mathbf{X}$ :

**Definition 2** (Singular Value Decomposition). *The SVD of  $\mathbf{X}$  is a triplet  $(\mathbf{V}, \mathbf{\Lambda}, \mathbf{U})$  with  $\mathbf{U}, \mathbf{V}$  unitary, such that*

$$\mathbf{X} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^T \quad (3.7)$$

The matrix  $\mathbf{U}$  and  $\mathbf{V}$  are the same as before and  $\mathbf{\Lambda} = \sqrt{\mathbf{D}}$ . This means that each data element can be written as a linear combination of the PCs:

$$x_i(t_j) = \sum_k v_k z_k(t_j) \quad (3.8)$$

In practice, the data matrix is often centered before the decomposition. The matrix  $\mathbf{C} = \mathbf{I}_m - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T$  is called the centering matrix and  $\tilde{\mathbf{X}} = \mathbf{X} \mathbf{C}$  is the centered version of  $\mathbf{X}$ . The mean subtracted or centered PCA is then applied to this new matrix.

It is often convenient to replace the covariance matrix  $\mathbf{C}$  ( $\mathbf{G}$ ) with a normalized version called the correlation matrix  $\tilde{\mathbf{C}}$  defined as:

$$\tilde{\mathbf{C}} = \mathbf{S}^{-1/2} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \mathbf{S}^{-1/2}, \quad (3.9)$$

where  $\mathbf{S}$  is the matrix with the standard deviations  $\sigma_i$  of each neuron in the diagonal. In general, the first PC passes through the centroid of the data if the mean is not subtracted, however, the same is not true for block matrices (matrices with clusters).

### 3.2.1 Characterization and interpretations of PCA

We now focus on useful facts and properties of PCA.

#### Best low rank approximation of the data

An important property of PCA, and the one that mostly justifies this work, is the following. Remember that the space of matrices is a vector space. There we can define a norm and a distance.

**Definition 3** (Frobenius norm). *The map  $\|\cdot\|_2 : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}$ , given by*

$$\|\mathbf{X}\|_2 = \left( \sum_i \sum_j \mathbf{x}_{ij}^2 \right)^{1/2}$$

*is called the Frobenius norm.*

The induced distance  $d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_2^2$  is just the sum of square errors of all the elements.

Given a data matrix  $\mathbf{X}$ , the PCA provides the rank  $k$  best approximation to the data in terms of the Frobenius distance, that is, let  $\mathbf{X}_k$  be the reconstructed matrix obtained by truncating the sum in (3.8) at the  $k$ -th term (or alternatively, by selecting the first  $k$  columns of  $\mathbf{V}$ ,  $\mathbf{D}$ , and  $\mathbf{U}$  in the SVD), then

$$f(\mathbf{Y}) = d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_2^2 \quad (3.10)$$

where  $\mathbf{Y}$  is a rank  $k$ ,  $n \times m$  matrix, reaches a minimum precisely at  $\mathbf{X}_k$ . That is, the data reconstructed from the first  $k$  PCs is as close to the data as it can be.

The minimization of (3.10) can be restated as a problem of training a neural network with 1 hidden layer with  $k$  units and inputs and outputs layers with  $n$  units so that the output tracks as closely as possible the input. This configuration is called an *Autoencoder* and the space spanned by the weights is the same as the PC subspace obtained by taking the first  $k$  eigenvectors [12].

### Geometrical objects associated to PCA

PCA also gives the orthogonal directions of maximal variance in the data. Indeed, the variance of each of the principal components is given by  $\text{var}(z_k) = (v_k^T \mathbf{X})(v_k^T \mathbf{X})^T = v_k^T \mathbf{C} v_k = \lambda_k$ . It can be shown, for example, by the method of Lagrange multipliers, that of all possible orthonormal sets of directions (vectors  $u_i$  such that  $\|u\| = 1$  and  $u_i^T u_j = 0$  for  $i \neq j$ ), the eigenvectors  $v_k$  are those which maximize the *quadratic form*  $q(x) = x^T \mathbf{C} x$ .

A quadratic form is a polynomial of degree two given, as seen before, by the equation  $q(x) = x^T \mathbf{C} x$ . When  $\mathbf{C}$  is the covariance matrix, the level curve  $q(z) = z^T \mathbf{C}^{-1} z = 1$  is the best fitting ellipsoid to the rotated data ( $z_k$ ). Its major axis (first eigenvector) shows the direction of greater variance or the line of best fit, while its minor axis (last eigenvector) shows the direction of least variance or worst fitting line.

Another geometrical object associated to PCA is the quadratic form in itself. For covariances matrices it is a paraboloid, however, for correlation matrices, the associated *quadratic* is an hyperbolic paraboloid (figure 3.1) as shown by the following theorem

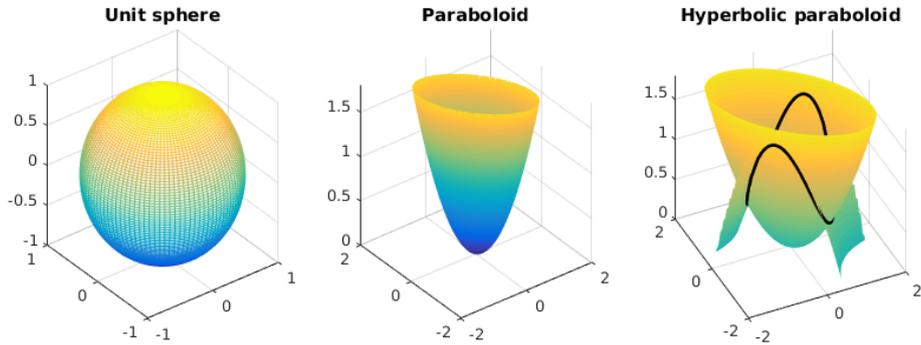


Figure 3.1: Geometrical objects associated to the PCA. Left. Sphere of orthonormal directions, the eigenvectors are those directions that maximize variance subject to the orthogonality constraint. Middle. The paraboloid is the graph of the quadratic form itself (see text). Right. When the correlation matrix is used instead, the space is a hyperbolic paraboloid.

**Theorem 2** (Correlation hyperboloid). *The associated quadric, for normalized (zscored) data is an hyperbolic paraboloid.*

*Proof.* Developing the matrix equation of the quadratic form we have

$$q(x) = \sum_i \sum_j c_{ij} x_i x_j,$$

this sum can be split however into 2 terms:

$$q(x) = \sum_i c_{ii} x_i^2 + \sum_{i \neq j} c_{ij} x_i x_j.$$

The first term is constant because all the data elements are normalized to have unit variance and so

$$q(x) = a + \sum_{i \neq j} c_{ij} x_i x_j.$$

which is the equation for an hyperbolic paraboloid. The relationship of this hyperboloid and the corresponding paraboloid is shown in figure (figure 3.1).  $\square$

Therefore, for each data matrix  $\mathbf{X}$  there is an associated ellipse, an sphere of unit directions, a paraboloid and, for correlation matrices, an hyperbolic paraboloid. These are all natural manifold in which the PC “live” and to which they are constrained. We use this fact as a guiding principle for developing our theory of null shapes in chapter 5. The observation that the natural spaces and the previous theorem, despite their simplicity have not been found in the literature before.

Finally, associated to PCA there is also a *hyperplane arrangement*. This last geometrical object will be treated later in chapter 6.

## Optimal compression

How many basis vectors are necessary to represent a given data point? The answer to this question is related to the optimal approximation issue dealt with before. This time however, we adopt a coding approach. Imagine the basis vectors as symbols and each data points as words. We associate here and in chapter 6 the problem of finding the PCA to the problem of finding a code with minimum *average length* (i.e. a Huffman code) [23]

Consider a set of vectors  $X = \{x_1, x_2, \dots, x_n\}$ , their *length entropy* is defined as

$$H(X) = \sum_i \|x_i\| \log \|x_i\|. \quad (3.11)$$

It was shown in [88] that it is  $\mathbf{V}$ , the matrix of eigenvectors of the covariance matrix of  $X$ , the element of  $SO(n)$ , the group of rotation matrices, that minimizes that functional. The minimal value is reached then at

$$H(\mathbf{C}) = \sum_i \lambda_i \log \lambda_i, \quad (3.12)$$

where  $\lambda_i$  are the eigenvalues of the covariance matrix. This turns out to be the *Von Neumann* entropy of the data

$$H(\mathbf{C}) = \text{Tr}(\mathbf{C} \log \mathbf{C}), \quad (3.13)$$

So PCA gives a code with minimum average length for the data points when they are taken as codewords and the basis vectors as symbols.

## PCA as a sequence of rank one updates

The covariance matrix can be written as a sequence of rank one updates, this point of view allows to understand how the eigenvectors and eigenvalue change as new samples of the population vector are added. Rank one update to a matrix  $\mathbf{A}$  is the matrix that results from adding a rank one matrix  $\mathbf{B} = uu^T$  to  $\mathbf{A}$  where  $u$  is a column vector; that is  $\mathbf{A}' = \mathbf{A} + uu^T$ . Therefore, for a covariance matrix  $\mathbf{C}$  we have

$$\mathbf{C} = \frac{1}{m} \sum_j \mathbf{X}_j \mathbf{X}_j^T \quad (3.14)$$

where  $\mathbf{X}_j$  is the  $j$ th column of the data matrix, that is, a snapshot of the population activity (population vector) at time  $j$ . Because the data that occupy us in this work is composed of smoothed time series, we can assume that the data

is continuous and also differentiable. So, for  $m$  large enough, we can define the function:

$$\mathbf{C}(t) = \int_0^t x(s)x(s)^T ds \quad (3.15)$$

where  $x \in \mathbb{R}^n$  is the population trajectory. Its derivative is the rank one matrix:

$$\dot{\mathbf{C}}(t) = x(t)x(t)^T. \quad (3.16)$$

Now, consider the modified eigenvalue-eigenvector problem:

$$C(t)v(t) = \lambda(t)v(t), \quad (3.17)$$

Differentiating we can find the following formal equations for the instantaneous change in the eigenvalues and eigenvectors as new samples arrive:

$$\dot{\lambda}_i(t) = (v_i^T x(t))^2 \quad (3.18)$$

$$\dot{v}_i(t) = (v_i^T x(t)) \sum_{j \neq i} \frac{v_j^T x(t)}{\lambda_i - \lambda_j} v_j \quad (3.19)$$

where  $v_i \in (S)_n$  (is constrained to the  $n$  dimensional sphere). We see therefore that the amount of change in an eigenvalues when a new sample is added is proportional to *how similar* the pattern in the population vector is to the given eigenvector, while the change in direction of the eigenvectors is higher the closer the eigenvalues. This evidence a repulsion phenomenon in which 2 eigenvalues can not be very close unless the data is orthogonal along those directions.

### 3.2.2 PCA of random matrices

It might be tempting to think, for example when developing the heterogeneity measures of chapter 6, that the spectrum of a randomly distributed matrix has a similar distribution as the variables. For example, how does the spectrum of a data matrix where the entries are samples of a Gaussian distribution looks like? It turns out that the spectrum has a specific shape that is independent of the nature of the data.

When the entries of our data matrix are drawn from a probability distribution, the matrix is a *random matrix*. Let's assume each data entries  $x_{ik}$  are iid real random variable,  $x_{ik} \sim \mathcal{N}(0, \sigma^2)$ , then, the sample covariance matrix is, as before (we use the notation of [6]),

$$\mathbf{S}_n = \frac{1}{n} \mathbf{X} \mathbf{X}^T, \quad (3.20)$$

where  $\mathbf{X}$  is a  $p \times n$  matrix. We assume, as is usual in the theory of high dimensional statistics, that  $p$  is proportional to  $n$ , or in other words  $p/n \rightarrow y < \infty$  as  $n$  tends to infinity.

In the context of high dimensional random matrices, the mean is not subtracted due to the fact that this is a rank one modification with no relevant impact on the basic properties. The Empirical Spectral Distribution of  $\mathbf{S}_n$  is defined as

$$F(x) = \frac{1}{m} \#\{j \leq m : \lambda_j \leq x\}, \quad (3.21)$$

Then we have the following

**Theorem 3** (Marchenko-Pastur). *The ESD of the sample covariance matrix (3.20) has a density function (i.e. its derivative)*

$$p(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2} \sqrt{(b-x)(x-a)} & \text{if } x \in [a, b] \\ 0 & \text{otherwise,} \end{cases} \quad (3.22)$$

where  $a = \sigma^2(1 - \sqrt{y})^2$  and  $b = \sigma^2(1 + \sqrt{y})^2$ . That is, all the eigenvalues are restricted in a bounded domain that depends upon the dimensions (there is a point mass at the origin too for  $y > 1$ ).

### 3.2.3 Functional PCA

The technically correct form of PCA for our data is called *functional PCA*, after all, the firing rates are functions of time which implies that the columns of the data matrix are statistically dependent; we use this fact for analytical purposes as many problems become tractable once put in terms of integral operators which are the natural objects in this context. A linear integral operator in a space of functions  $X$  (like  $L^p$  seen before) is a map  $A$  on that space given by:

$$A[u](x) = \int_{\mathbb{R}}^n W(x, y)u(y)dy, \quad (3.23)$$

with  $u \in X$ , where the function  $W(x, y)$  is called the *kernel*. If each row of our data lives in a function space, we define the covariance kernel as

$$C(t, s) = \sum_i x_i(t)x_i(s), \quad (3.24)$$

where  $x_i$  are our data functions. The functional PCA eigenvalue problem is set up then as

$$\int_{\mathbb{R}} C(t, s)\phi_k(s)ds = \mu_k\phi_k(t). \quad (3.25)$$

This will be expanded in our analysis of the null shapes in chapter 5.

### 3.3 Differential geometry

Now we focus on the study of curves as they are our main object of interest in this thesis. A curve is a continuous map  $\gamma : [a, b] \rightarrow \mathbb{R}^n$  given by  $\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_n(t))$ ,  $t \in [a, b]$  is called the parameter. If it is infinitely differentiable we call it smooth. A curve is regular if  $\|\dot{\gamma}(t)\| > 0$  for all  $t \in [a, b]$ . A curve is closed if  $\gamma(a) = \gamma(b)$ , otherwise is open. If the curve does not cross itself it is called simple. We assume without loss of generality that  $[a, b] = [0, 1]$ .

#### 3.3.1 Parametrization

The length of a curve  $\gamma$  or arc length between  $[0, t]$ ,  $t \in [0, 1]$ , is given by

$$s(t) = \int_0^t \|\dot{\gamma}(\tau)\| d\tau; \quad (3.26)$$

The total length of the curve is then  $L = s(1)$ . A continuous monotonically increasing (bijective) function  $\phi : [0, 1] \rightarrow [a, b]$  is called a homeomorphism. The curve  $\gamma(t)$  is the same as  $\tilde{\gamma}(s) = \gamma(\phi(t))$  seen as a set of points, i.e. it traverses the same path. We say that  $\tilde{\gamma}$  is a reparametrization of  $\gamma$ . Reparametrization gives rise to the equivalence relation  $\alpha \sim \beta$  iff there exists  $\phi$ , homeomorphism, such that  $\alpha(t) = \beta(\phi(t))$  for all  $t$ .

The natural reparametrization for a curve is given by the arc length and it is called the *arc length parametrization*. Indeed, the map  $s : [0, 1] \rightarrow [0, L]$  is a homeomorphism and therefore has an inverse that can be used to find a valid reparametrization [37]. We assume, from now on, that the curves are arc length parametrized unless otherwise stated.

#### 3.3.2 Frenet frame

The tangent to a curve  $\gamma(s)$  is given by its first derivative  $\tau(s) = \gamma'(s)$  (we use  $' = \frac{d}{ds}$ ,  $s$  the arc length parameter); it is by definition a unit vector. The second derivative is orthogonal to the tangent vector and therefore defines the *normal*  $n(s)$  to the curve at each point. The ordered pair  $(\tau(t), n(t))$  span a moving coordinate system and are called together *the Frenet frame*.

The vector differential equation with variable coefficients called Frenet equation is

$$\begin{pmatrix} \dot{\tau}(s) \\ \dot{n}(s) \end{pmatrix} = \begin{pmatrix} 0 & \kappa(s) \\ -\kappa(s) & 0 \end{pmatrix} \begin{pmatrix} \tau(s) \\ n(s) \end{pmatrix}; \quad (3.27)$$

it completely defines the evolution of the curve in terms of the moving frame. The function  $\kappa(s)$  is called the curvature and its inverse  $1/\kappa(s)$  is called the osculating circle; the best fitting circle to the curve at that point.

If  $\theta(s)$  is the angle of  $\tau(s)$  with a basis vector, then the curvature is defined as the rate of change of that angle  $\kappa(s) = \theta'(s)$ . In the general case

$$\kappa(t) = \frac{\det(\dot{\gamma}(t), \ddot{\gamma}(t))}{\|\dot{\gamma}(t)\|^3} \quad (3.28)$$

This equation also works for higher dimensions. For higher dimensions, say,  $n$ , we have  $n - 1$  curvatures. The Frenet framework also works in that case and the curvatures can be computed iteratively [38]. The osculating plane is, in some sense, the best low dimensional approximation to the curve *locally*, for that reason it does not come as a surprise that it is related to PCA as, indeed, is proved in [3] where it is shown that, for points in a small neighborhood of a point of a curve, the PCA space is the same as the osculating plane.

### 3.3.3 Contact of curves: Vertices and flat regions

Singularity theory [14] gives us a framework to rigorously define the points of interest in a curve and develop tests for their identification. We now expose the basic concepts used the chapter 5 to define vertices and flat regions.

Given a functional  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  and a curve  $\gamma(t)$ , we study the function  $g(t) = F(\gamma(t))$ .  $\gamma$  and  $F^{-1}$  have  $k$ -fold contact at  $g(t_0)$  if  $g^{(i)}(t_0) = 0$  for  $i = 0, 1, \dots, k-1$  but  $g^{(k)} \neq 0$ ;  $g^{(i)} = \frac{d^i g}{dt^i}$  and  $g^0(t) = g(t)$ . We focus in two important functions:

- $F(x) = \|x - u\| - \|u - p\|$ , the circle with center  $u$  and radius  $p$ . The contact with this function tell us how circular the curve is at  $t_0$
- $F(x) = (x - p)^T u$ , the line passing trough  $p$ , orthogonal to  $u$ . It tell us how flat the curve is at  $t_0$ .

A *vertex* is a point of a curve for which there is a circle having 4 (ordinary vertex) or at least 5 (higher vertex) contact with the curve at that point. Vertices are points in which the curve is rounder than the surrounding points. A flat point is one in which the curve has 3 fold contact (ordinary) or 4 fold contact (higher) with the tangent line at that point.

The circle function gives rise to the *square distance function*  $f(s) = \|\gamma(s) - u\|^2$ . By studying the derivatives of this functions, we can conclude that the unique circle with 3 point contact is the osculating circle. Also, that for the osculating circle to have higher contact, we need  $\kappa(s) \neq 0$ ,  $\kappa'(s) = 0$  (4 point contact) and  $\kappa'(s) = \kappa''(s) = 0$  (higher contact).

By studying the *height function*  $h(s) = \gamma^T(s)u$  we can conclude that sufficient and necessary conditions for a flat point (inflections) are  $\kappa(s) = 0$  (ordinary) and  $\kappa(s) = \kappa'(s) = 0$  (higher)

### 3.4 Semiorders

A semiorder is a kind of order relationship that arises often in science [98]. Consider the problem of comparing 5 means from 5 different samples. We want to know which ones differ significantly and which ones do not. Let us assume, that our statistical tests give the result shown in figure 3.2. All the means covered by a line are statistically equal, this is what is called the symmetric part of the semiorder; it is an *equivalence relation*. No line is completely covered by another one because this relation is transitive.



Figure 3.2: Left. The problem of comparing means. Right. The corresponding decision pattern

Suppose, in addition that you sort the means from left to right in a descending way. That is the asymmetric part of the semiorder. It is a *partial order* in that it is transitive and reflexive but antisymmetric. Therefore any ordering having a symmetric and an antisymmetric part is a semiorder. In our case, given that during a putative response to an event, neurons can interact at some time segments but not at others, the semiorders is the most natural structure to keep track of whether they do interact or, on the contrary, fire with different patterns.

In the case of three neurons, for example, responding to an event, there are 5 ways in which they can respond. They can either completely overlap, be disjoint or partially overlap having some of them belonging to two ensembles at the same time. One possible scenario in which this can happen is in the *mixed selectivity* framework [78] (neurons being selective to more than one feature of the stimulus; neurons sharing the feature they are selective to)

### 3.5 Heat equation and diffusion

In order to analyze the PCA structure of the covariance matrices we will make use of then spectral theory of the Laplace operator. Therefore, we will assume

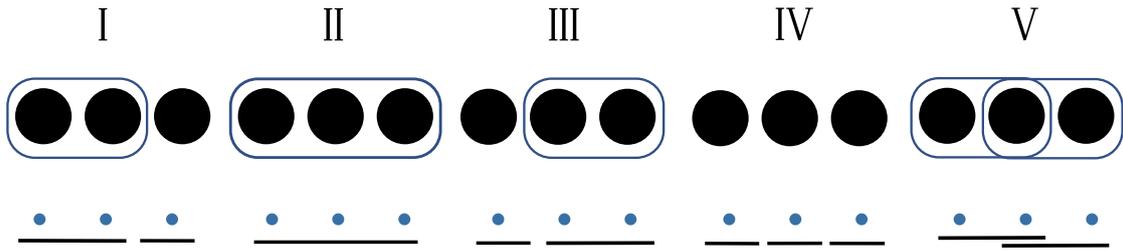


Figure 3.3: All the possible interactions for 3 neurons

a continuous time domain and, therefore, technically the results concern what is called *functional PCA*. The discretization, however, does not affect our conclusions.

### 3.5.1 The laplacian

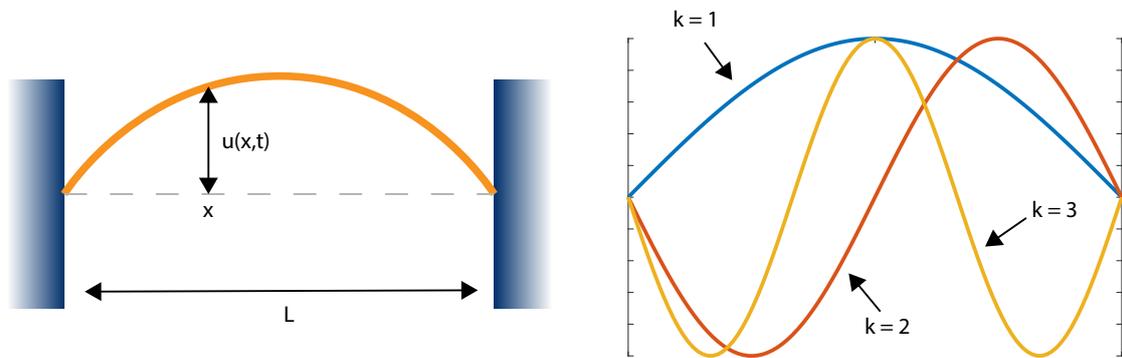


Figure 3.4: Left. Physical situation - a string fixed on its extremes. Right. Eigenfunctions of the Laplacian

It is convenient to introduce the laplacian operator in the context of the familiar physical situation of a string fixed at its extremes. Consider figure 3.4, the vibrations of the the string fixed at  $x = -L/2$  and  $x = L/2$  are described by the wave equation. Let  $u(x, t)$  be the height of the string at position  $x$  and time  $t$ , then, the vibration is described by the equation

$$\partial_{xx}u = \frac{\rho}{T}\partial_{tt}u \quad (3.29)$$

where we adopt the convention  $\partial_{tt} = \frac{\partial^2}{\partial t^2}$  for the differential operator.  $\rho$  is the density of the string and  $T$  is the tension. For a derivation of this equation see [32]. The situation of the string fixed at both extremes is represented by the initial

conditions

$$u(x, t) = 0, \quad \text{for } x = \pm L/2 \quad \text{for all } t \quad (3.30)$$

In order to solve this equation we use the separation of variables technique, that is, we look for solutions of the form

$$u(x, t) = \phi(x)\psi(t). \quad (3.31)$$

Replacing into the equation (3.29) we get

$$\psi(t) \frac{d^2}{dx^2} \phi(x) = \frac{\rho}{T} \phi(x) \frac{d^2}{dt^2} \psi(t), \quad (3.32)$$

which after reorganizing the terms becomes

$$\frac{1}{\phi(x)} \frac{d^2}{dx^2} \phi(x) = \frac{\rho}{T} \frac{1}{\psi(t)} \frac{d^2}{dt^2} \psi(t) = \mu. \quad (3.33)$$

As a consequence of that, we can solve for the time and spatial components of the equation separately. In this paper we are only interested in the spatial part which gives what are called standing waves which, incidentally, will be our principal components in the forthcoming sections. The spatial part from equation (3.33),

$$\frac{d^2}{dx^2} \phi(x) = \mu \phi(x) \quad (3.34)$$

is the eigenvalue problem for the *Laplacian* operator  $\Delta = \frac{d^2}{dx^2}$  which is pervasive in many areas of the basic and applied sciences. The solutions have the form

$$\phi(x) = A \sin(\sqrt{\mu}x) + B \cos(\sqrt{\mu}x), \quad (3.35)$$

subject to the initial conditions  $\phi(\pm L/2) = 0$ . Replacing the initial conditions in the solution (3.35) we find that one of the following equations must be satisfied (but not both). Either

$$A \sin\left(\frac{\sqrt{\mu}L}{2}\right) = 0, \quad B = 0$$

or

$$B \cos\left(\frac{\sqrt{\mu}L}{2}\right) = 0, \quad A = 0.$$

This implies  $\sqrt{\mu}L/2 = k\pi/2$ ,  $k = 1, 2, 3, \dots$  or

$$\mu = \frac{k^2 \pi^2}{L^2}. \quad (3.36)$$

The eigenfunctions or modes of vibration are, in turn

$$\phi_k(x) = \begin{cases} B \cos\left(\frac{k\pi}{2}\right), & k = 1, 3, 5, \dots \\ A \sin\left(\frac{k\pi}{2}\right), & k = 2, 4, 6, \dots \end{cases} \quad (3.37)$$

Note that the spectrum is discrete even though the problem is continuum, this is a basic fact of harmonic analysis due to the fact that the domain is bounded. The eigenfunctions are shown in figure 3.4.

### 3.5.2 Relationship with the heat kernel

Another physical situation that comes into this analysis and that is closely related with the previous one is that of the finding the distribution of temperatures in a given material over time. The dynamics of such phenomena is described by the heat equation

$$\partial_t u - \kappa \partial_{xx} u = 0 \quad (3.38)$$

where  $\partial_t u = \frac{\partial u}{\partial t}$  and  $\kappa$  is the thermal diffusivity. We will use this equations in more than one context in this paper. Now we will focus on its solutions. Again, we use the separation of variables technique to find

$$\frac{d}{dt} \psi(t) = \frac{\kappa}{\phi(x)} \frac{d^2}{dx^2} \phi(x) = \mu, \quad (3.39)$$

which again, contains the laplacian eigenvalue problem in its spatial part.

## 3.6 Further reading

Detailed treatment of the matrix theory and advanced linear algebra can be found in [59]. There are many nice accounts on the properties of PCA, a good starting point is a recent review by Jolliffe and Cardima [44] and the classical book on the topic [43] although the authors find the development in [17] more than satisfactory. The details about hyperplane arrangements can be found in [2] and [11].

## Chapter 4

### Description of the physiological data

## 4.1 Experimental setup and methods

We used data from previous studies of decision making[33] and action suppression[83].

### 4.1.1 Data collection

The data collection process has been described elsewhere in detail [83]. Subjects (adult male Long-Evans rats) were trained in the task described in the following section. After achieving stable performance ( $> 70\%$  correct choices on Go trials), the data collection was performed. Wide band signals were recorded from 21 implanted tetrodes that targeted basal ganglia structures: Striatum (Str), globus pallidus (GP), subthalamic nucleus (STN) and Substantia Nigra pars reticulata (SNr), and the parafascicular nucleus in the thalamus.

### 4.1.2 Task structure

During the task, hungry rats were in an operant chamber with 5 nose poke holes on the wall and a food dispenser opposite to it (figure 4.1). In *Go trials*, the rat should poke its nose in one of the central holes(Nose In) illuminated at random (Light On). After a variable delay, 500ms to 1200ms, the Go cue (GO, 65dB, 50ms) instructed it to move swiftly (Nose Out - Side In) either to the left (1kHz tone) or to the right (5kHz tone). A limited holding period encouraged the rats to respond fast. Correct trials, those where the movement is done within the holding time and in the right direction, were immediately rewarded.

For some analyses we used the stop trials in which a second signal after the go cue instructs the rat to cancel the movement. Correct and failed stop trials are determined according to whether the stopping was successful. Those trials were compared with fast and slow go trials determined using the reaction time distribution (see [94]).

### 4.1.3 General methods

All the analyses were performed in Matlab R2018a. After manually spike sorting and cell type classification, the spike trains were pooled across trials, sessions and subjects, and aligned to the event of interest (correctgo) for the contralateral side in most of the analyses performed in this work. The PSTH (PETHs, technically, Peri-Event, Time Histograms) of the aligned spike trains was computed, averaging over trials and computing the spike counts with bin size of 50ms with step size of 10ms and smoothed with a square moving window of 20ms. We performed the analyses under different pre-processing conditions as part of the analysis in the following sections.

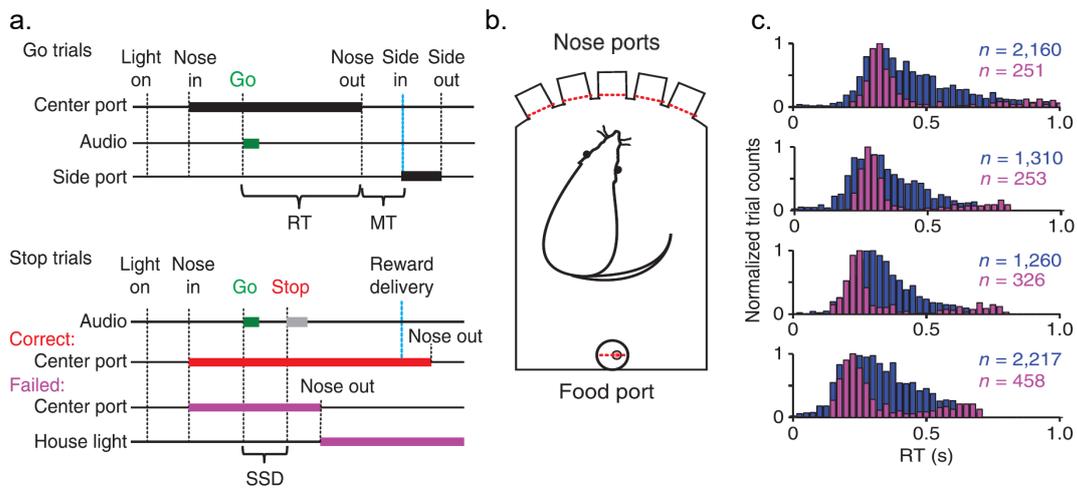


Figure 4.1: a. Task structure (for details, see [83]). Go trials. at the start of the trial the rat pokes its nose into a give port (Nose In). After the go cue is given, the rat moves (Nose out) its nose into the left or right ports; a reward is given in correct trials. Stop trials. In some trials, after the go cue a second signal indicates the cancellation of movement. b. Chamber configuration with the nose and food (reward) ports. c. Fast and slow trials are determined by choosing the fast and slowest parts of the reaction time distribution. RT: reaction time. MT: movement time.

Notably, the results do not change with other firing rate estimation methods like SDF or BARS [46]; or the parameters like window or step size. The estimated firing rates were finally cut in a symmetric 1s interval around the event and organized in a data matrix with neurons as rows and time as columns. Other methods are developed as part of the forthcoming sections.

## 4.2 Overview of the PCA information for the basal ganglia around movement initiation

For the computation of PCA we arranged the mean firing rates into a matrix as described in the previous section; each row corresponding to a neuron and the covariance, eigenvectors and projections were computed as in section 3.2. Figure 4.2 shows the summary PCA information for the different regions in the basal ganglia.

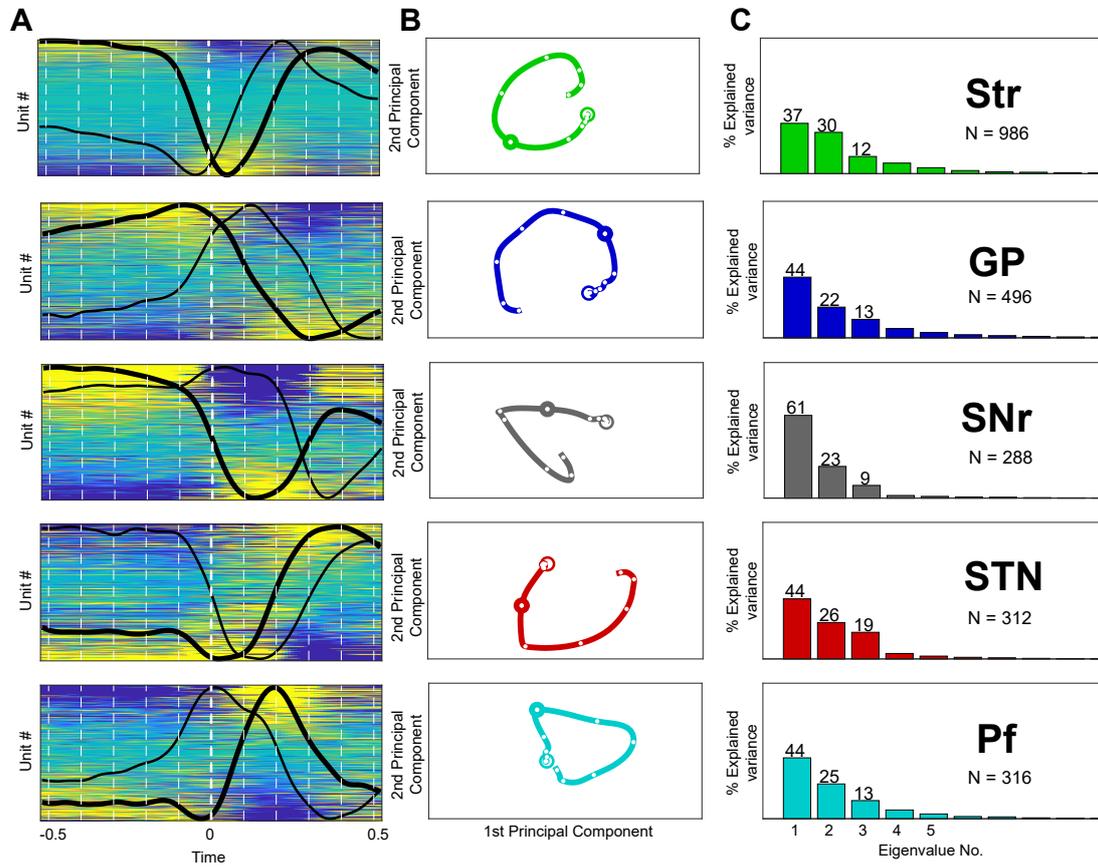


Figure 4.2: Summary of the PCA of the data. A. Heat maps showing the z-score of each population sorted by the components of the first eigenvector of the covariance matrix. The first and second principal components (PCs) are shown in thick and thin black lines. Dashed white lines indicate 100ms intervals. The firing rates are aligned to movement initiation as indicated before. B. Two dimensional trajectories. Movement time is marked by a big circle. 100ms intervals along the trajectory are indicated by white dots. c. Percentage of explained variance for each region. Each panel contains the total number of recorded neurons.

# Chapter 5

## The geometry of neural trajectories

### 5.1 Chapter overview

In this chapter we study the main aspects of the shape of the low dimensional PCA neural trajectories, which we call the *principal trajectories*, in the basal ganglia during movement initiation.

We start by introducing a rigorous account of the aspects that influence the shape but are outside traditional frameworks of shape analysis (see for example [86]). Doing this considerably simplifies our endeavor in the forthcoming section and give us a feeling of what aspects to pay attention to in the analysis.

It is important, for example, to understand what are we expecting to see in the average firing rate in terms of the firing rate patterns of individual trials; it turns out that the patterns in the average data are low pass filtered versions of the single trial patterns and the filtering depends, as expected, of the distribution of the individual events in each trial. In general, we find out that events far away only contribute irregular/sharp points that are irrelevant due to the fact that the corresponding explained variance is negligible.

A very important and almost always ignored aspect in the study of these trajectories is the presence of irregularities; such fact invalidates any other analysis that depends on the differential geometry of curves. We show how to avoid these difficulties.

In section 5.3 we develop a novel account on curvature. It turns out that curvature is the most conspicuous aspect of the shape of the trajectories (others being, for example, the speed). We study how the covariance or overlap between the individual patterns in simplified situations contribute to the curvature; the results, some of which are counter-intuitive, pave the way for the future multivariate approach in terms of ensembles.

We develop then in section 5.4 the main study of the shape in terms of com-

plementary views of the shape information. We reserve the term shape to account for the aspects of the trajectories that are invariant to certain transformations or group actions (i.e. rotations should not change the shape). We coin the term “form” to local aspects of the shape like points of interest.

Finally, the question of the constraints on the shape by the PCA computation is approached reaching important conclusions about the expected trajectories and eigenvalue distributions for random firing rates. This result has not been previously published in the literature.

## 5.2 Preliminary aspects of the shape of the principal trajectories

### 5.2.1 The geometric nature of the neural trajectory

Understanding the geometric nature of the principal trajectories should be the first step in the study of their shape. By geometric nature we mean, where, in the mathematical landscape dictated by geometry, do they lie as mathematical objects. The main result here is that they are not curves in the traditional use of the term: they contain irregularities.

**Definition 4** (Neural trajectory). *The neural trajectory is defined as the set of points traversed by the firing rates of  $n$  recorded neurons in the neural space. Mathematically, it is given by the map  $\gamma : [a, b] \rightarrow \mathbb{R}^n$ , expressed in coordinates as*

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_n(t))^T, \quad (5.1)$$

where we assume  $a = -\epsilon$  and  $b = \epsilon$ .

For the purposes of this work we understand each coordinate of the curve as the *estimated firing rate* of the  $i$ th neuron in some interval  $[-\epsilon, \epsilon]$  around an alignment event. More specifically, our input data is the result of three different operators applied to the raw spike trains: Alignment, Expectation and Filtering:

$$\hat{\gamma}_i(t) = \left\langle \int_{-\infty}^{\infty} A_e [K(t - \tau) s(\tau)] d\tau \right\rangle, \quad (5.2)$$

where

$$s(t) = \sum_k \delta(t - t_k) \quad (5.3)$$

is the spike train at a given trial and  $\langle \cdot \rangle$  is the expectation operator. The filtering kernel is usually Gaussian or square, or a composition of both. The conventional choice, the Gaussian kernel, has immediate consequences that are stated

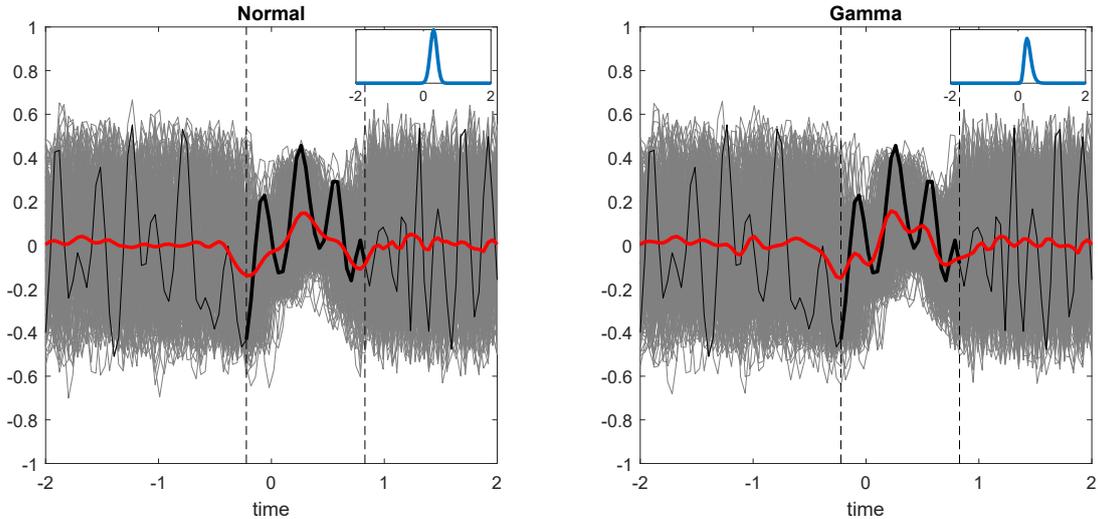


Figure 5.1: Averaging effect of responses locked to an event occurring at times drawn with a normal (left) and gamma (right) distributions with mean  $\mu = 0.3$  and standard deviation  $\sigma = 0.2$ . The red trace is the average of 200 trials, the black trace is the original pattern. Activity outside the dotted lines is not locked to any particular event.

as a lemma below and whose proof can be easily deduced from the regularizing properties of Gaussian convolutions or the heat kernel (this means, it takes any discontinuous functions and makes it smooth)

**Lemma 1.** *If the kernel  $K(t, s)$  is Gaussian, then the map  $\gamma(t)$  is continuous and smooth with high probability*

Therefore, such map satisfies the definition of a smooth curve given in chapter 3. The alignment operator is given by a translation in the time domain  $A_e[f(t)] = f(t - t_e)$ . A consequence of this is that neural responses to an event present at some time  $t_a$ , relative to the alignment time  $t_e$  and drawn from some probability distribution, will be low-pass filtered versions of the putative stereotypical response of that neuron.

As an example, consider figure 5.1. Inside the region bounded by the dotted lines there is a simulated neural response locked to an event occurring at times  $t_a$  drawn from a normal (left) and gamma (right) distribution. The red trace is the average over trials and the black trace is the original pattern; note that the former is a low pass filtered version of the later. The gamma distribution was chosen to approximate the reaction time distributions which are usually asymmetric [83]. Patterns with no particular time locked response will cancel out (region outside

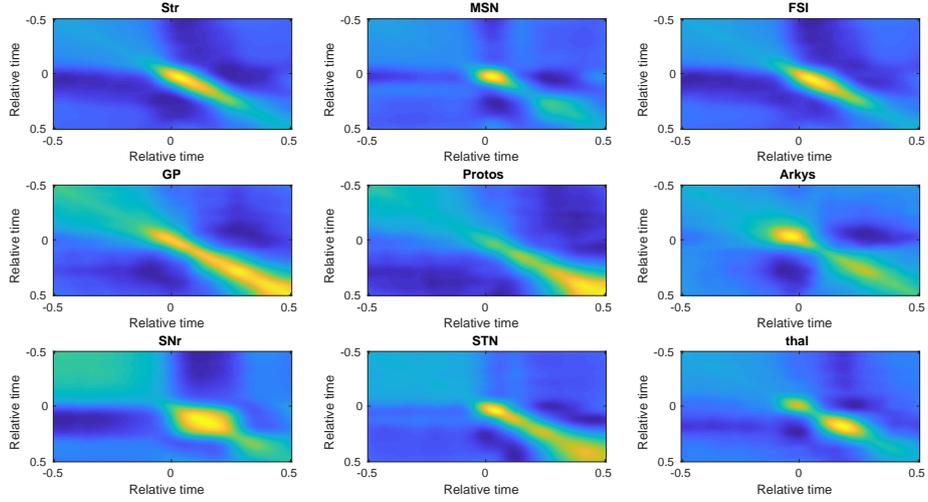


Figure 5.2: Time-Covariance matrices of the different regions. Each element of these matrix corresponds to the dot product of two population vectors at different time points.

the dotted lines).

The resulting firing rate will be composed then of patterns with different degree of filtering and attenuation depending upon how well time-locked are they to the alignment event. Our trial structure, shown in figure 4.1, gives a series of potential alignment events; details can be found in [83]. For the purpose of this chapter, we align the firing rates to the *Nose out* event and, hence, the responses to *GO* will be filtered by the distribution of reaction times (RT). Accordingly, the responses to *side in* will be filtered by the distribution of movement times (MT).

The effects of such filtering can be seen in the Autocovariance operator  $\mathbf{G}$  (figure 5.2). the diagonal of the matrix is given by the length of the population vector at each time (i.e.  $\mathbf{G}_{ii} = \|\gamma(t_i)\|^2$ ); we call this the *activity mass* at time  $t_i$ . Prior to the movement event ( $t = 0$ ), the activity mass is small compared to the time around movement, additionally, the extend of the correlations is greater for times close to movement. This has the consequence that *activity prior to movement is bound to explain less variance relative to the activity around movement*.

Also from the covariance matrix we can describe broadly the patterns of activation of different regions during movement. Those range from approximately uniform mass around and after movement (GP, protos, STN), point masses (MSN,

STN, Arkys) and broad, concentrated masses (STN, Str). This preliminary observation will be refined later in the chapter.

## 5.2.2 The low dimensional PCA trajectories can be irregular

Now we focus on pathological aspects of the trajectories. The main one observed in our data set is *irregularity*.

**Definition 5** (PCA trajectory). *Let  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$  be the eigenvectors of the covariance matrix  $\mathbf{C}$  associated with the two largest eigenvalues. The PCA trajectory associated to the neural trajectory  $\gamma(t)$  is the projection of the later in the subspace spanned by  $\mathbf{u}$  and  $\mathbf{v}$ , that is*

$$z(t) = (\mathbf{u}^T \gamma(t), \mathbf{v}^T \gamma(t))^T. \quad (5.4)$$

The PCA trajectory is then two dimensional. Remember that a regular curve is one in which the tangent vector  $\|\dot{\gamma}(t)\|$  never vanishes (a point traveling along the trajectory never “stops”). It turns out that PCA trajectories are, in general, irregular. It is imperative then to deal with violations to regularity at the outset if we want to be able to use the bulk of results from differential geometry<sup>1</sup> available in the literature as they apply, invariably, only to regular curves (for example, the curvature is undefined at those points).

What are the sources of irregularity in the PCA trajectory? Trivially, any moment in which the high dimensional population goes silent is an irregular point. That is an unlikely situation with real data as neurons have a nonzero probability of firing at any time in the interval during each trial.

A more troublesome source of irregularity comes from the situation illustrated in figure 5.3, we call this the *dimensionality cathastrophe* alluding to a similar situation found in the singularity theory of differentiable maps [14]. It comes from the fact that, by definition, some of the variance of the curve will be along directions orthogonal to the PCA plane and therefore, their images will be overlapping in the low dimensional projections (see figure 5.3).

There are two qualitatively different situations in which this occurs in our data and which will be dealt with differently.

**Definition 6** (Minor singularities). *Minor singularities are due to responses locked to distant events. This are called minor because they are unlikely to contribute in*

---

<sup>1</sup>There is a beautiful theory of irregular curves but its use is not considered necessary and, therefore, goes beyond the scope of this monograph

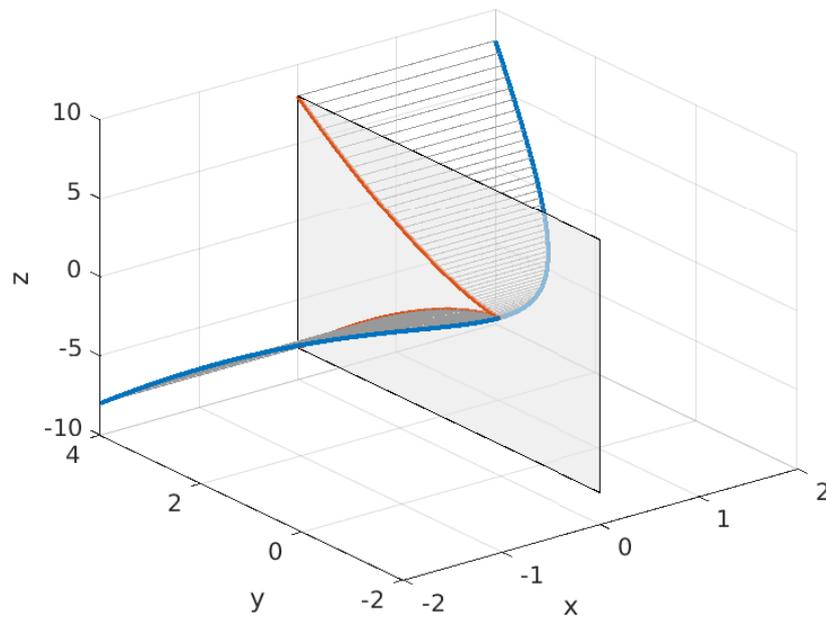


Figure 5.3: Irregularity coming from singularities due to low dimensional projections. The curve in blue is  $\gamma(t) = (t, t^2, t^3)$ . The low dimensional projection onto the plane  $x = 0$  (orange) displays a type of singularity called a cusp. There, the tangent vector vanishes.

any significant way to the PCA trajectory. Contributions to the trajectory are understood in terms of arc length (see below).

As we saw, responses locked to distant events (i.e. GO) are bound to explain less variance as they are attenuated by the corresponding relative event times distributions. Therefore, most of this variance will *live* in the *orthogonal complement* of the principal subspace. The practical consequence of this is a high frequency oscillation characterized by many irregular points where the trajectory stops as it traverses the corresponding hidden dimensions (figure 5.4). Minor singularities are, therefore, expected byproducts of the alignment process.

**Definition 7** (Major singularities). *Major singularities are all those singularities that are not explained by the averaging process.*

A pragmatic definition characterization of singularities, amenable to numerical purposes, is given now. Remember that the arc length function of the curve  $z$  is given by

$$z(t) = \int_{-\epsilon}^t \|\dot{z}(\tau)\| d\tau. \quad (5.5)$$

The total length  $L$  of the curve is then  $z(\epsilon)$ . We characterize singular points by a slowing down of the curve below a threshold given as a fraction of the total length of the curve

**Definition 8** (Irregular point). *An irregular point of the curve  $z$  is a  $t_i \in [-\epsilon, \epsilon]$  such that*

$$\|\dot{z}(t_i)\| < \eta L. \quad (5.6)$$

The irregular points of the trajectories for each region are shown in figure 5.4. The names *irregular point* and *singularity* are used interchangeably in this text. It is worth asking where slow down comes from. From now we can partially answer that question by observing the arc length function of the PCA trajectory in comparison with the arc length function of the full trajectory (figure 5.5).

The initial slowing down, due to the low pass filter mentioned above is observed in the high dimensional trajectory (dotted line), however, it does not stop when the PCA trajectory does, suggesting that at those times *the trajectory visits dimensions that are not accounted for the principal subspace.*

### 5.2.3 Geometric pre-processing: reparametrization and smoothing

We now eliminate the singular points for the rest of the analysis. Furthermore, we ignore any speed information as it does not contribute to the shape of the curve (as

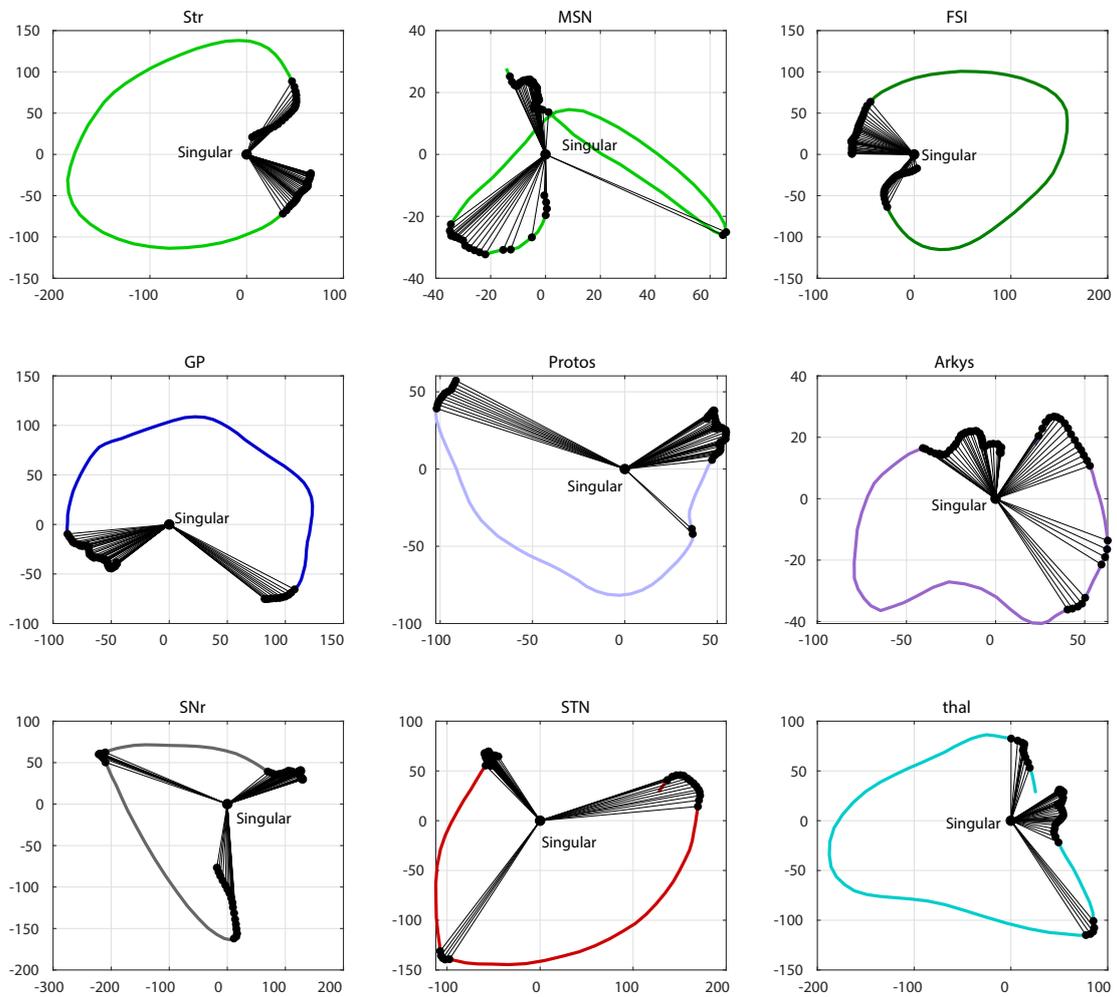


Figure 5.4: Singularities of the different regions for  $\eta = 0.01$ , that is, points that contribute less than 1% to the length of the curve. Some of them, the ones corresponding to the initial segments will be removed in future analyses

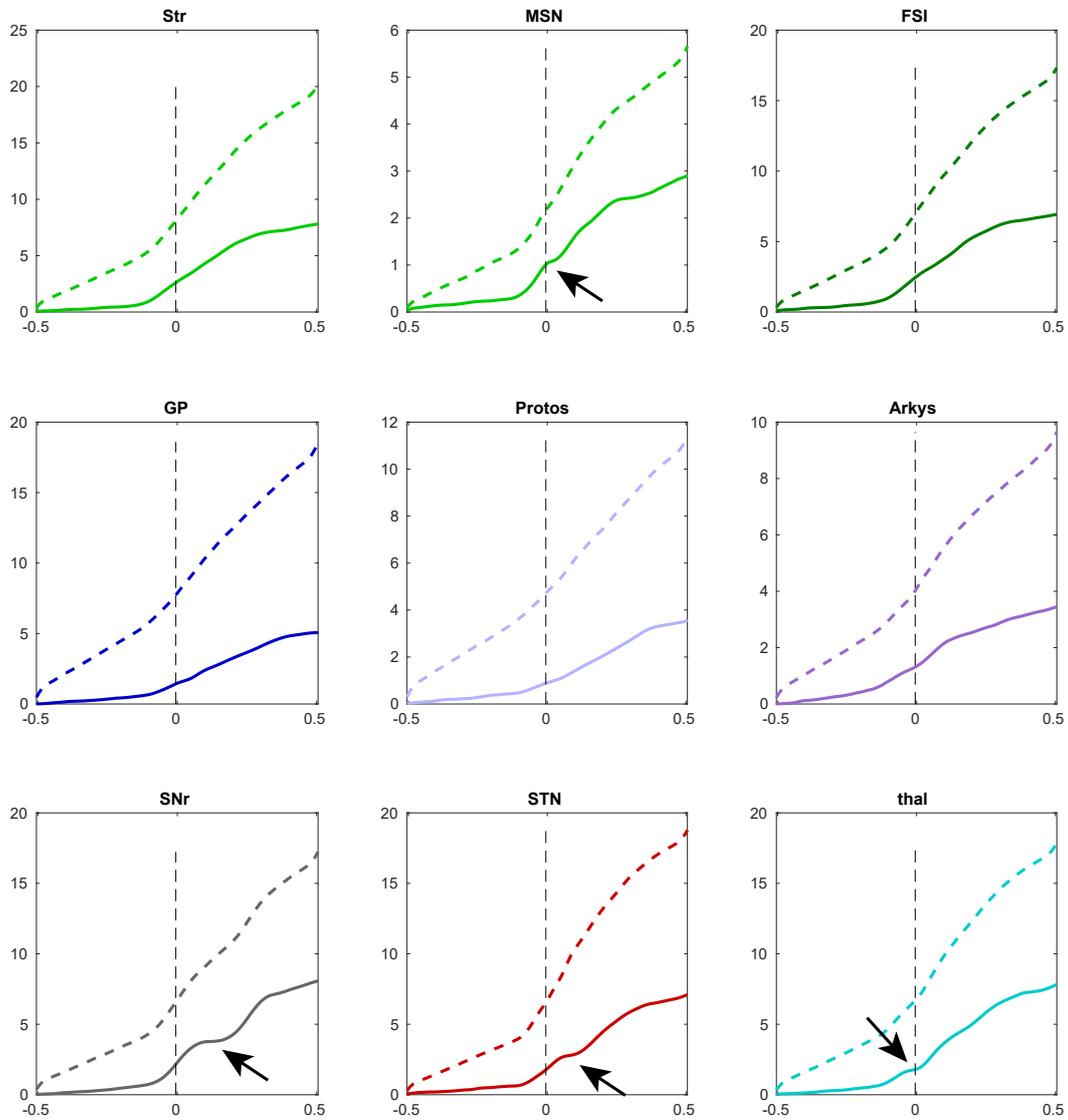


Figure 5.5: Arc length of the PCA trajectory (solid line) and of the neural trajectory (dotted line) as a function of time. Observe the times at which the PCA trajectory stays still (arrows).

it is well known, see chapter 3). We preserve, nevertheless, the profiles in figure 5.5 as a guideline about how to traverse the shape once analyzed in the forthcoming sections. The algorithm for reparametrization of the curve is given in section 3, now we concentrate in the question of how such procedure affects the individual firing rates and, more importantly, the covariance structure of the data.

Consider figure 5.6. The arc length reparametrization algorithm gives a map  $\phi : [0, L] \rightarrow [-\epsilon, \epsilon]$  such that  $\|\dot{z}(t)\| = 1$  for all  $t \in [-\epsilon, \epsilon]$ . Now we show the following theorem:

**Theorem 4.** *The covariance matrix  $\mathbf{C}$  remains unchanged after reparametrizations.*

*Proof.* The entry  $\mathbf{C}_{ij}$  of the covariance matrix is given by

$$\mathbf{C}_{ij} = \int_{-\epsilon}^{\epsilon} \gamma_i(\tau)\gamma_j(\tau)d\tau.$$

but  $\tau = \phi(s)$ ,  $s \in [0, L]$ . Note that  $L = \phi^{-1}(\epsilon)$  and  $0 = \phi^{-1}(-\epsilon)$  (because the reparametrization is invertible, section 3), therefore

$$\mathbf{C}_{ij} = \int_0^L \gamma_i(\phi(s))\gamma_j(\phi(s))\dot{\phi}(s)ds = \int_{-\epsilon}^{\epsilon} \gamma_i(t)\gamma_j(t)dt,$$

the last term due to the substitution rule of integration. □

Consequently, the eigenvectors are the same and the principal components are just reparametrizations of the original PCA. From now on we will assume arc-length parametrized curves. In figure 5.6, we show the result of the reparametrization of the STN's trajectory. The distortion due to the reparametrization in the firing rate is shown in the right most panel. Observe that this distortion is only due to a warping of the time domain; the previous result should be apparent from the fact that the warping is the same *for all neurons*.

As it turns out, shape is a low frequency phenomenon [90], therefore, we remove high frequency details of the trajectory. This is done by the following procedure. First we compute the *angle function* of the curve  $\theta(t)$ . It is defined as the angle made by the tangent vector and the abscissa at each time point. As can be seen in figures 5.7, 5.8 and 5.9, the angle shows high frequency oscillations at the beginning of the curve and here and there in the middle. Filtering those out will give a bandlimited version of the trajectory [9] and will leave the shape unchanged.

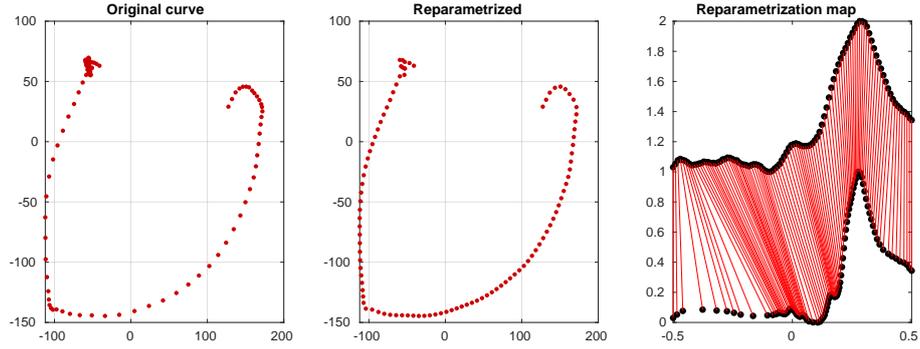


Figure 5.6: Effect of the reparametrization on the firing rates. Left. Original curve, the dots show the non-uniform parametrization. Center. Reparametrized curve traversed at a unit speed. Right. The reparametrization map

Also in the middle panels of those figures, we can observed jumps in the turning angle; those correspond to sharp points that will be defined later. The procedure of low pass filtering is performed then iteratively using an Savitzky–Golay filter. The iterations are made through segments bordered by the sharp points. This avoid higgs-like artifacts in those locations and allows us to preserve the sharp points for future analysis.

Let  $\theta'(t)$  be the filtered angle function, the curve can be recovered then by integration

$$x = c_x + \int \cos(\theta(s))ds$$

$$y = c_y + \int \sin(\theta(s))ds$$

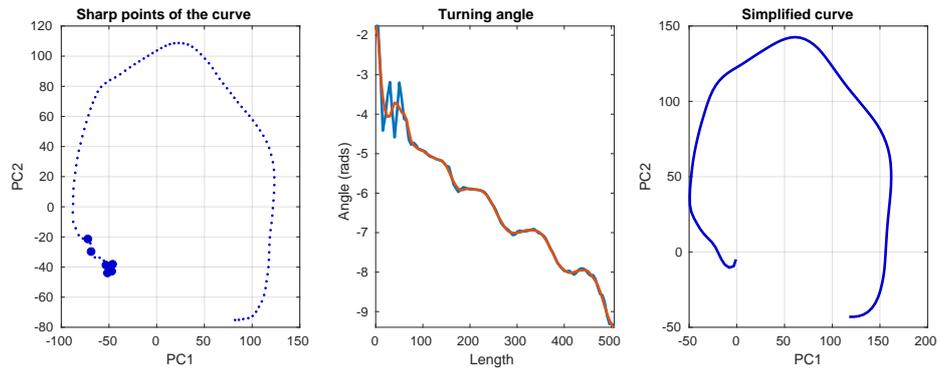


Figure 5.7: Simplification of GP trajectory. Left, original trajectory showing the sharp points and high frequency features. Center. Angle function and its smoothed version using a Savitzky–Golay filter of frame size 11 and degree 3. Right. Simplified curve.

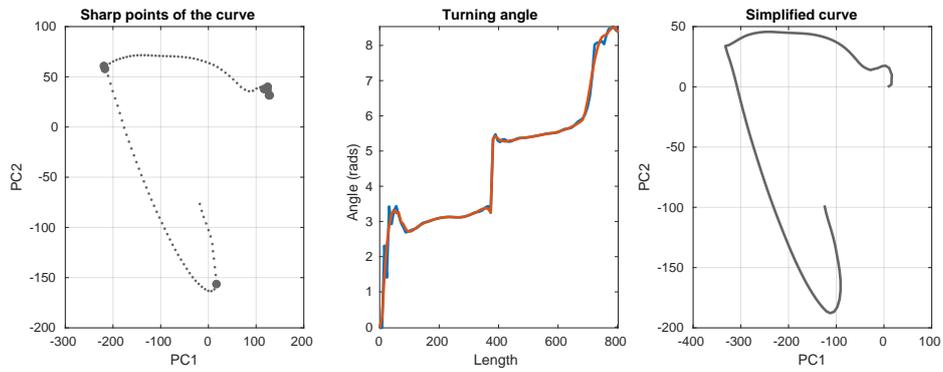


Figure 5.8: Simplification of SNr trajectory. Note that the sharp points are preserved.

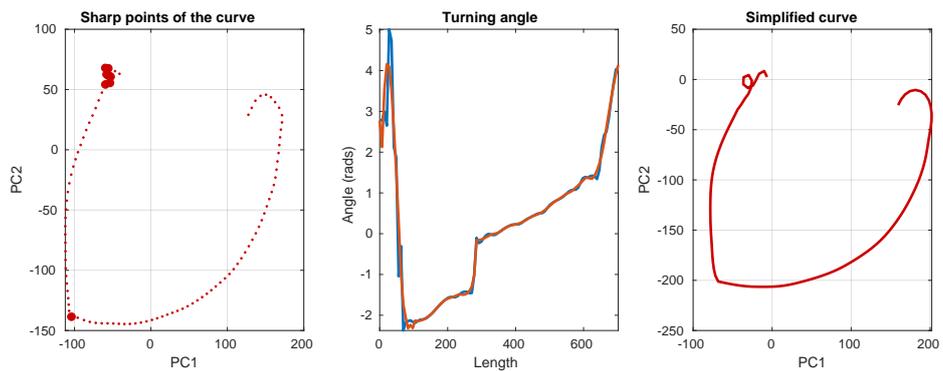


Figure 5.9: Simplification of STN trajectory

### 5.3 The curvature

Our interest is mainly on the shape of curves from the perspective of neural codes. Representing a neural code as a trajectory or a curve means that we are interpreting each neuron as a dimension and, in the curve, as a coordinate of the given parametric curve  $z(t)$ . However, many geometric quantities are not readily accessible to an interpretation from the coordinate point of view. That is the case of the curvature for which there is no formula, to the best of our knowledge, that relates the two coordinates in a simple, intuitive way. We are interested, in particular, in the relation between curvature and the degree of “overlap” between the individual patterns displayed by the individual coordinates (i.e. firing rates, principal components).

In this section we focus in the arc-length parametrized, simplified, 2 dimensional PCA trajectory. The curvature is defined as the rate of change of the angle function, that is the angle between the tangent  $\mathbf{t}(s)$  and an arbitrary vector, with respect to the length  $s$ .

**Definition 9.** *Given a vector  $\mathbf{u}$ , the function*

$$k(s) = \frac{d\theta}{ds} \tag{5.7}$$

*is called the curvature of the curve  $\gamma \subset \mathbb{R}^2$ , with tangent  $\gamma'(s) = \frac{d\gamma}{ds}$ .  $\theta(s)$  is the angle between  $\mathbf{u}$  and  $\gamma'(s)$ , that is  $\theta(s) = \cos^{-1}\langle \gamma'(s), \mathbf{u} \rangle$*

when taking  $\mathbf{u} = \mathbf{e}_i$ , equal to one of the canonical basis vectors, we get

$$\cos \theta(s) = \langle \gamma'(s), \mathbf{e}_1 \rangle = \gamma'_1(s),$$

that is, for a unit speed curve, the cosine of the angle with the first canonical basis vector is the first component of the tangent to the curve. Therefore

$$\frac{d \cos \theta}{ds}(s) = -\sin \theta \frac{d\theta}{ds}(s) = \frac{d^2 \gamma_1}{ds^2}(s) = \gamma''_1(s),$$

But, in the unit speed curve,

$$\sin \theta = \gamma'_2(s),$$

and, consequently

$$\hat{k}(s) = -\frac{\gamma''_1(s)}{\gamma'_2(s)}, \tag{5.8}$$

which we name the **coordinate description of the curvature** because it helps understanding how the individual profiles of the the components produce the final

observed curvature (The standard formula is discussed in the preliminaries). An immediate insight is that, in terms of neurons and firing rates, increases of curvature happen when one of the neurons' firing rates accelerates faster than the other. Of course there is a dual definition based in the complementary ratio

$$\hat{k}^*(s) = -\frac{\gamma_2''(s)}{\gamma_1'(s)},$$

The relation between these two curvatures becomes apparent by the fact that the tangent and the normal to a unit speed curve are othogonal, therefore

$$\gamma_1''\gamma_1' + \gamma_2''\gamma_2' = 0,$$

from which we get

$$\frac{\gamma_1''}{\gamma_2'} = -\frac{\gamma_2''}{\gamma_1'}$$

or  $k(s) = -k^*(s)$  for all  $s$ .

**About the sign of the curvature** Depending upon the agreed direction of positive curvature, the appropriate fraction should be used that correspond to the desired sign.

Note that either function becomes undefined at the extreme values of each of the coordinates  $\gamma'_{1,2}(t^*) = 0$  and therefore, neither of them is an accurate representation of the curvature at those points, however, it is possible to recover the original curvature by inserting the appropriate points instead of the singularities

$$k(s) = \begin{cases} \hat{k}(s), & \text{if } \gamma_2'(s) \neq 0 \\ -\hat{k}^*(s), & \text{otherwise} \end{cases} \quad (5.9)$$

In the following examples we will study different scenarios for 2 dimensional curves.

**Example 1.** *The circle is the only curve that accelerates as fast as it speeds. Indeed, for the unit circle parametrized by  $\gamma(t) = (\cos t, \sin t)$ ,*

$$k(s) = -\frac{-\cos t}{\cos t} = 1,$$

*as expected.*

**Arbitrary parametrizations** For an arbitrary parametrization  $\gamma = \gamma(t(s))$ , we have  $\frac{d\gamma}{ds} = \frac{d\gamma}{dt} \frac{dt}{ds} = \frac{1}{\|\dot{\gamma}\|} \frac{d\gamma}{ds}$  (which can be also derived from the traditional formula for the cosine), therefore

$$\frac{\dot{\gamma}_2}{\|\dot{\gamma}\|} \frac{d\theta}{ds} = \frac{d}{ds} \left( \frac{1}{\|\dot{\gamma}\|} \dot{\gamma}_1 \right). \quad (5.10)$$

**Example 2.** For a general circle parametrized by  $\gamma(t) = (R \cos t, R \sin t)$ ,

$$k(s) = \frac{1}{R} - \frac{-\cos t}{\cos t} = \frac{1}{R},$$

where  $R$  is the radius of curvature.

**Example 3.** Whenever  $\gamma_1(t) = \pm \gamma_2(t)$ ,  $\frac{1}{\|\dot{\gamma}\|} \frac{d\gamma_1}{dt} = \text{sgn}(\gamma_1)$ , therefore,

$$k(s) = 0.$$

**Example 4.** A slightly more complicated but yet analytically treatable example, comes from assuming partially overlapping firing rates shaped as cosine bumps:  $\gamma_i(t) = \cos(t - \mu_i)$ ,  $t \in (-\pi/2, \pi/2 + \mu)$ . Setting  $\mu_1 = 0$  and  $\mu_2 = \mu$ , replacing in equation 5.10, we get

$$k(s) = -\frac{\langle \dot{\gamma}, \ddot{\gamma} \rangle \sin t - \|\dot{\gamma}\|^2 \cos t}{\|\dot{\gamma}\|^3 \sin(t - \mu)},$$

which, after some massaging becomes

$$k(s) = -\frac{\sin \mu}{\|\dot{\gamma}\|^3}, \quad (5.11)$$

with

$$\begin{aligned} \|\dot{\gamma}\|^2 &= 1 + \cos \mu [\cos \mu - 2 \cos t \cos(t - \mu)] \\ &= 1 + \cos \mu [\cos \mu - 2\gamma_1(t)\gamma_2(t)] \end{aligned} \quad (5.12)$$

The theoretical and computed curvatures are shown in figure 5.10. The extreme curvature points coincide with the extreme points of the covariance function  $C_{12} = \gamma_1(t)\gamma_2(t)$  of the two neurons. Indeed,

$$\frac{dk}{dt} = 0 \Rightarrow \cos(2t - \mu) = 0$$

therefore, in the selected domain,

$$t = \frac{\pi n + \mu}{2}, n \in \{-1, 0, 1\}$$

are the values at which the extreme points lie for both functions. The latter is the crossing point between the two bumps (see figure 5.10)

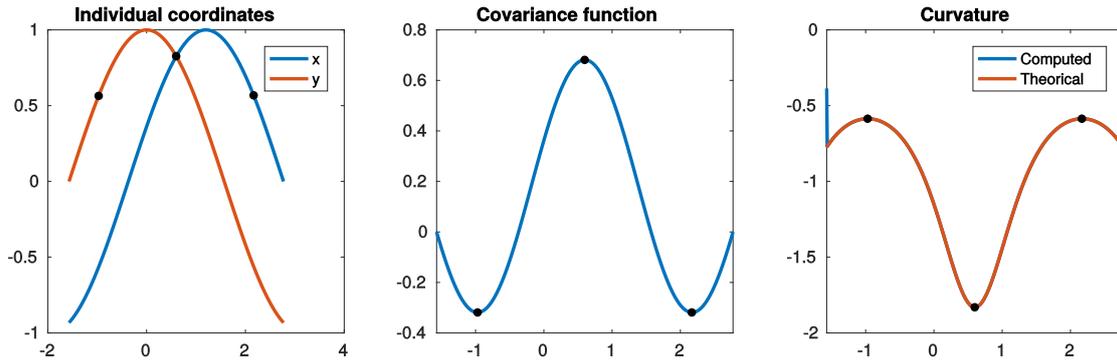


Figure 5.10: Curvature of bumps. For simple bumps the curvature reflects the covariance functions to some extent.

The corresponding trajectory in example 4 is shown in figure 5.12. To fully appreciate our interpretation of curvature in the example, we need to reparametrize the curve (see preliminaries). The reparametrization is shown in figure 5.11. Note that the reparametrization map (left) induces a distortion of each of the components of the trajectory (i.e. the individual firing rates), in order to traverse it at unit speed; it is nevertheless, the same curve (Aspects of invariance are studied in previous section).

Under this map, it can be seen that the normal (right, dashed line) is always smaller (in absolute value) than the tangent in points with low curvature; this can be also seen in figures 5.12 and 5.13. The numerator on  $\hat{k}(s)$  can be seen as the "desire" of the curve to turn in the new direction. The denominator can be seen as its par with respect to continuing or keep going. The dual  $\hat{k}^*(s)$  is interpreted along the same lines thinking about the numerator as the desire of "not turning", heralded by the activity of the first neuron vs the "desire" of not keep going. In both cases the balance of those forces gives a perfectly circular trajectory.

### 5.3.1 Simulation of specific patterns of curvature

The curvature of the trajectory reflects the nature of the overlap between the individual firing rates. Consider idealized gaussian bumps as neural activations. We parametrize the different patterns using 5 parameters: width  $\sigma$ , amplitude  $A$ , relative phase  $\phi$ , orientation, asymmetry and number of bumps. Of those, the asymmetry is the only one that will not be modeled by a gaussian distribution; instead, a specific distribution will be created for that purpose. In this section we study the behavior of the curvature as one firing rate pattern changes its phase with respect to another, under different values of the mentioned parameters.

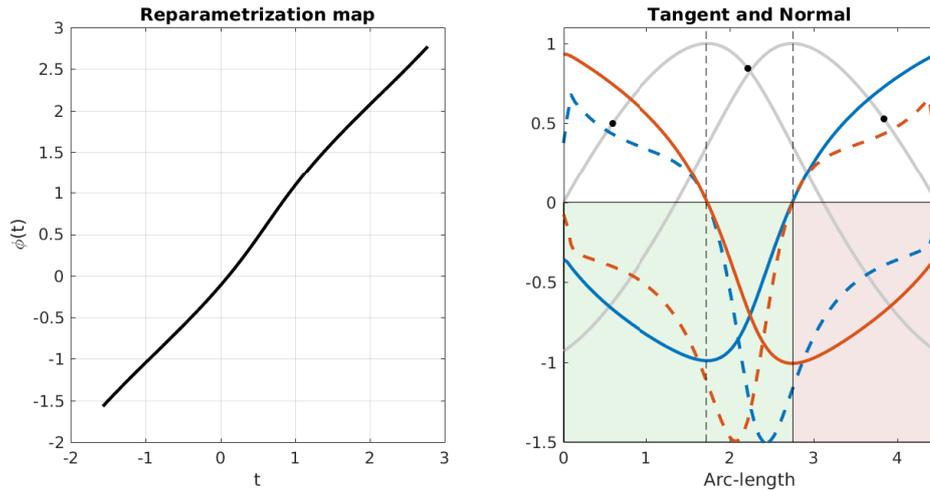


Figure 5.11: Details on the interpretations of the curvature. Continuous colored lines correspond to the tangent's coordinates. Dashed lines the normal. The red region corresponds to a part in which the direction of comparison is inverted due to a change in direction.

**General shifts** The simplest simulation is, again, cosines overlapping. We already saw that the curvature is related to the covariance in some cases. In the general case, formula 5.9 allows to tell when the curvature is going to be high: when the speed of change is slow in one coordinate but there is an extrema in the other.

**Simple gaussians** By fixing all the parameters and changing only the overlap, in a gaussian bump, we obtain general intuitions of the behaviour of the curvature as a function of the overlap (figure 5.15).

**Different widths** The case of bumps with different widths is illustrated in figure 5.17. In this case, the thin one will drive the curvature on top of the slowest change of the wide one.

**More than one bumps and different amplitudes** Different amplitudes shear the shape towards the coordinate with more mass. The effect of multiple bumps can be ambiguous but when they are of higher frequency than the single pattern it reduces to the previous case (figure 5.16)

**Asymmetric** The asymmetry in shape drives the curvature in different ways depending of the part that is overlapping with the symmetric pattern (figure 5.18)

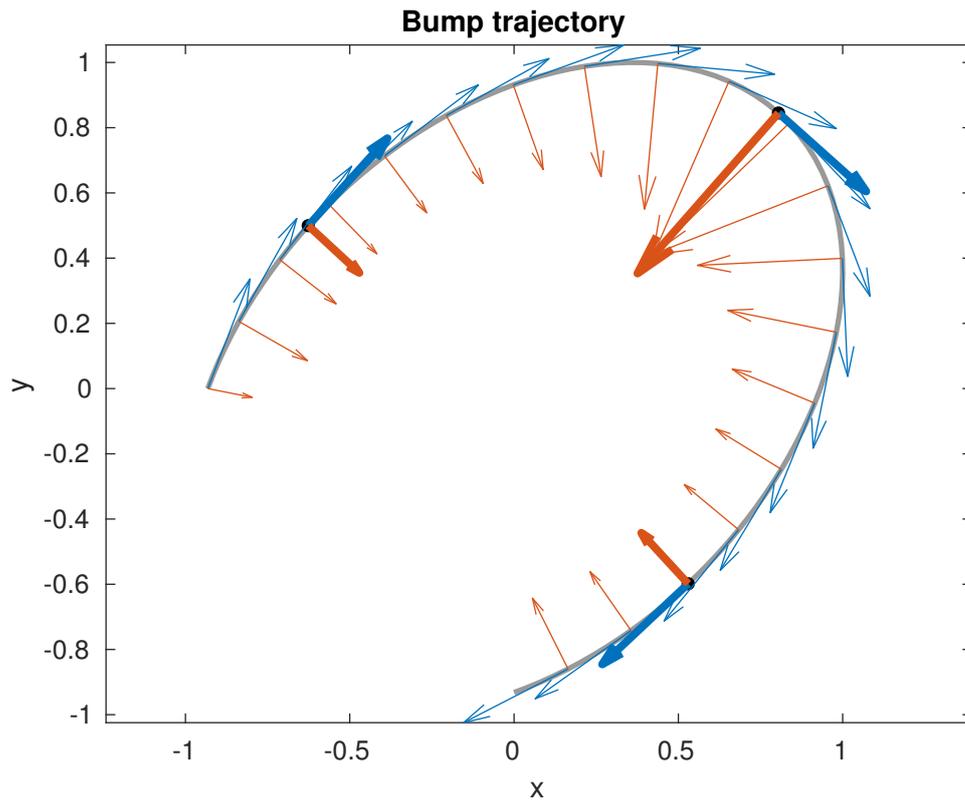


Figure 5.12: Trajectory of the overlapping cosine bumps trajectory. The points of high curvature are characterized by a normal that is bigger than the corresponding tangent. The tangent represents the “desire” to keep going in the same direction while the normal is the “desire” to turn.

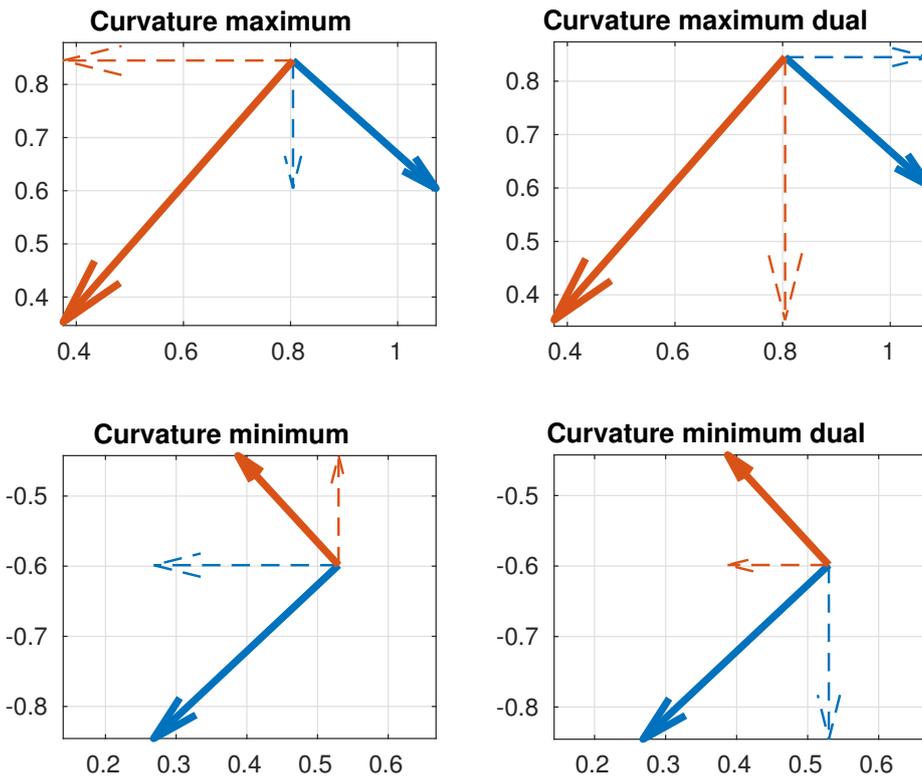


Figure 5.13: The points of high curvature (top panels) are characterized by a high "turning" vs. "keep going" ratio (left) or a low "not turning" vs. "not keep going" one (right). Low curvature ones have higher "keep going" components and therefore tend to be closer to a line.

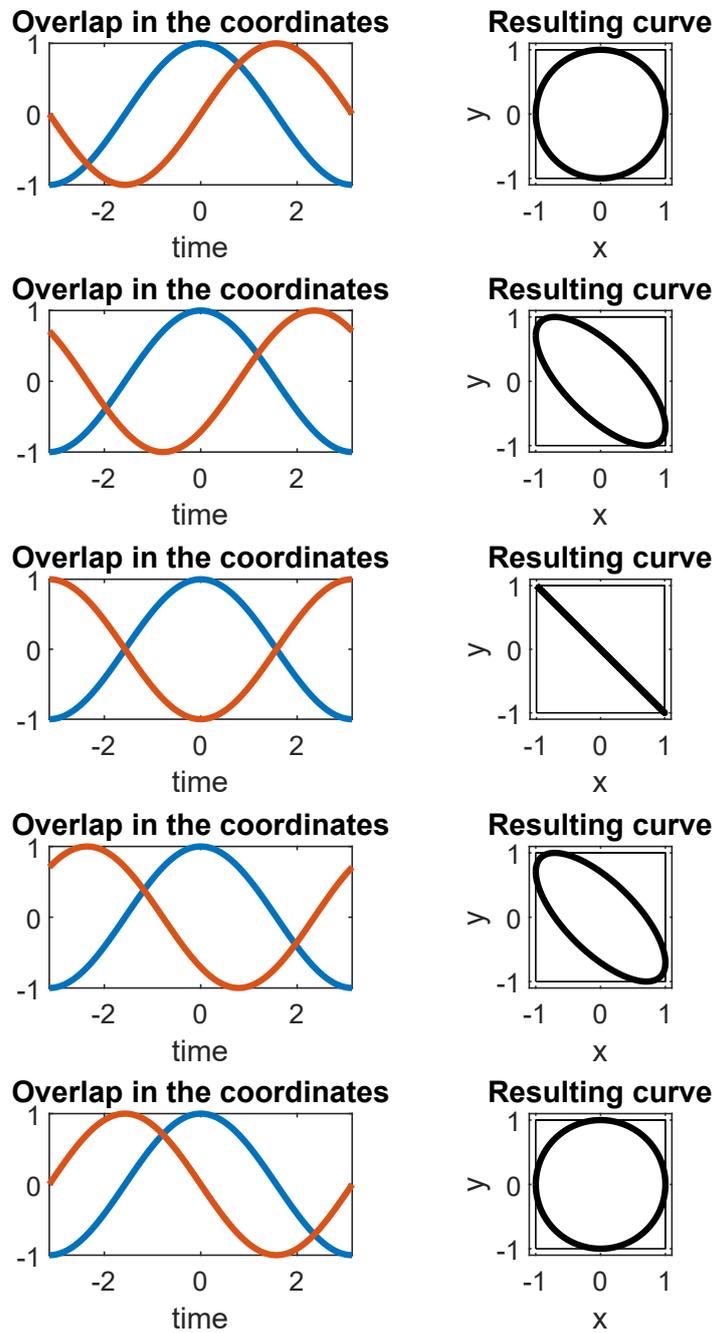


Figure 5.14: The relative phase between two sinusoidal waves give rise to curvature depending on how they overlap.

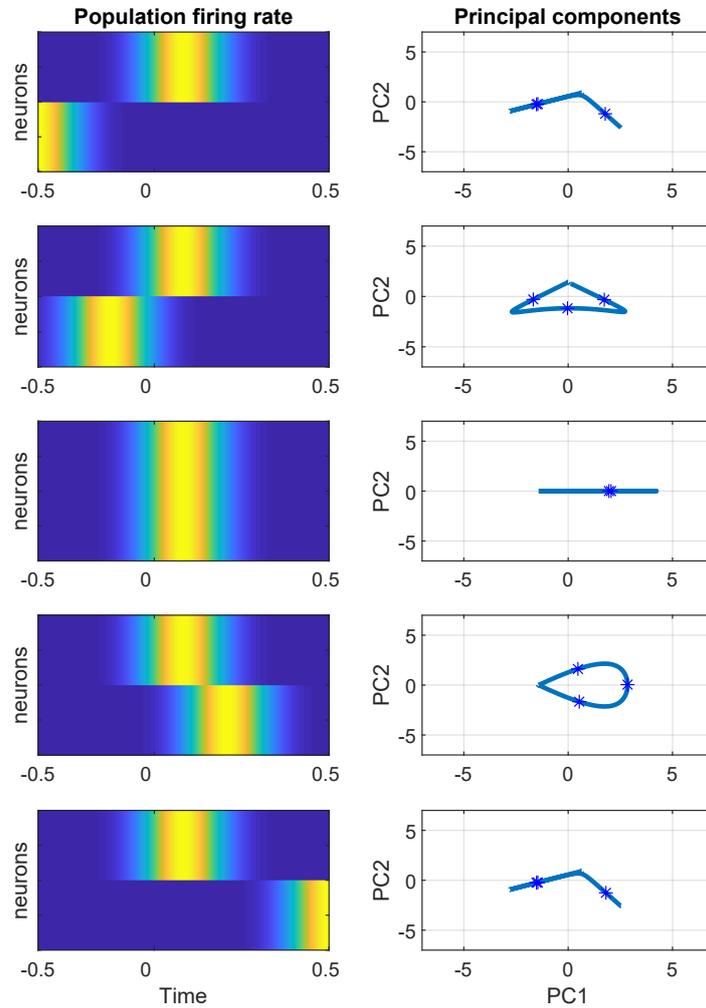


Figure 5.15: Two gaussian bumps with different overlaps. The trajectory is shown to the left. Observe that the trajectory at first, when the two bumps have almost not overlap, has two branches which are irregular, that is, the trajectory stops and backtracks to return to the origin as each traverses each bumps. A different kind of irregularity appears when the two bumps are completely correlated, there the trajectory is a line that is traversed forward and back depending upon the specifics of the firing rate.

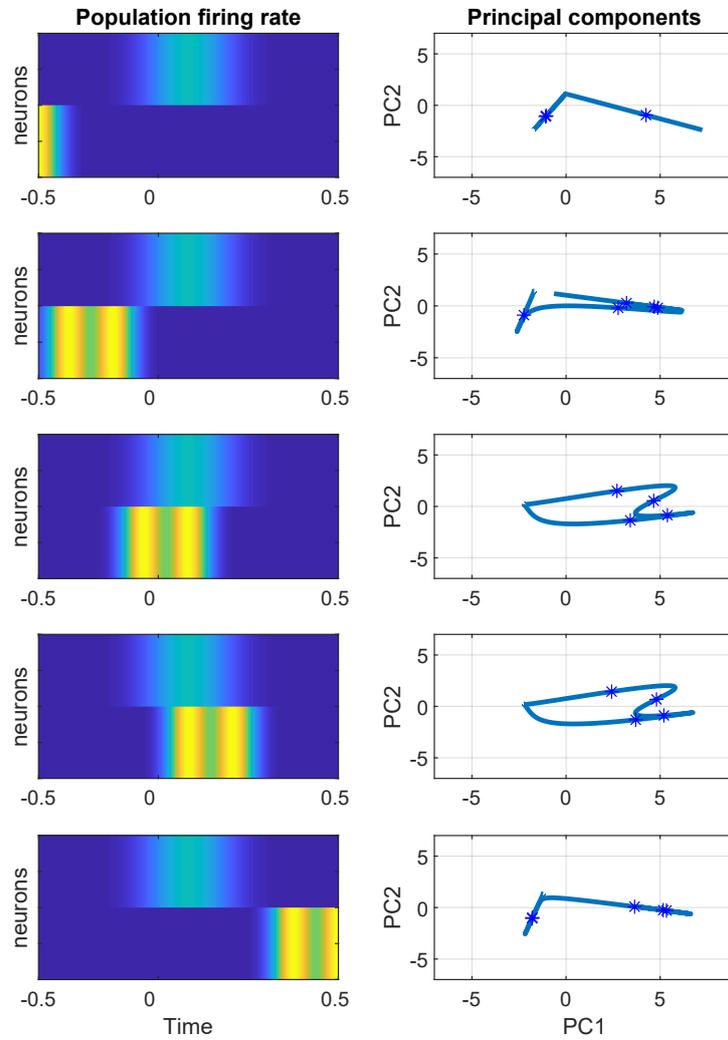


Figure 5.16: Two bumps against one. In this case note that the curvature of the trajectory reflects the second fast changes of the thin bumps at some times. The parts that are fully synchronized with the wide bump reflect a line but the part that is out of sync has a smaller curvature

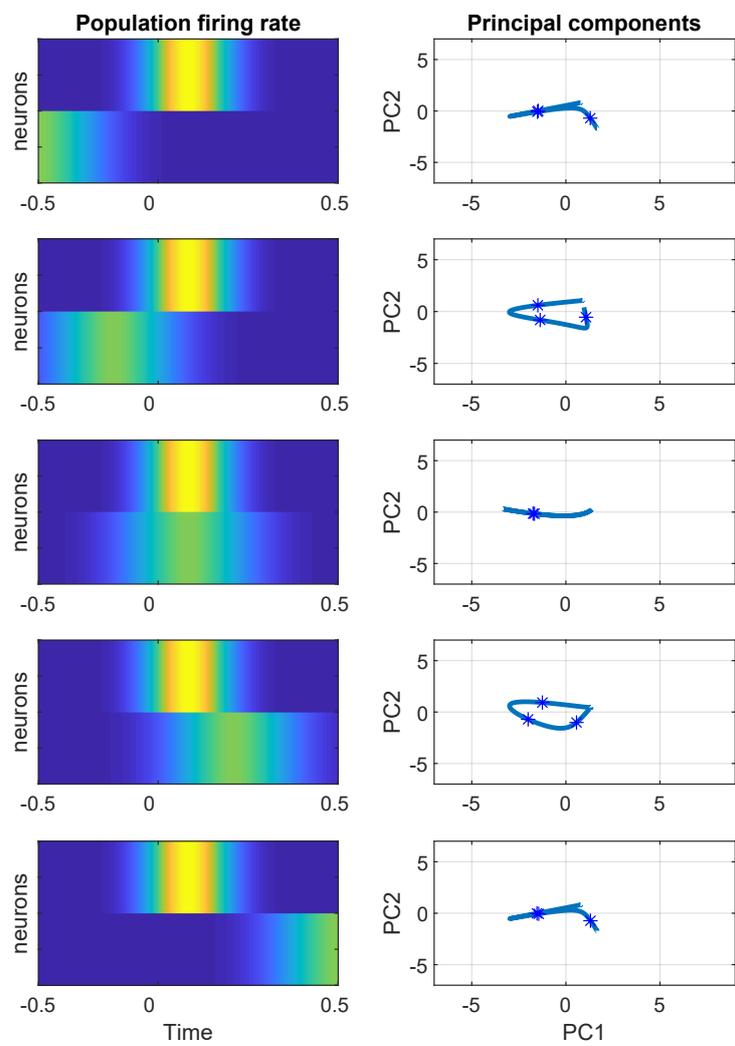


Figure 5.17: Two bumps of different widths

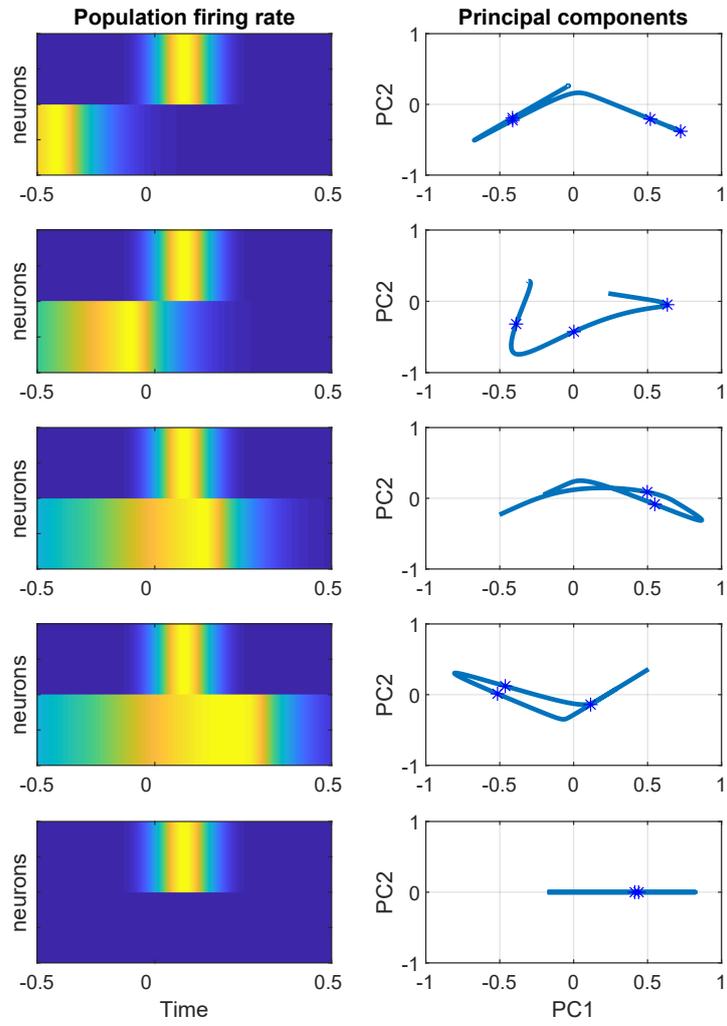


Figure 5.18: One symmetric bump against an asymmetric one

## 5.4 The shape of the neural trajectories

In this section we study the shape and the form of the trajectories. The difference is made on account of two different purposes and underlying theories with overlapping schools of thought. The former adopts a geometric invariance point of view [86], the later stems from the study of singularities or catastrophes in differentiable maps [90].

The shape is then defined as what is left once all the non-shape related information is *factored out*; its aim is global in the sense that each shape is a point in an abstract space[86]. This factoring out is rigorously implemented by grouping the trajectories in equivalence classes that are invariant to transformations encoded as group actions[86].

The form is local, focused on the specific features like sharp points, singularities or flat regions. It is discrete in the sense that focuses in the points of interest but continuous in the sense that it studies the behavior of the trajectory in the vicinity of those points. From the point of view of this chapter, the shape approach will be used with comparison purposes while the form will guide the study of the activity of specific ensembles in relation to a putative population code.

### 5.4.1 Low dimensional invariance

We deal now with the shape of the low dimensional PCA trajectory. We created already in the previous section our first equivalence class. The shape is invariable to reparametrizations, therefore, all the reparametrizations of a given curve have the same shape; are in the same class. The class is characterized as follows: Given two trajectories  $z_1, z_2 \in C([- \epsilon, \epsilon], \mathbb{R}^2)$ , if there exist a diffeomorphism  $\phi : [- \epsilon, \epsilon] \rightarrow [- \epsilon, \epsilon]$  such that  $z_1(t) = z_2(\phi(t))$  for all  $t$ , then  $z_1$  and  $z_2$  have the same shape. The arc length parametrization is our *representative* of the class.

This condition is sufficient but not necessary; there are other transformations that preserve the shape and that should be accounted for. They are illustrated in figure 5.19. The first step in that direction is to choice an appropriate representation of the shape that facilitates the construction. If we accept the arc length parametrization as an acceptable preprocessing step, the angle function  $\phi(t)$  introduced in the previous sections constitutes a natural choice for the representation.

We now construct what is called the *pre-shape* space by dealing with the different group actions.

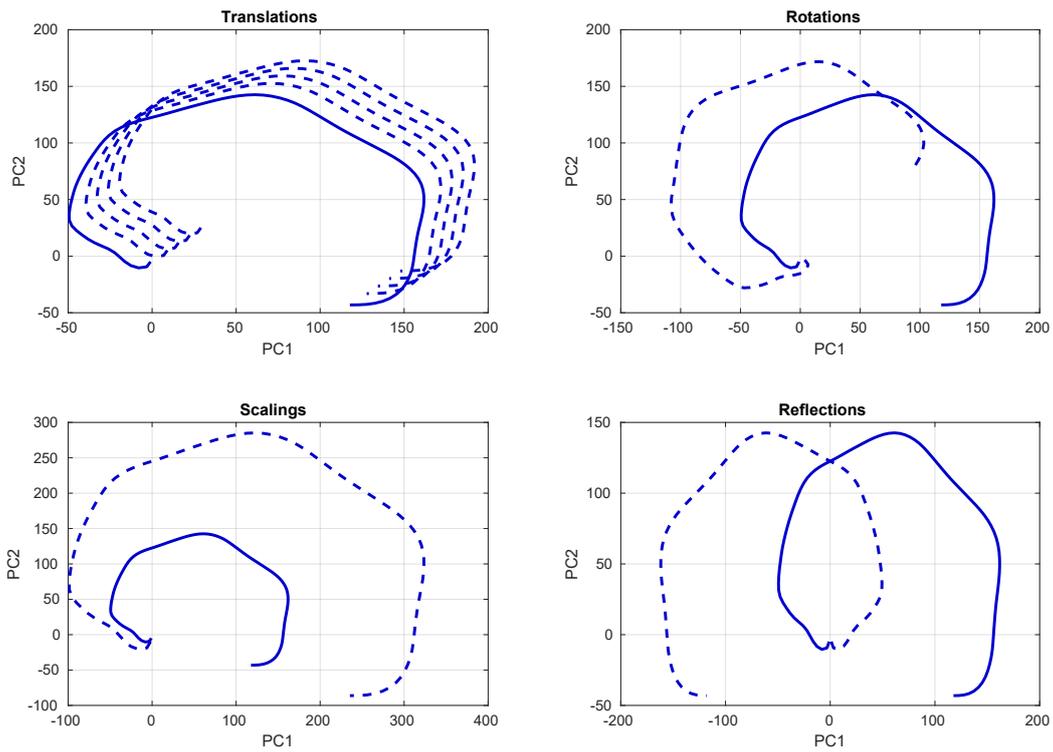


Figure 5.19: Example of shape preserving transformations for the GP. From top left clockwise. The translations of the trajectory, the rotations, scaling and reflections; all give an equivalent shape

Translation group:

The angle function is already invariant to translations. Indeed, let  $z_1(s) = z_2(s) + c$ ,  $c \in \mathbb{R}^2$ . Then, the tangent vectors coincide  $\dot{z}_1(s) = \dot{z}_2(s)$  for all  $s$ .

Scaling group:

The quotient with the scaling group is taken by normalizing all curves to have length  $L = 2\pi$ .

Rotation group:

Rotations act like translations of the angle function [86], that is, a rotation by an angle  $\phi_0$  is given by  $\phi'(t) = \phi(t) + \phi_0$ . We normalize then so that the *average rotation* is  $\pi$ , that is

$$\frac{1}{2\pi} \int_0^{2\pi} \phi(s) ds = \pi. \quad (5.13)$$

Symmetric group:

Not all the reflections should be identified. A more detailed account on reflections will be give in the next section. The reflection group of interest is the one generated by taking the canonical basis as root vectors. It is isomorphic to  $\mathbf{Z}_2$ . We account for this invariance by constructing the *orbifold* of the given space including in the representation all  $2^2$  reflections in the corresponding mirrors or by picking up a representative using the results of future section about concave/convex shapes.

We have constructed then the following space:

**Definition 10** (PCA trajectory preshape space). *The space*

$$\mathcal{O} = \left\{ \phi \in \mathbb{L}([0, 2\pi], \mathbb{R}) : \frac{1}{2\pi} \int_0^{2\pi} \phi(s) ds = \pi. \right\} \setminus \mathbf{Z}_2 \quad (5.14)$$

*is called the PCA preshape space.*

We use this space to make comparison between shapes and to do statistics on shapes. In order to compute the similarity between shapes we need to measure the length of geodesic paths between them. Fortunately, we have the following theorem [86].

**Theorem 5.** *Geodesics between shapes  $\phi_1$  and  $\phi_2$ , in the space  $\mathcal{O}$ , are given by straight lines of the form*

$$\alpha(\tau) = (1 - \tau)\phi_1 + \tau\phi_2, \quad (5.15)$$

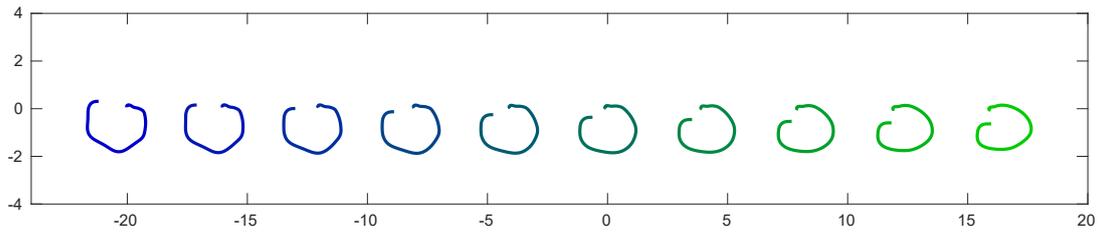


Figure 5.20: Path between Str and GP. The fact that the transformation is barely noticeable implies that they are close in the shape space

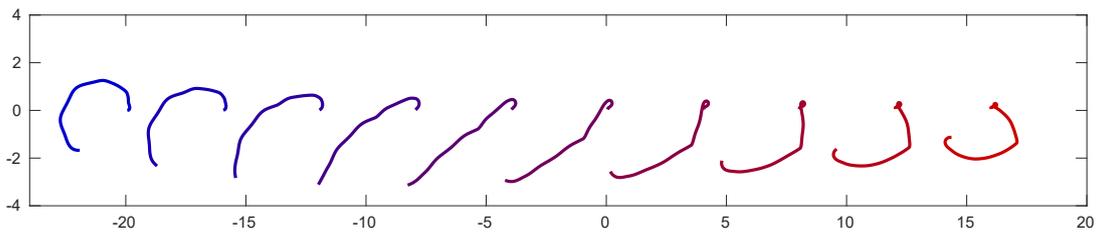


Figure 5.21: Path between GP and STN. The transformation between the GP and STN is more dramatic, notice that this is the shortest path in the shape space

with  $\tau \in [0, 1]$ .

Geodesic path between Str and GP is shown in figure 5.20. The smooth transition between them mean that they are close in the shape space. On the other hand, the transition between STN and GP is forced to pass through a line in order to get to the other shape. Even though discontinuous functions are perfectly allowed in the definition of the space, this behavior could be an artifact caused the sharp point in the STN (artifact in the sense that the geodesic is forced to go out of the space). By smoothing out the sharp points we found that the showed paths remain unchanged.

The pairwise comparison of shapes shown in figure 5.22 reveal an interesting pattern and is the more important result of this section: The trajectories form to clear groups, STN, SNr, Arkys and MSNs against GP, Str, FSIs, Protos and Pf. Interestingly *The FSIs and the whole striatum have indistinguishable shapes*. Similarly, *GP and Protos have indistinguishable shapes*. Both cases refer as a subpopulation having the same shape than the whole.

In order to gain a better understanding of those shape relations we turn now to a more detailed study of the shape. We aim to answer the following questions in the sections to come: What are the defining aspects of those groups? Could

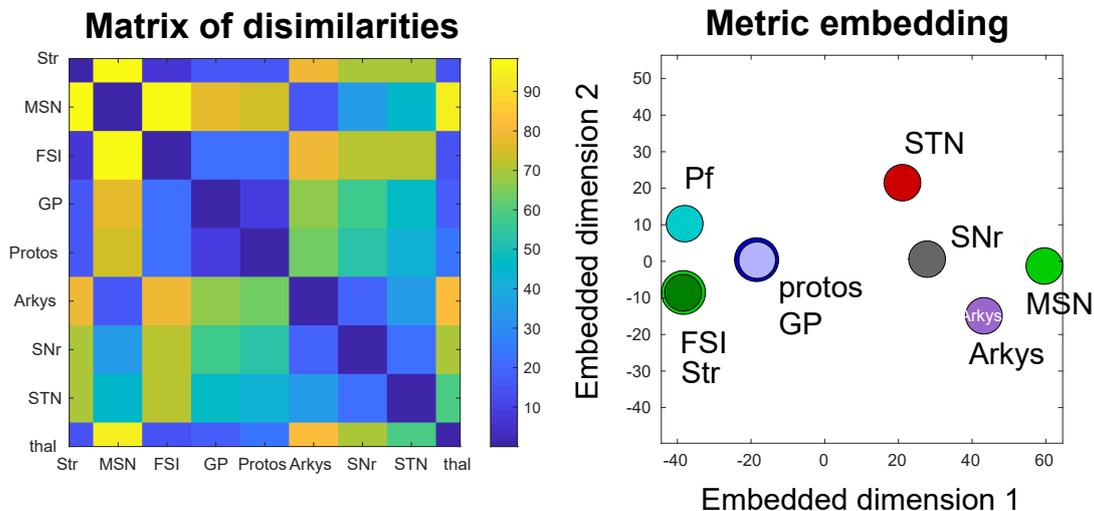


Figure 5.22: Left. Dissimilarity matrix of the trajectories in the shape space. Right. Multidimensional embedding in a two dimensional plane that approximately preserves the distances. Note the 2 clusters and the overlap between sub-population and super populations in proto-GP and FSI-Str

those similarities in shape be interpreted in terms of the neural code?

### Elastic shapes and spheres

One important caveat that can be observed in figure 5.21 is that the path that includes a shape with a sharp point is forced to open up and the close again. This phenomenon might make the path longer that it should be, intuitively, and is caused by the rigidity in the parametrization. Other examples of theoretical shape spaces are studied in the appendix

### 5.4.2 The form of the trajectory: salient features.

Now we study the form, the salient features of the shape of each trajectory. The more important information about a curve, as was seen before, is the curvature; it does not come as a surprise then that the morphology of the trajectory is studied by taken a closer look at the curvature. We assume that points at which the curve turns are points of interesting events in the population code. In particular we are interested in 4 kinds of points that, along with the singular points studied before, give a complete account of the different features of the curve: sharp points, vertices, flat points and arcs.

**Definition 11** (Sharp points). *A sharp point is defined as a discontinuity in the turning function  $\phi(s)$ . Specifically, a value of the parameter  $s = s_0$  such that*

$$\lim_{s \rightarrow s_0^-} \phi = \phi_1 \neq \phi_2 = \lim_{s \rightarrow s_0^+} \phi. \quad (5.16)$$

This theoretical definition is rarely satisfied but a more practical description comes in terms of the derivative:

**Definition 12** (Sharp point criterion). *A sharp point is a point  $\phi(s_0)$  that satisfies*

$$\dot{\phi}(s_0) > \delta/h, \quad (5.17)$$

where  $\delta$  is an arbitrary angle and  $h$  is the step size for a given partition of the interval  $[-\epsilon, \epsilon]$ .

In summary, a sharp point is a point that turns an angle higher than  $\delta$  in a single step (in terms of the arc length) of size  $h$ .

We give now a characterization of the next feature of interest in terms of 3 equivalent definitions

**Definition 13** (Vertex). *A vertex is a point with 4 order contact with its osculating circle*

The details in this definition can be seen by plotting the osculating circles at each point of the curve 5.23; at these points the trajectory is rounder and they are indications of turns that are less obvious than the sharp points. A device useful for the identification is called the *evolute* which is the locus of all the centers of curvature; more details about evolutes are given in the preliminaries.

Figure 5.24 shows the evolute of all the target trajectories and the corresponding vertices. The following result is apparent

**Lemma 2.** *The vertices are singular (sharp) points of the evolute.*

Also apparent from the figure is the fact that the different regions have different vertex structure. Vertices at the end of the curve are likely due to filtered out activity of future events. In order to identify the vertices we look for maxima of the curvature which, in the case of a unit speed curve is the magnitude of the acceleration. The figure 5.25 shows the identified vertices of the regions. Those of greater interest are the ones that are not close to the extremes we will call them *internal*. Based on internal vertices we have this characterization:

- No vertices: SNr, STN: We do not count points twice, therefore, the points classified as irregular are not included

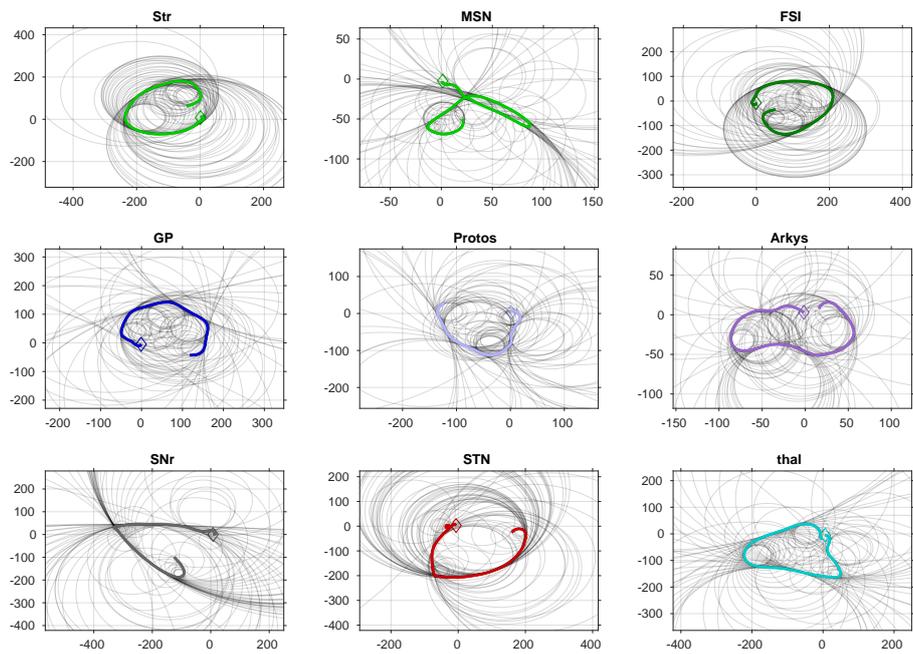


Figure 5.23: Contact with the osculating circles. The osculating circles are tangent circles with radius  $1/k$  where  $k$  is the curvature of the curve. Circles with higher contact are accentuated.

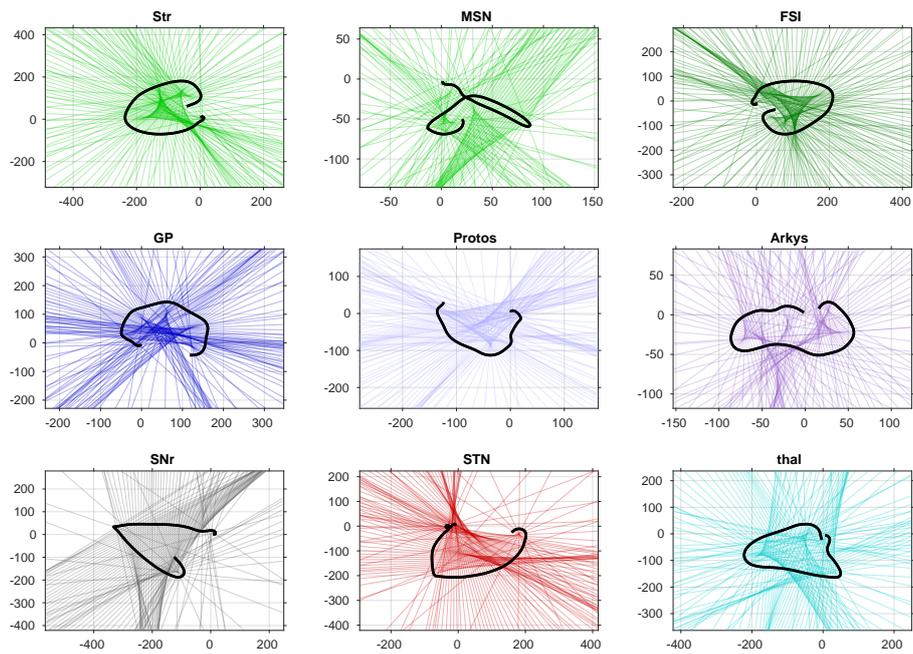


Figure 5.24: Evolute of the different regions. The evolute is generated as the envelope of the normals to the curve at each point. Due to the complex structure of the trajectories, an explicit computation of the evolute is unfeasible.

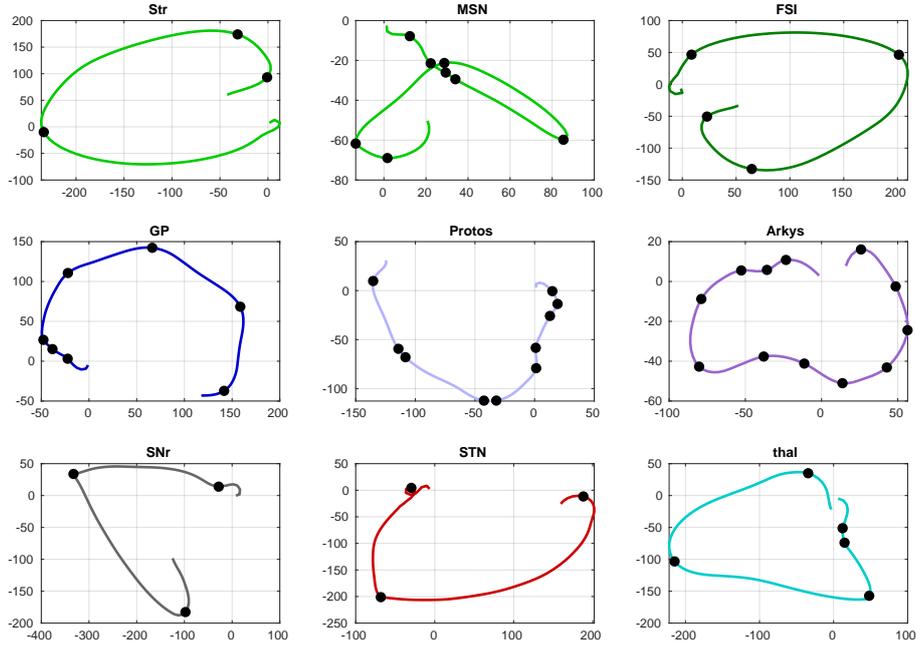


Figure 5.25: Vertices of the different trajectories. Internal vertices are more likely to be related to a putative neural code.

- Two vertices: Str, MSN and FSI
- Three vertices: GP, Protos and Arkys. Importantly, while the 3 vertices of the GP are convex, both, in protos and arkys there are concave regions bordered by 2 vertices that we are counting here as 1.

The rest of the points in the curve are classified either as flat or generic.

**Definition 14** (Flat point). *Flat points are points with high contact with the tangent line at that point.*

A flat point is then points where the trajectory is very similar to a line. Qualitatively it is important to know whether the flat points are inflection points or higher flat points. A good criteria was developed in the preliminaries and is stated here;

**Definition 15** (Flat point criterion). *A flat point  $s_0$  is an inflection point iff*

$$k(s_0) = 0 \text{ and } \dot{k}(s_0) > 0. \quad (5.18)$$

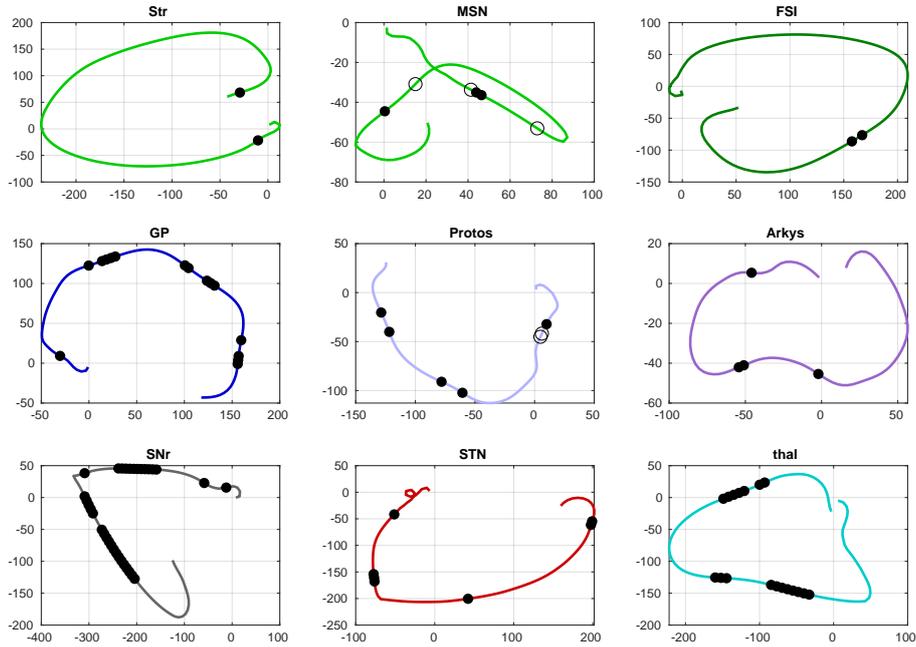


Figure 5.26: Flat regions

*It is a higher flat point iff*

$$k(s_0) = 0 \text{ and } \dot{k}(s_0) = 0. \quad (5.19)$$

The flat points of the regions are shown in figure 5.26. Inflection points are plotted as open circles, higher points as filled ones. The SNr, GP and Pf stand out as having *flat regions*. Their significance in the code will be studied in the next section.

**Definition 16** (Generic arcs). *Points not classified either as singular, vertex or flat, are called generic arcs.*

## 5.5 Null shapes

In this section we address the question: what is the expected shape for a random trajectory? We argue that such shape is heavily constrained by the correlations induced in the time domain. We start by exploring significant and analytically tractable situations some of which could be considered the null model for the trajectory. We then derive precise formulas for the principal components and the eigenspectrum under two different null models: Gaussian and exponential covariance matrices. We show to which extent the data fits into each of those models.

The general strategy followed in this analysis is to transform the *functional PCA* problem into an equivalent heat or diffusion equation and apply the known harmonic theory of the laplacian operator to find a solution.

### 5.5.1 PCA of gaussian SDFs

The computation of the Spike Density Function of a spike train can be seen as some sort of *cooling* of point heat sources in a spatial domain. Consider figure 5.27. In the usual physical situation, we start with a spike of high temperature at a given position of space  $x$  and want to know what is the distribution of heat after a time  $t$ . Let such distribution be  $u(x, t)$ ,  $(x, t) \in \mathbb{R} \times (0, \infty)$ , then the problem is solved by finding the solutions to the partial differential equation

$$\partial_t u - \kappa \partial_{xx} u = 0 \quad (5.20)$$

where  $\partial_t u = \frac{\partial u}{\partial t}$  and  $\kappa$  is the thermal diffusivity. We assume that the spatial domain is one-dimensional.

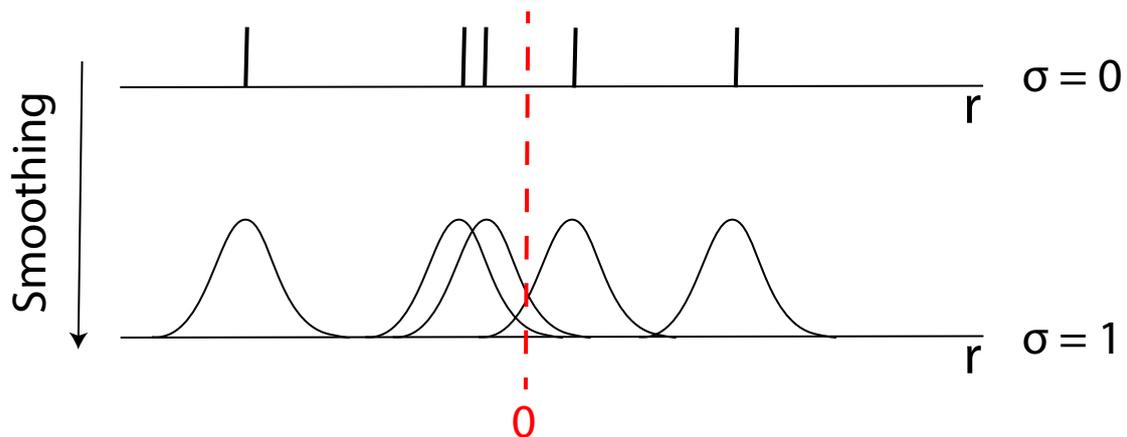


Figure 5.27: The concept of heat kernel smoothing.

In order to put the problem of smoothing in terms of the heat equation we treat the time relative to a given event (cue, movement, etc...) as a spatial domain and call it  $r$ . The new “time domain” will be the kernel bandwidth  $\sigma$ . In those terms, the initial value problem for a firing rate  $u(r, \sigma)$  (firing rate at position  $r$  relative to the event and smoothing level  $\sigma$ ) becomes

$$\partial_\sigma u - \kappa \partial_{rr} u = 0 \quad (5.21)$$

$$u(r, 0) = f(r) \quad (5.22)$$

The function  $f(r)$  is our initial, unsmoothed spike train. It is known that the general solution to this equations has the form

$$u(r, \sigma) = \int K(r - r', \sigma) f(r') dr', \quad (5.23)$$

where

$$K(r - r', \sigma) = \frac{1}{\sqrt{4\pi\kappa\sigma}} \exp(-(r - r')^2/\kappa\sigma) \quad (5.24)$$

is the heat kernel. Notice that for  $\sigma = \sigma'^2$ ,  $u(r, \sigma)$  is the SDF of our spike train.

We now study the effect of smoothing on PCA. For simplicity, we will smooth Gaussian noise, the previous observations apply just as well. For a population of  $N$  noisy neurons with i.i.d spike counts  $x_i(r)$  with  $\mathbb{E} x_i = 0$  and  $\mathbb{E} x_i(s)x_i(t) = \delta(s-t)$ , the covariance function of the corresponding SDFs is

$$\begin{aligned} C(r, r', \sigma) &= \frac{1}{N} \mathbb{E} \sum_i \int K(r - \tau, \sigma) x_i(\tau) d\tau \int K(r' - \tau', \sigma) x_i(\tau') d\tau' \\ &= \frac{1}{N} \sum_i \int \int K(r - \tau, \sigma) K(r' - \tau', \sigma) \mathbb{E}[x_i(\tau') x_i(\tau)] d\tau d\tau' \end{aligned}$$

which, due to the i.i.d assumptions becomes

$$C(r, r', \sigma) = \int K(r - \tau, \sigma) K(r' - \tau, \sigma) d\tau.$$

The integral can be evaluated on the interval  $\tau \in (-\infty, \infty)$  to get

$$C(r, r', \sigma) = \frac{1}{4\sqrt{2\pi\sigma}} \exp(-(r - r')/2\sigma) \quad (5.25)$$

which is again a heat kernel. The advantage of this approach is now obvious. From the last section we know that the heat kernel can be decomposed as

$$C(r, r', \sigma) = \frac{1}{c} \sum_k e^{-\lambda_k \sigma} \phi_k(r) \phi_k(r'), \quad (5.26)$$

where  $\phi_k(r)$  are eigenfunctions of the Laplace operator. They are also our principal components. More precisely, it is clear that they are solutions to the eigenvalue equation

$$\int C(r, r', \sigma) \phi_k(r') dr' = \mu_k \phi_k(r), \quad (5.27)$$

with eigenvalue  $\mu_k = e^{-\lambda_k \sigma}$ . The principal components are therefore

$$\gamma_k(r) = \sqrt{\mu_k} \phi_k(r). \quad (5.28)$$

It can be shown that computing the population covariance and projecting the firing rate on the principal directions is equivalent to this result. In order to get specific eigenfunctions, we need to consider the interval around the event and some boundary conditions on the extremes.

**Theorem 6.** *For an interval of length  $L$  with boundary conditions  $u(-T/2, \sigma) = 0$  and  $u(T/2, \sigma) = 0$ , for all  $\sigma > 0$ , the eigenvalues and eigenfunctions are*

$$\lambda_k = \frac{k^2 \pi^2}{L^2} \quad (5.29)$$

and

$$\phi_k(n) = \frac{2}{L} \sin\left(\frac{k\pi n}{L}\right). \quad (5.30)$$

Note that the frequency of the eigenfunctions do not depend upon the properties of the kernel. In order to test the theoretical predictions we simulated  $N = 100$  noisy neurons in  $T = 100$  trials and computed the covariance function and the eigenfunctions. The theoretical and empirical results agree as expected (see figures 5.28 and 5.29).

The trajectories and eigenvalues are shown in figure 5.30. For sufficiently large number of samples/neurons, the empirical trajectories approach the theoretical result.

### PCA of a random walk

The previous analysis can be extended or even interpreted as the case in which the population, as a high dimensional trajectory, makes a random walk in the high dimensional neural space. In that case, as can be seen from figure 5.31, the eigenfunctions and eigenvalues behave as solutions of the Laplacian eigenvalue problem.

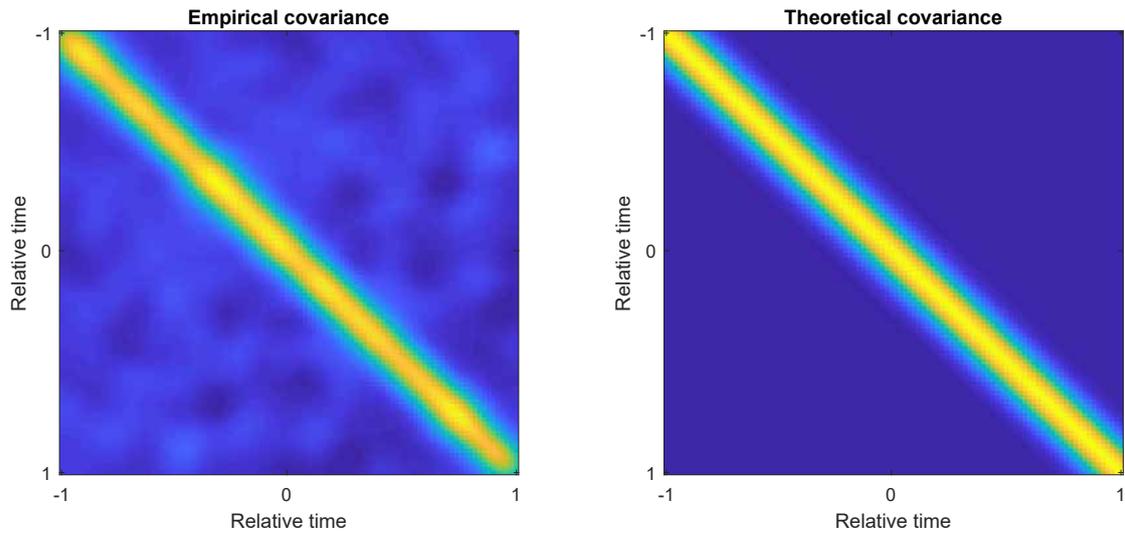


Figure 5.28: Empirical(left) and theoretical(right) covariance of smoothed noise

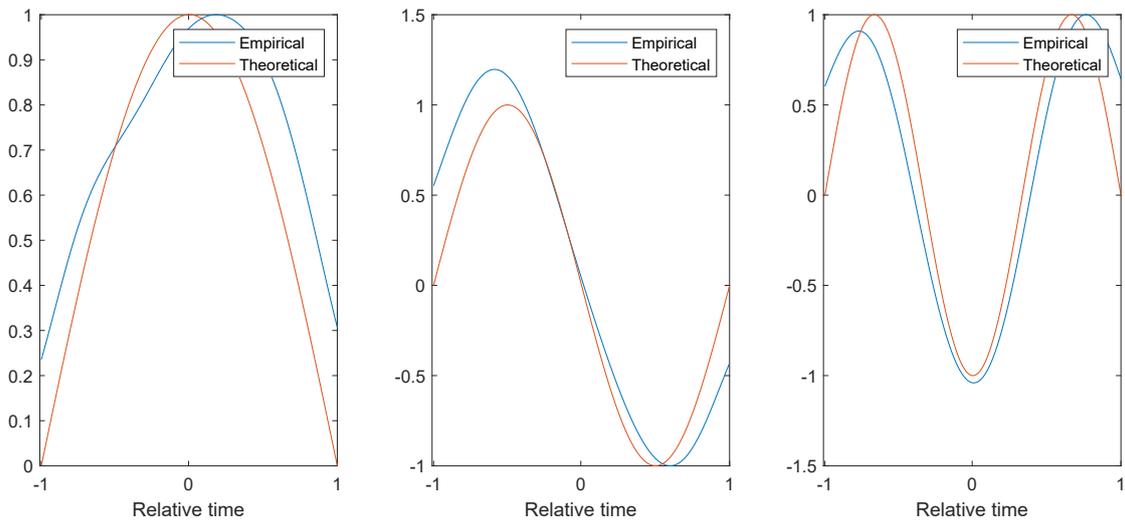


Figure 5.29: Empirical(left) and theoretical(right) eigenvectors. From left to right, first, second and third eigenvectors. The discrepancy is due to finite size effects.

## 5.5.2 Two models of autocorrelation

Now we want to obtain a more general result. We postulate two models of covariance matrix and for each one we present the process of relating it to the Laplacian operator.

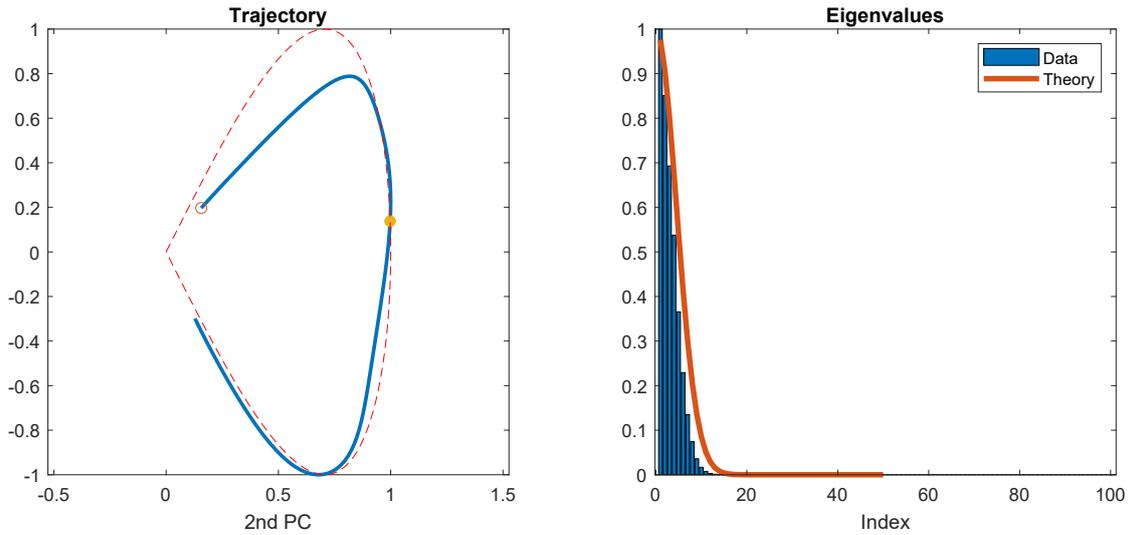


Figure 5.30: Trajectory (left) and eigenvalues (right). Both agreeing asymptotically with the theory

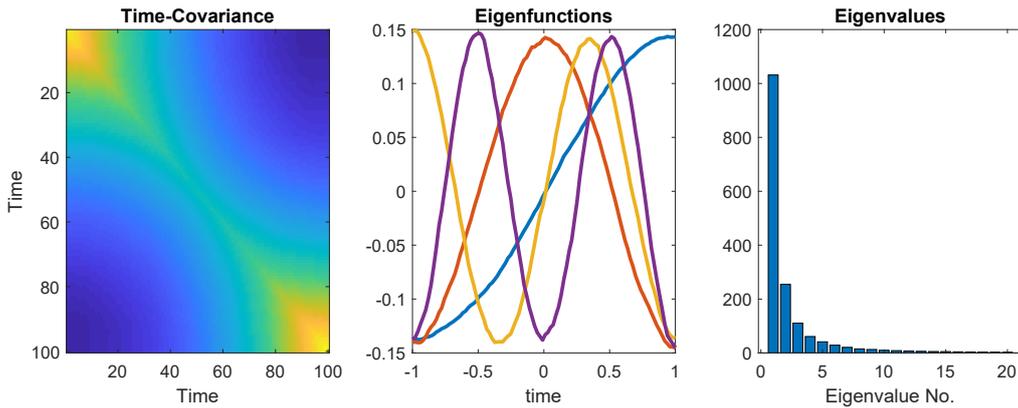


Figure 5.31: PCA of a random walk. Left. Covariance matrix of the random walk. Center. The eigenvectors are clearly Laplacian eigenfunctions. Right. The eigenvalues show exponential decay

### Gaussian autocorrelation

The case of a Gaussian autocorrelation is illustrated by a traveling bump in the population and it can be approached analytically building upon the tools developed in the previous section. Traveling bumps or traveling waves are a common occurrence in the brain as well as in models of the brain, so it is convenient that precise theoretical predictions can be obtained related to the speed of the bump.

Specifically, we are considering a data matrix with patterns similar to the one shown in the figure 5.32. Our goal is to describe the covariance matrix and its eigenvectors, to finally understand the neural trajectories obtained from this kind of data.

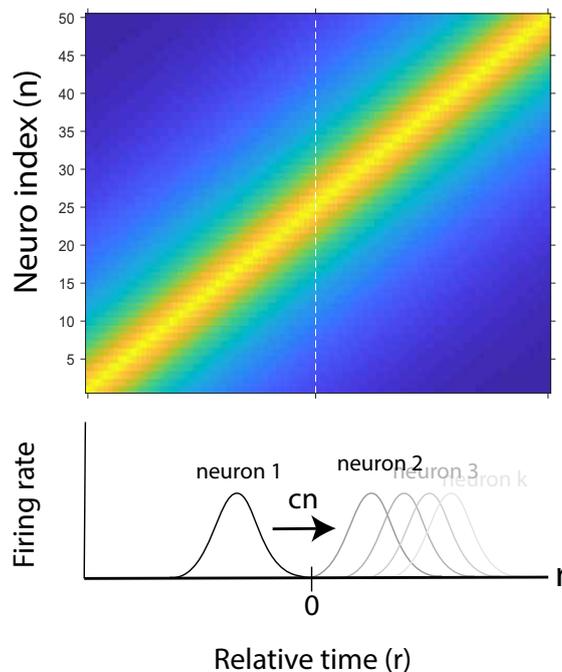


Figure 5.32: A traveling bump in the population

The traveling bump satisfies the following partial differential equation called the transport equation:

$$\partial_r u + c \partial_n u = 0 \quad (5.31)$$

$$u(r, 0) = g(r). \quad (5.32)$$

Remember that we treat our relative time ( $r$ ) as an spatial domain and the evolution in the equation treats the neuron index as a continuous time domain so the equation is interpreted as moving a bump ( $g(r)$ ) centered at  $r = -1$  for neuron  $n = 0$  to  $r = 1$  for neuron  $n = 50$  (in the example of figure 5.32) with speed  $c$ . The solution to this equation has the form

$$u(r, n) = g(r - cn) \quad (5.33)$$

The population covariance matrix has the form

$$C(n, m) = \int_{-\infty}^{\infty} g(r - cn)g(r - cm)dr, \quad (5.34)$$

which, as we show now, have the form  $C(n, m) = C(|n - m|)$  for a Gaussian initial condition. It is isotropic and depends only on the distance between neurons at positions  $n$  and  $m$ .

**Theorem 7.** *Let  $g(r) = (1/\sqrt{2\pi\sigma^2}) \exp(-r^2/(2\sigma^2))$ , then the population covariance matrix is*

$$C(n, m) = C(|n - m|) = \frac{1}{2\sqrt{\pi}\sigma} \exp(-c^2(n - m)^2/(4\sigma^2)) \quad (5.35)$$

*Proof.* First, making the change of variables  $r' \rightarrow r - cn$  and using the Gaussian in the equation 5.34, we get

$$C(n, m) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} \exp(-r'^2/(2\sigma^2)) \exp(-[r' + c(n - m)]^2/(2\sigma^2))dr \quad (5.36)$$

$$= \frac{1}{2\pi\sigma^2} \exp(-c^2(n - m)^2) \int_{-\infty}^{\infty} \exp(-r'^2/(2\sigma^2)) \exp(-cr'(n - m)/(2\sigma^2))dr, \quad (5.37)$$

which is integrated to

$$C(n, m) = \frac{1}{2\sqrt{\pi}\sigma} \exp(-c^2(n - m)^2/(4\sigma^2)). \quad (5.38)$$

This function, along with the empirical one from the data in figure 5.32, are shown in figure 5.33.  $\square$

In the same manner as in the previous section, this covariance matrix can be interpreted as the heat kernel

$$K(\tau, n, m) = \frac{1}{c} \frac{1}{4\pi\tau^2} \exp(-|n - m|^2/4\tau), \quad (5.39)$$

with  $\tau = \sigma^2/c^2$ . Also, as before, it can be decomposed as

$$K(\tau, n, m) = \frac{1}{c} \sum_k e^{-\lambda_k\tau} \phi_k(n)\phi_k(m), \quad (5.40)$$

where  $\{\lambda_k, \phi_k\}$  are eigenpairs of the Laplace-Beltrami operator. Specifically, for the appropriate initial conditions, they have the form

$$\lambda_k = \frac{k^2\pi^2}{L^2} \quad (5.41)$$

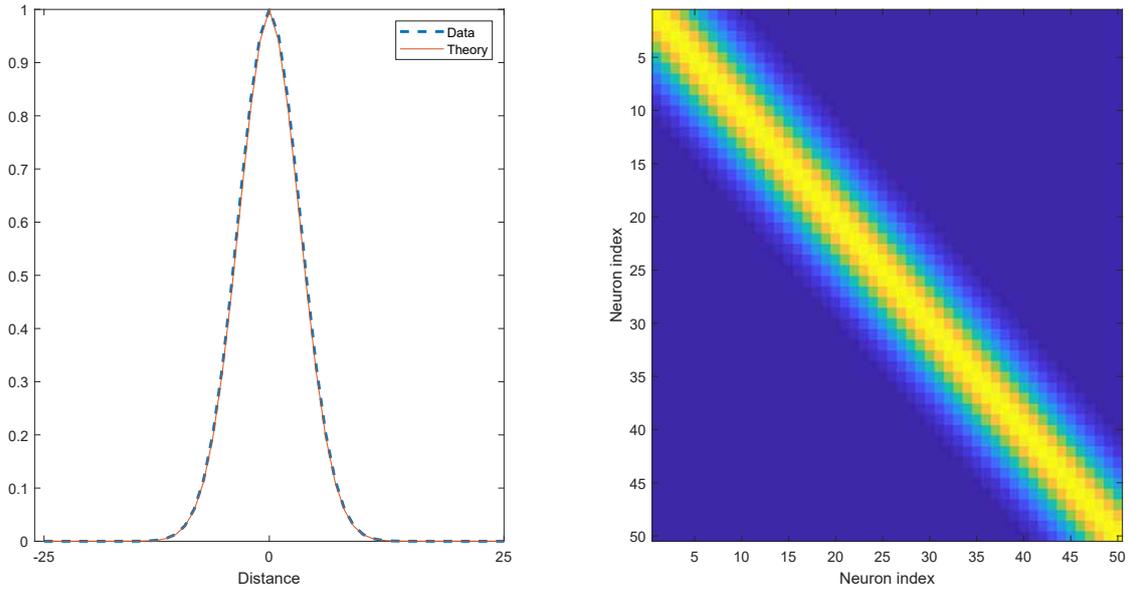


Figure 5.33: Covariance matrix of the traveling bump. Left is the profile of the covariance matrix as a function of the distance between neighboring neurons. Right. The diagonal form of the covariance matrix

and

$$\phi_k(n) = \frac{2}{L} \sin\left(\frac{k\pi n}{L}\right). \quad (5.42)$$

Note that the frequency of the eigenfunctions does not depend of the specifics of population. The corresponding theoretical and empirical eigenfunctions are shown in figure 5.34

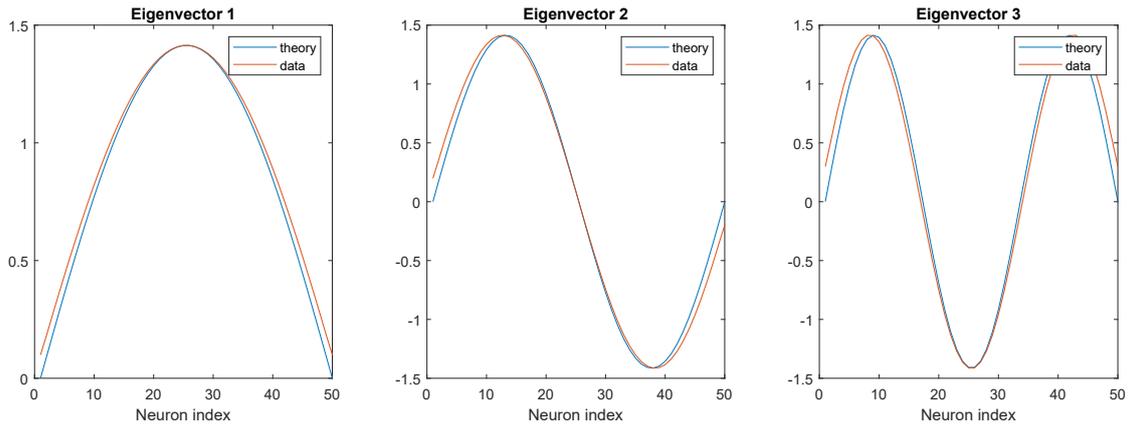


Figure 5.34: Eigenvectors of the covariance matrix

The eigenvalues are therefore

$$\lambda_k = e^{-\mu_k \sigma} \quad (5.43)$$

The trajectories, given by the projections

$$\gamma(r) = \int_{-\infty}^{\infty} \phi(n)u(n, r)dn \quad (5.44)$$

have a stereotypical shape reminiscent of the Lissajou curves (figure 5.35). Indeed, they are Lissajou curves for the appropriate relationship between the eigenfunction index and the size of the domain  $L$ .

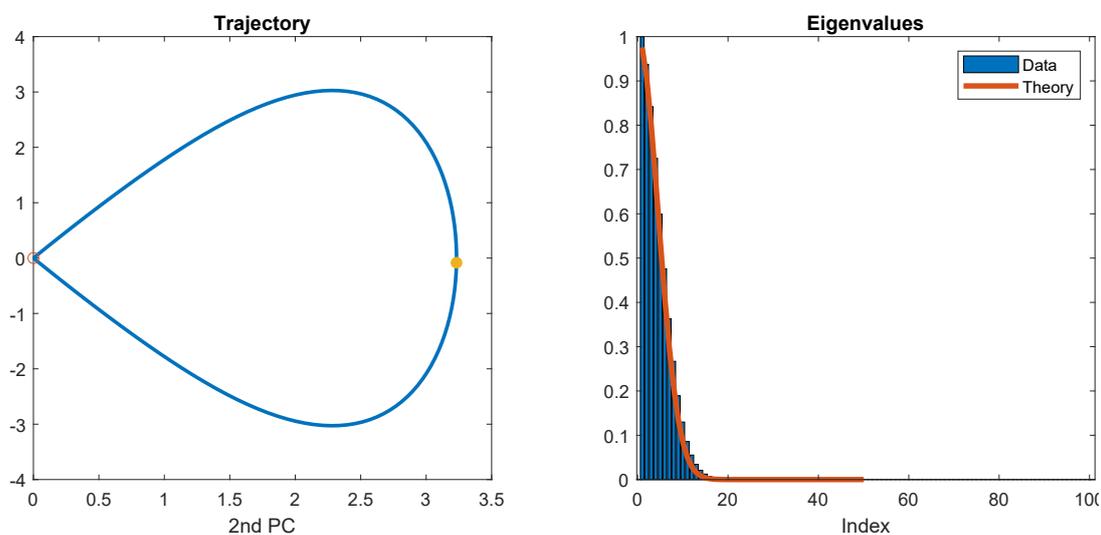


Figure 5.35: Left. Stereotypical shape of the trajectory. Right. The eigenvalues as given by the theory

### Exponentially decaying covariances

The previous section showed similar eigenfunctions for different kind of data. Why is that? What is common in those scenarios? The answer is: the time covariance matrix display short range correlations, therefore, another reasonable assumption about the time correlation functions is that it decays exponentially with the distance between two given points. More specifically, we assume a model time correlation matrix of the form

$$C(r, s) = C(|r - s|) = K \exp(-\sigma|r - s|), \quad (5.45)$$

then, the eigenvalue problem is

$$K \int_{-\infty}^{\infty} C(|r-s|)\phi_k(s)ds = \tilde{\lambda}_k \phi_k(r), \quad (5.46)$$

this can also be expressed more succinctly as

$$C * \phi_k = \lambda_k \phi_k, \quad (5.47)$$

where  $C * \phi$  represents the spatial convolution and  $\lambda_k = \tilde{\lambda}_k/K$ . Remember that we still treat the time domain as a spatial domain. This equation is also deeply connected with the Laplacian eigenvalue problem studied before but in a slightly different way; it is the Laplacian eigenvalue problem in disguise. In order to see this, let us apply the Fourier Transform (FT) on both sides

$$\widehat{C * \phi_k} = \widehat{C} \widehat{\phi_k} = \lambda_k \widehat{\phi_k}, \quad (5.48)$$

where we used the properties of the FT of the convolution. The FT of the covariance function in (5.45) is known to be  $\hat{C}(\xi) = 2\sigma/(\xi^2 + \sigma^2)$ , therefore, we can express equation (5.48) as

$$\xi^2 \hat{\phi} + \sigma^2 \hat{\phi} = \frac{2\sigma}{\lambda_k} \hat{\phi}. \quad (5.49)$$

We can now apply the inverse FT to get the following differential equation

$$-\frac{d^2}{dr^2} \phi_k(r) = \left( \frac{\sigma}{\lambda_k} - \sigma^2 \right) \phi_k(r) = \mu_k \phi_k(r), \quad (5.50)$$

which is the one dimensional eigenvalue problem for the Laplacian operator. It is known (see section 2) that on a bounded domain of length  $L$  (for example in an interval around a given event in our case), the spectrum is discrete and the eigenfunctions are given, again, by

$$\phi_k(x) = \begin{cases} B \cos(\sqrt{\mu_k}x), & k = 1, 3, 5, \dots \\ A \sin(\sqrt{\mu_k}x), & k = 2, 4, 6, \dots \end{cases} \quad (5.51)$$

and

$$\mu_k = \frac{k^2 \pi^2}{L^2}. \quad (5.52)$$

The eigenvalues of the integral operator can therefore be obtained from

$$\lambda_k = \frac{2\sigma}{\mu_k + \sigma^2}. \quad (5.53)$$

In figure 5.36 we show the eigenvectors for a simulated covariance matrix with the proposed exponential structure. As it can be seen, they correspond with the theoretical results.

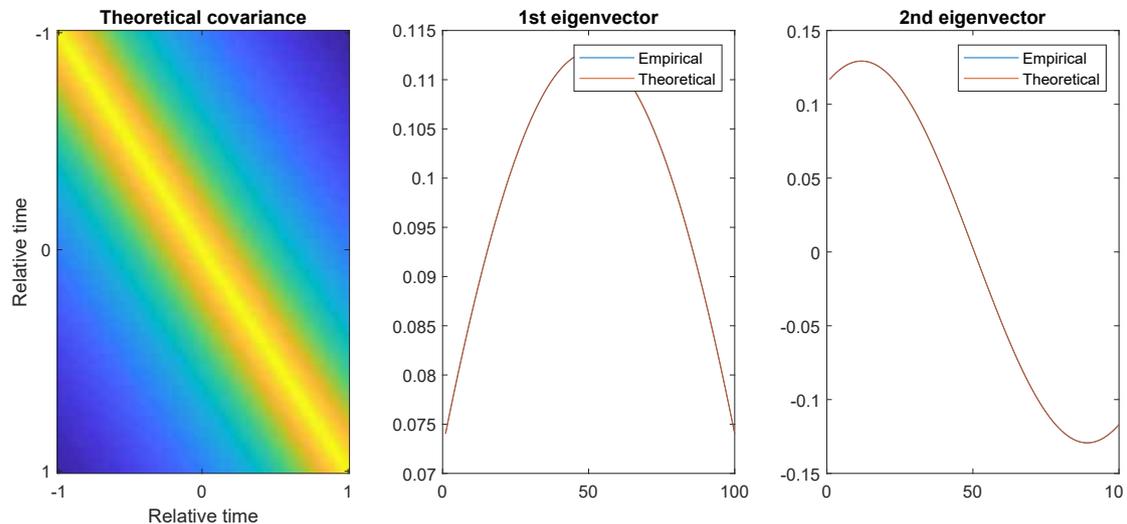


Figure 5.36: Eigenvectors of an exponentially decaying covariance matrix

The principal components, properly scaled, are therefore

$$\gamma_k(r) = \sqrt{\lambda_k} \phi_k(r), \quad (5.54)$$

which, along with the spectrum (eigenvalues) are well explained by the theory (figure 5.37). The trajectory again has the stereotypical shape seen before.

### 5.5.3 The basal ganglia trajectories behave like laplacian eigenfunctions

Firing rates will surely show the phenomenon just described as they are, after all, functions of time. One way to check to which degree they behave like the null functions presented before is by observing whether their principal components satisfy the Laplacian eigenvalue problem, that is, whether their second derivative is a multiple of themselves.

Figures 5.38, 5.39 and 5.40, show the degree of matching between the first PCs and their derivatives. The agreement is higher for later PCs (figure 5.40 for comparison). The striatum and striatal subpopulations (MSNs and FSIs) behave more like null shapes for all PCs. By the 3rd PC, the equation is satisfied almost everywhere except, and the same is true for later eigenvectors, in the initial segment that precedes movement in which the correlations are destroyed or attenuated.

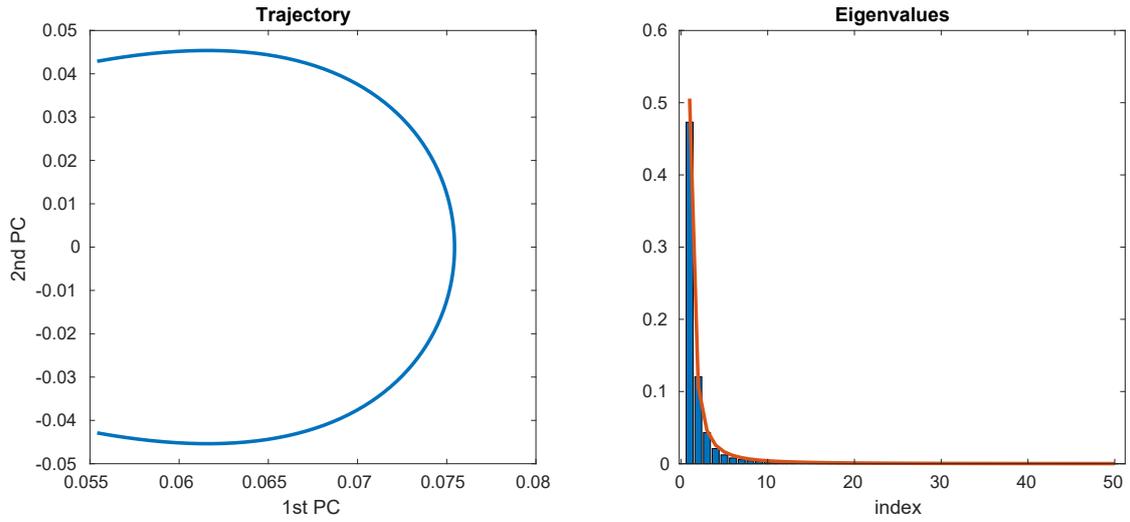


Figure 5.37: Trajectory and eigenvalues

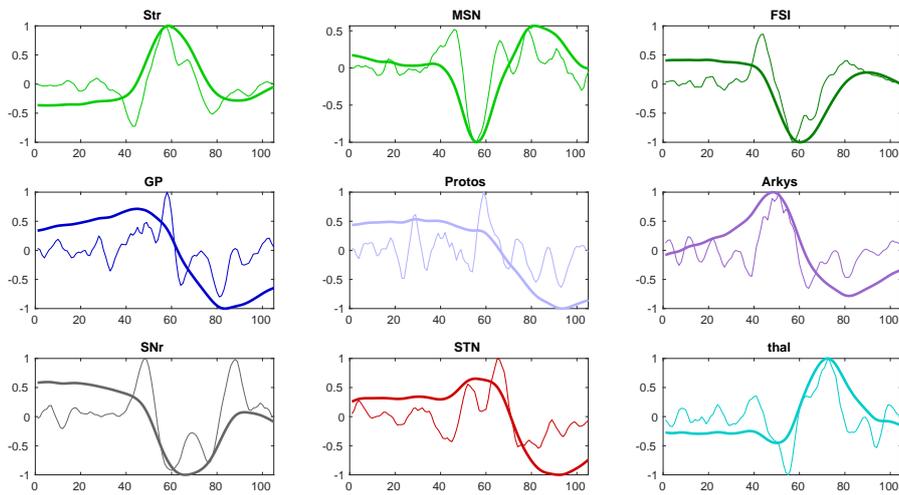


Figure 5.38: By plotting the each principal component along with their second derivative, the degree of satisfaction of the Laplacian eigenvalue problem can be assessed. Each panel correspond to the first PC for each trajectory.

This observation suggests two methods to develop null spectra for each of the regions. The first one is to fit one of the two model covariance matrices presented in the previous sections to the data and compute the eigenvalues that correspond to each of the theoretical eigenfunctions by using equations 5.43 or 5.53. The second one is to fit the second derivative to one of the PCs times the eigenvalue

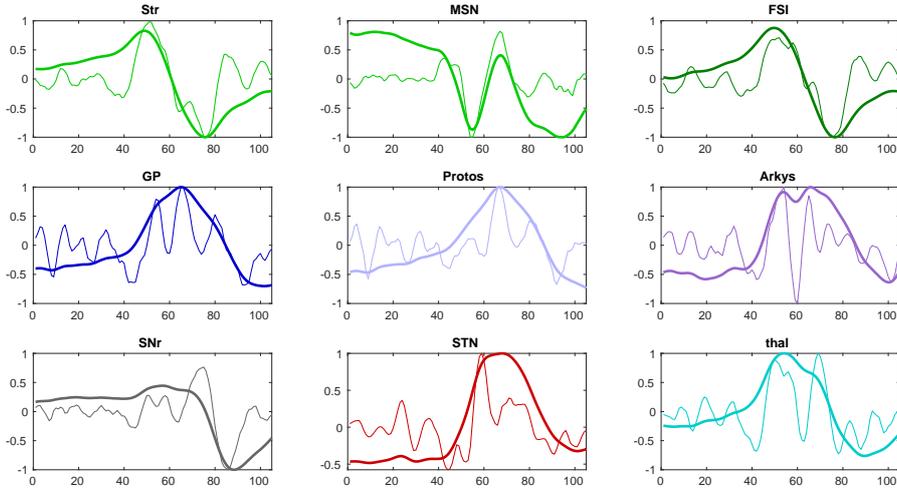


Figure 5.39: Laplacian eigenvalue satisfaction PC2

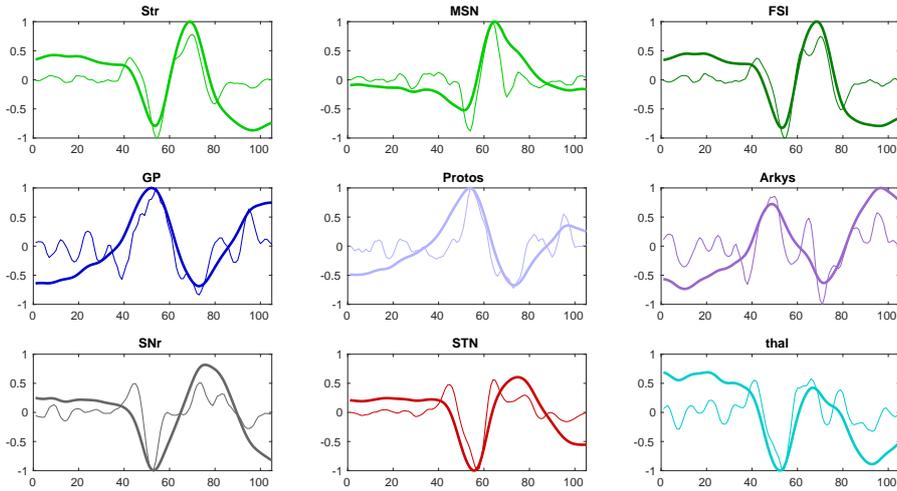


Figure 5.40: Laplacian eigenvalue satisfaction PC3

and then use that information to derive the rest of the relevant information.

We summarize the theoretical formulas obtained in this section:

- Eigenfunctions: for measurement interval of length  $L$  the eigenfunctions are:

$$\phi_k(x) = \begin{cases} B \cos(\sqrt{\mu_k}t), & k = 1, 3, 5, \dots \\ A \sin(\sqrt{\mu_k}t), & k = 2, 4, 6, \dots \end{cases} \quad (5.55)$$

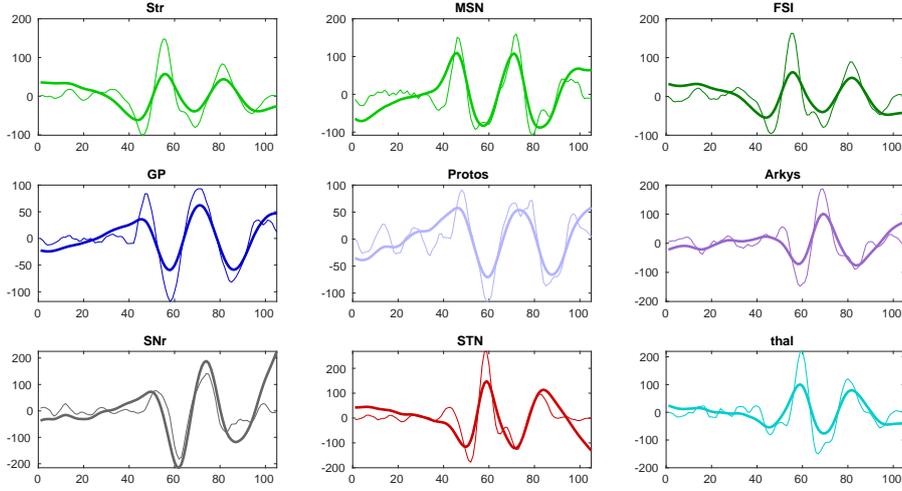


Figure 5.41: Laplacian eigenvalue satisfaction PC5. Later PCs have greater degree of matching. They behave more as null shapes

- The  $k$ th Laplacian eigenvalue  $\mu_k$  is:

$$\mu_k = \frac{k^2 \pi^2}{L^2} \quad (5.56)$$

- If the null covariance model is assumed to be exponential (circus tent function), and  $\lambda_k$  is the Laplacian eigenvalue, then the eigenvalues of the PCA are

$$\lambda_k = \frac{2\sigma}{\mu_k + \sigma^2}. \quad (5.57)$$

- If the null covariance model is assumed to be Gaussian, and  $\mu_k$  is the Laplacian eigenvalue, then the eigenvalues of the PCA are:

$$\lambda_k = e^{-\mu_k \sigma} \quad (5.58)$$

The theoretical fits for the different brain regions are shown in figure 5.42.

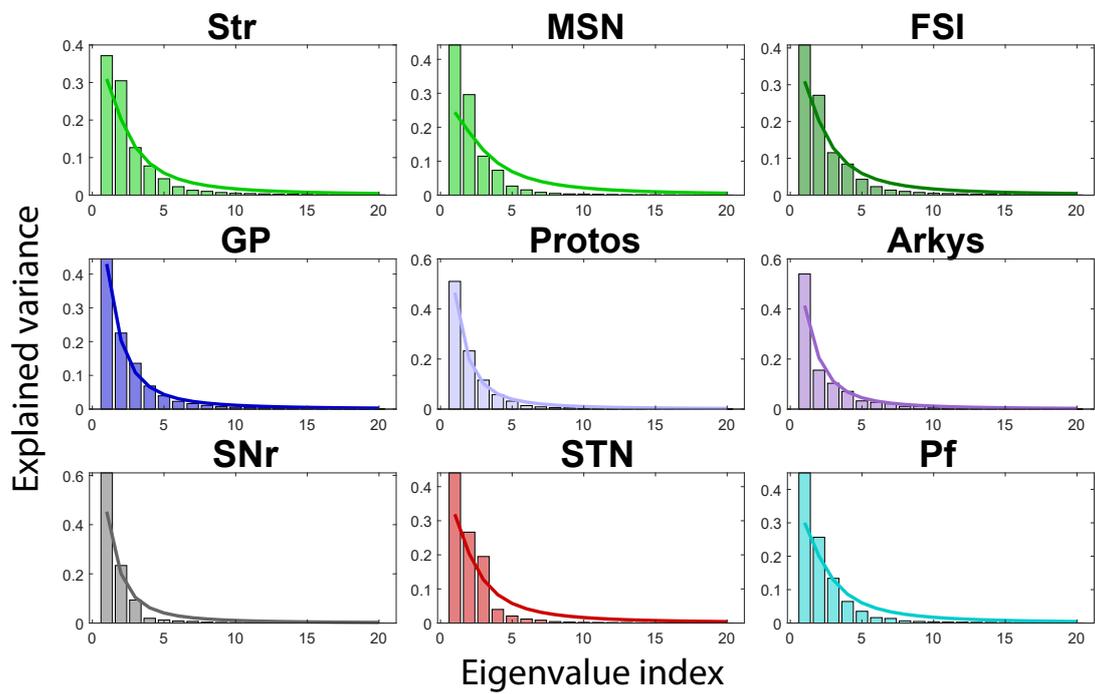


Figure 5.42: Empirical (bars) and theoretical (lines) spectra.

## Chapter 6

# Population interpretation and neural ensembles

## 6.1 Chapter overview

In this chapter we argue that the best approach to understand the structure of the principal components and, by extension, that of the low dimensional trajectories, is hierarchical. Such a hierarchy, in general, is not equivalent to the one a traditional clustering algorithm might find; instead we show that it is a very specific one that has important mathematical properties.

The most important result of this section is a *bijection* between the principal components and combinatorial entities called “Catalan objects”. Not only we show that there is a one to one correspondence between the eigensystem of an Hermitian matrix and those objects, but also that such relation has physiological significance in terms of neural ensembles or assemblies. To the knowledge of the authors, these results are new in the literature.

The importance of such bijections lies in the fact that now the principal components are associated to more than 200 mathematical objects and concepts amply described in the literature [87]. This opens up a new way of studying and visualizing population codes. Indeed, the mathematical properties of those objects are immediately transferred to the principal components by means of the bijection, providing a change in perspective that can provide insights about the data that will be hidden otherwise.

From a more practical standpoint, we use this methodology to attain the following goals:

1. Identify the groups or ensembles that drive the trajectory in different directions in the phase space
2. Quantify the contributions of each group to the each eigenvalue
3. Identify orthogonal or overlapping subspaces for the evolution of the population trajectory

We start by showing how the neural ensemble view is important and instrumental in understanding the curvature of the neural trajectories. We proceed then to present the theoretical grounds for the hierarchical approach to the principal components. In section 6.3 we show a recursive interpretation of PCA thanks to an important theorem in linear algebra. Then, in sections 6.4 and 6.4.1 we present the main motivations and concepts behind the PCA ensembles; this is just a geometrical reading of the sign changes of the eigenvectors of the covariance matrix.

In section 6.5 we finally present our proof of the bijection. This is done in a constructive way, i.e. by providing a specific set of rules that yield a unique Catalan object from a particular system of eigenvectors and the other way around. The

chosen Catalan entity is a semiorder. This has the advantage of endowing the population with a natural ordering structure in terms of explained variance and overlaps. The discovery of this relation is one of the main contributions of this work.

We end the chapter by relating the semiorder to the way the low dimensional subspace is arranged in the high dimensional one and addressing the issue of mean subtraction. We argue that mean subtraction adds a new dimension to the analysis but it has to be dealt with carefully because it also complicates the interpretation of the components of the eigenvectors.

## 6.2 Neural ensembles are important for the understanding of curvature

As a first step in understanding the relation between the shape and the population activity, remember that the curvature of a high dimensional arc-length parametrized curve  $\gamma(s)$  is the magnitude of the acceleration, that is,

$$k(s) = \|\gamma''(s)\| = \sqrt{(\gamma_1''(s))^2 + (\gamma_2''(s))^2 + \dots + (\gamma_n''(s))^2}. \quad (6.1)$$

Additionally, for a  $n$  dimensional curve, there are  $n - 1$  generalized curvatures. Two observations are in order.

Firstly, the results of [3] suggest that, for a very small interval around the event of interest (i.e. movement initiation), the principal subspace is very similar to Frenet frame of the curve at the time of movement. Therefore, non-zero higher curvatures suggest higher dimensionality of the curve itself.

Secondly, the presence of ensembles implies that the curvature could be approximately explained by the local pattern of firing of the ensemble. By local we mean that they have effectively the same curvature at the point  $s$  but can differ later. To see this, let us assume that there are  $k \ll n$  patterns in the individual neurons, say  $\tilde{\gamma}_1(s), \tilde{\gamma}_2(s), \dots, \tilde{\gamma}_k(s)$ , and  $G_i$  neurons displaying the pattern  $i$ ,  $i = 1, 2, \dots, k$ , then

$$k(s) \approx \sqrt{(G_1 \hat{\gamma}_1''(s))^2 + (G_2 \hat{\gamma}_2''(s))^2 + \dots + (G_k \hat{\gamma}_k''(s))^2}. \quad (6.2)$$

Moreover, based upon the observations in the previous chapter and as a preliminary study of the data, we can group the neurons in those with high and small curvature, again, locally. Three cases will be of interest. Either all the curvatures are equal and constant, which gives a circle (a vertex); either they are zero (all patterns are changing slowly ( a flat region)); or there are two subgroups, one slow and one fast (a sharp point).

Even so, as we will see, the trajectories are not two dimensional in general and they will have significant torsion or higher curvatures at times; in those times, an additional dimension is necessary to fully explain the data (irregular points).

## 6.3 Preliminary insights in the hierarchical nature of PCA

The previous observations justify study of the ensembles contributing to the curvature in the PCA. We will see that the similitude in firing is encoded in the components of the eigenvectors of the covariance matrix, however, before studying them we should recognize the hierarchical structure of PCA. The starting point is the Courant-Fisher theorem [59] which provides a recursive definition of the

eigenvalues of an Hermitian matrix, which the covariance matrix happens to be. We just state the min-max, real version of it:

**Theorem 8** (Courant-Fisher). *The eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  of a hermitian matrix  $\mathbf{C}$  are*

$$\lambda_i = \min_{\dim \mathcal{X} = n-i+1} \max_{x \in \mathcal{X}, \|x\|=1} x^T \mathbf{C} x. \quad (6.3)$$

When  $i = 1$ ,  $\mathcal{X} = \mathbb{R}^n$ , that is, the full euclidean space.

This seemingly complicated statement implies the following: Eigenvectors of the covariance matrix can be computed recursively as follows (figure 6.1)

- If the matrix is a block matrix, separate the different blocks.
- For each block, compute the largest eigenvalue
- Project the data into the  $n - 1$  dimensional subspace (hyperplane) orthogonal to the corresponding eigenvector
- Repeat the procedure for the projected data.

Therefore, in each iteration you compute the eigenvectors for the projected data which has lost one dimension. We obtain this way a sequence of subspaces each of which has some percentage of the explained variance left on it.

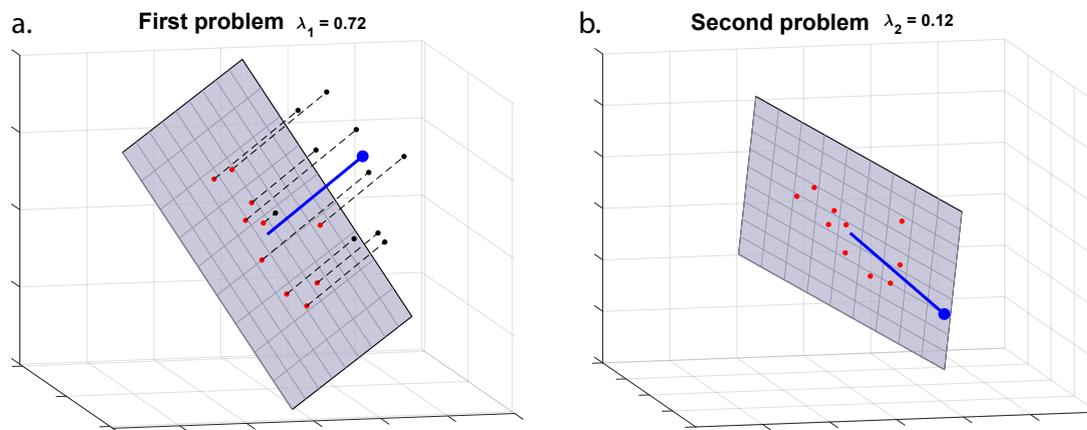


Figure 6.1: a. The original data in black with  $n = 3$ , the first eigenvector is shown in blue, the orthogonal subspace is shown in light blue with the data projected on it. b. The first eigenvector of the projected data is the second eigenvector of the original data.

The trees that we will build in the forthcoming sections contain the structure of such recursion but, in order to find an algorithm to construct them we need

to analyze the structure of the eigenvectors. In the next section we, additionally, relate this to the ensemble structure of the population. In order to do so, it is convenient to assume that the mean is not subtracted from the data.

Observe that, if the data is not subtracted, the first eigenvector will pass through the centroid of the data [70] and the projection onto the first hyperplane in the previous procedure will be equal to the mean subtraction. Not subtracting the mean, however allows to single out cases in which the matrix has a block structure (completely independent patterns) which is important for the proof below.

## 6.4 The induced system of hyperplanes

The definition of PCA in terms of eigenvectors or directions of maximal variance implies a dual view of a best fitting system of hyperplanes; the traditional one is in terms of eigenvectors (i.e. one dimensional subspaces), the dual one is in terms of  $n - 1$  dimensional subspaces. This view was indeed present in the first major works about the topic, in Pearson's 1901 paper [70] and in Hotelling's comprehensive work [41]. In this section we use this fact to study the ensemble structure of the population code. We start by clarifying in which sense we mean that a system of hyperplanes is *best fitting* to the data.

Given a set of basis vectors  $\mathcal{B} = \{v_1, v_2, \dots, v_n\}$ , the *induced system of hyperplanes* is defined by the equations  $v_i^T x = 0$ ,  $i = 1, 2, \dots, n$ ; it is *central* and *essential* [2] (we refer the reader for the preliminaries section for the fundamentals about hyperplanes). A good way of measuring how fit a point  $x \in \mathbb{R}^n$  is to basis  $\mathcal{B}$  and, by extension, to the induced system of hyperplanes is by the following distance function introduced in [22].

$$\epsilon^2(x) = - \sum_i z_i^2 \log z_i^2, \quad (6.4)$$

where  $z_i = v_i^T x$  are the coordinates of the point in the basis  $\mathcal{B}$ . In [22] it is considered the "*distance*" between the point and the basis. This function reaches a maximum when the vector has a  $45^\circ$  angle with all the hyperplanes and that is the farthest it can be from all of them at once. It turns out that PCA minimizes the average distance of all the data points from the given hyperplane arrangement.

**Theorem 9.** *Give an set of points  $X = \{x^1, x^2, \dots, x^m\}$ , the basis given by the principal directions minimizes the average distance given by*

$$f(X) = \frac{1}{m} \sum_k \epsilon^2(x^k) = - \frac{1}{m} \sum_k \sum_i |z_i^k|^2 \log |z_i^k|^2, \quad (6.5)$$

with  $z_i^k$  given as before.

*Proof.* The proof follows from *Jensen's inequality*. The function  $\epsilon^2(x)$  is concave and therefore it satisfies  $\epsilon^2(\sum_i x_i) \geq \sum_i \epsilon^2(x_i)$  or

$$-\frac{1}{m} \sum_k \sum_i |z_i^k|^2 \log |z_i^k|^2 \leq -\frac{1}{m} \sum_k \left( \sum_i |z_i^k|^2 \right) \log \left( \sum_i |z_i^k|^2 \right),$$

observe that  $\sum_i |z_i^k|^2 = \|z^k\|^2$ . We can replace then the original problem with the equivalent one of minimizing the right hand side of the previous expression, so the new optimization problem can be stated as:

$$\min_{V \in SO(n)} \hat{f}(X) = - \sum_k \|z^k\|^2 \log \|z^k\|^2, \quad (6.6)$$

which by the *minimal average length property* of the PCA (chapter 3), is reached when

$$\hat{f}(X) = - \sum_k \lambda_k \log \lambda_k \quad (6.7)$$

[88], where  $\lambda_k$  is the  $k$ -th eigenvalue of the covariance matrix of  $X$  and  $V$  is the matrix of eigenvectors.  $\square$

This best fitting hyperplane system can be seen as sequence of linear decoders that partition the set of neurons (coordinate axes) into different classes, the neural (geometrical) ensembles. We will come back to that interpretation later but first we will develop some tools to understand its high dimensional structure. This is instrumental in understanding the journey of the trajectory in the high dimensional space; the parts of the hyperplane arrangement that it visits will correspond to the neural ensembles.

First we find out where in this hyperplane arrangement lie each neuron (i.e. coordinate), then we study the locations of groups of neurons. Remember that a hyperplane  $H_i$  divides the space into two half spaces  $\{H_i^+, H_i^-\}$ , we follow the convention as in [2] and assume that half spaces are closed.

Given a system of hyperplanes  $\mathcal{H}$ , a *face* is the intersection of the half spaces taking either 1 or the 2 half spaces for each hyperplane. Maximal faces are called *chambers*. Each face  $F$  has an opposite face  $\bar{F}$ , the set  $\{F, \bar{F}\}$  is called a projective face. We select a set of positive faces as the set of faces that are in  $H_1^+$ , that is, in the positive half space associated to the first eigenvector. The hyperplanes that border a chamber are called *walls*.

In the neural space we associate each canonical basis vector of the euclidean space to a neuron (i.e.  $\text{neuron}_i = e_i$ ), that is, each coordinate axis is a neuron. We have the following important observation.

**Lemma 3.** *Each maximal positive chamber of the induced system of hyperplanes is occupied by a single neuron (figure 6.2)*

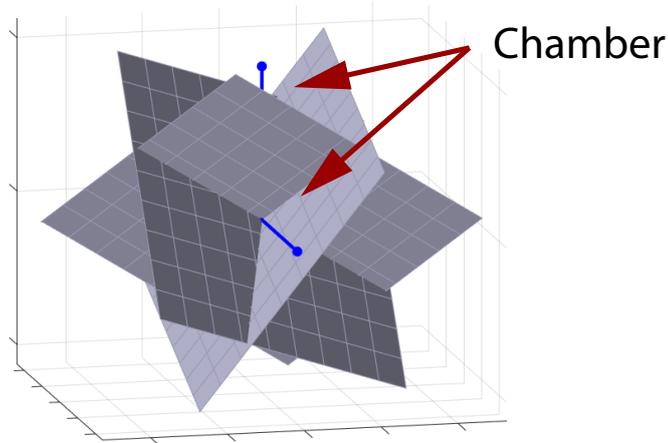


Figure 6.2: Induced hyperplane arrangement and neuron locations. The shaded hyperplanes are the induced hyperplane arrangement. The location of the neurons is shown in red. Blue lines represent single neurons.

*Proof.* Let us assume that two canonical basis vectors belong to the same maximal chamber. If they are not in any of the walls, their angle with each of the walls should be acute and therefore the angle between the two of them should be acute too but that is a contradiction of the fact that they are elements of the canonical basis.  $\square$

The relation between neurons and chambers suggest an encoding of the population activity. Associated with each chamber there is a sign sequence that uniquely identifies it. This sign sequence reflects the different half spaces intersected that result on that chamber. We have therefore, the definition

**Definition 17** (Chamber code). *A chamber code for a neuron is a sequence of  $+s$  and  $-s$  that uniquely identify the chamber that neuron lies on.*

We will use this to locate the ensembles participating in a given point of interest of the curve.

#### 6.4.1 A change in perspective: neural ensembles as faces of the hyperplane arrangement

Last showed that single neurons occupy chambers or maximal faces of the induced hyperplane arrangement. If single axis represent a single neuron then the subspace

(plane) spanned by two canonical vectors represent the joint firing of two neurons, etc. Those subspaces are faces of the *canonical coordinate arrangement*. It differs from the induced hyperplane arrangement in that it is generated by the canonical vectors and not the eigenvectors of the covariance matrix.

**Definition 18** (Geometrical ensemble). *Given  $k$  indexes  $i_1, i_2, \dots, i_k$ ,  $k \in [n]$ , a geometrical ensemble is the cone  $p = \alpha_1 e_{i_1} + \alpha_2 e_{i_2} + \dots + \alpha_k e_{i_k}$  with all  $\alpha_j \geq 0$ , that is, the linear combinations of the corresponding canonical vectors with positive coefficients.*

Each geometrical ensemble is a face of the canonical hyperplane arrangement. It represents a group of neurons or, in terms of the newly developed concept of combinatorial codes[25], a codeword. Additionally, belonging to a geometrical ensemble is an equivalence relation. Examples of geometrical ensembles are the first octant in the 3 dimensional euclidean space, the plane  $x = 0$ , the  $x$ -axis.

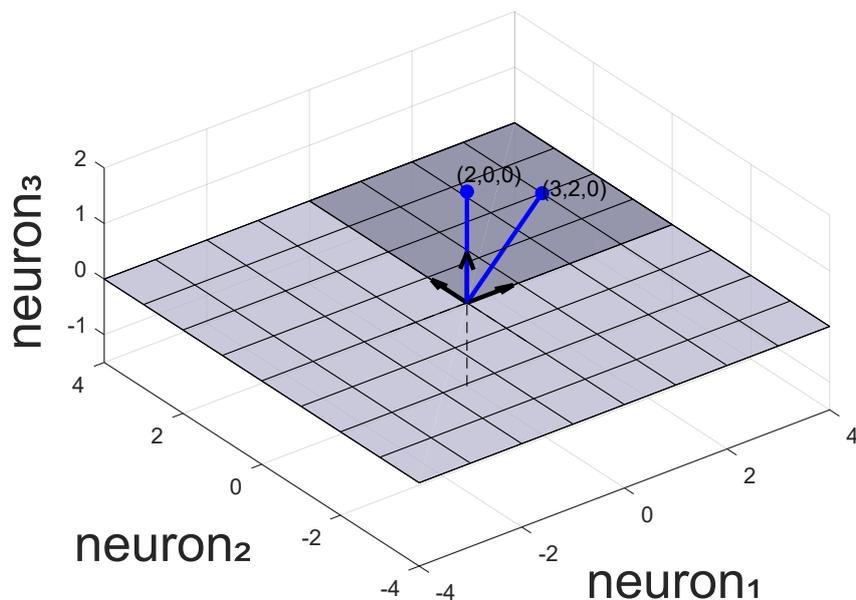


Figure 6.3: Example of a geometric ensemble. The point  $(3, 2, 0)$  represents the situation of neuron 1 having a firing rate of 3spikes/sec and the neuron 2, 2spikes/sec. The shaded area is the face corresponding to the geometric ensemble. Any point in that area corresponds to the coactivation of neurons 1 and 2 and not the neuron 3. The point  $(0, 0, 2)$  and the hole ray parallel to the  $z$ -axis, corresponds, similarly, to the activation of only one neuron: neuron 3.

Consider a point in the interior of a face. In terms of firing rates, it corresponds

to the situation in which the corresponding neurons are firing together. For example, the point  $p = (3, 2, 0)$  in figure 6.4 means that neuron 1 has firing rate 3, neuron 2 has firing rate 2 and neuron 3 is not firing at all, hence, it represents the ensemble 12. That ensemble is also represented by any positive multiple of  $p$  or any point with nonzero values in the first and second component and zero in the third.

Note that each hyperplane of the induced hyperplane arrangement separates two geometrical ensembles by itself, one on each half space. Thinking about each hyperplane as a linear classifier, it divides the set of canonical basis vectors in to two groups. It is important to note here the change of perspective; instead of separating points, as in traditional classification problems, we separate axes or, more specifically, coordinate rays of the high dimensional space.

The traditional view of separating points is not very informative because those points, in our case, are part of a curve and, consequently, no ad-hoc classes can be assigned to them individually. On the other hand, the fact that a subset of coordinate rays is separated from another indicates a partition of the population into two ensembles; therefore, the trajectory on each side is said to be *visiting* the corresponding ensembles.

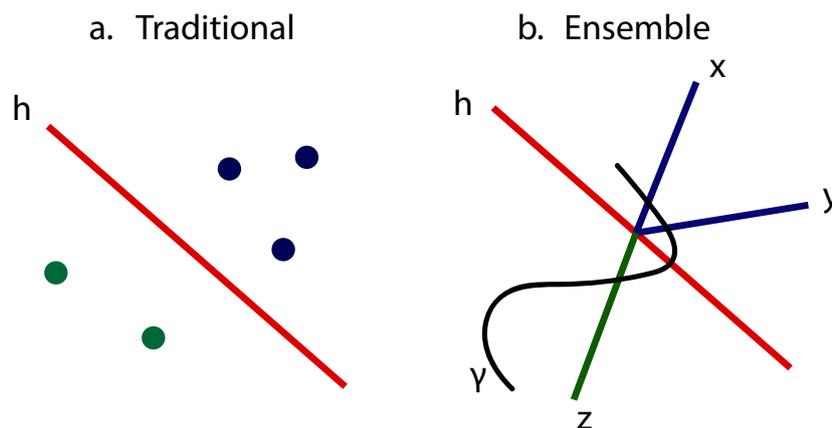


Figure 6.4: Example of the change of perspective in relation to the view of the induced hyperplane arrangement as a set of linear classifiers. a. In the traditional classification problem, hyperplane  $h$  separate the classes of points, green and blue. b. In the geometric ensemble perspective, the hyperlane separates sets of coordinates. Here,  $xy$  is an ensemble that has been selected out by  $h$ . In black, the trajectory visits the corresponding ensembles ( $xy$  and  $z$ )

As a final fact about the geometrical ensembles, it is important to note that the set of geometrical ensembles, as the set of faces has a partial order (see chapter 3) given by inclusion. So, we say that the ensemble  $S_1 \leq S_2$  iff  $S_1 \subset S_2$ , that is, if

$S_1$  is contained in  $S_2$ . For example, for ensembles 12 and 123, we have  $12 \leq 123$ . If we restrict ourselves to the positive orthant of the high dimensional space, we say that the partial order is a *lattice* (has a maximum and minimum element). The case of the positive orthant is the one in which we work with firing rates which are, by definition, positive.

## 6.5 Catalan objects: overlap trees and the neural semiorder

As seen in the previous section, interpreting the classification in terms of the ensembles allows us to study the dimensionality reduction aspects of PCA from a new perspective: By studying the ensembles that are singled out by adding each hyperplane of the induced hyperplane arrangement one at a time. This process can be encoded in 3 related tree structures.

### 6.5.1 The difference tree

The difference tree, associated to a hyperplane arrangement  $\mathcal{A}$  is defined recursively by the algorithm 1 which receives the eigenvectors  $v_i, i = 1, 2, \dots, n$  as an input. The iteration is performed for all eigenvector of until certain  $k$  that obeys a specified threshold (like a minimum amount of explained variance).

---

#### Algorithm 1 Difference tree

---

```

1: belonging  $\leftarrow$  node0 for all neurons
2: nodes  $\leftarrow$  node0
3: for  $i = 0$  to  $k \leq n$  do
4:   for all nodes do
5:      $u \leftarrow$  neurons in current node
6:      $h^+ \leftarrow j$  s.t.  $v_{i,j} > 0$  for  $j \in u$ 
7:     belonging[ $h^+$ ] = new node
8:     add new node to current node as a child

```

---

During the algorithm, each node contains a subset of the indexes corresponding to the neurons in the data; each neuron can only belong to a unique node, this restriction is enforced by having the relationship encoded in the variable *belonging*. At the beginning the tree has only one node which contains all the indexes. The neurons with positive coefficients in the  $i$ th eigenvectors are assigned to a new node that is then attached as a (right) child of the current node. The ambiguity of sign is resolved by assigning the choosing the half space with higher *compound contribution*.

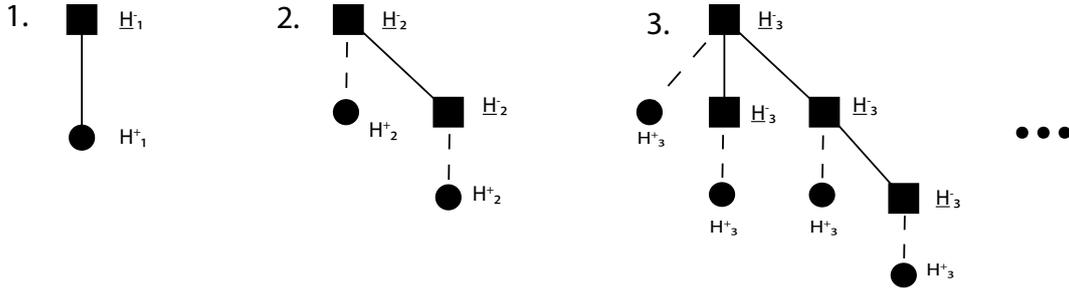


Figure 6.5: Growing of the difference tree assuming a population that at each step has neurons on both sides of all the hyperplanes

**Definition 19** (Compound contribution). *The Compound contribution of a set of neurons (indexes)  $I$  to a specific eigenvector  $u$  is given by*

$$c = \sum_{i \in I} |u_i|^2 \quad (6.8)$$

From the hyperplane perspective, each step encodes which geometrical ensembles occupy each chamber in a sequence of *sub-arrangements* of the original hyperplane arrangement. For example, in the first step of the first for loop, the hyperplane arrangement consists of a single hyperplane, the one corresponding to the first eigenvector. For data with positive firing rates, all the components of the eigenvector will be either zero or positive, therefore, the first step creates a new node with all the neurons, again. Geometrically, this means that the first eigenvector will be in the positive octant and the orthogonal plane will have all the coordinate rays, the whole octant, on the same side.

The next hyperplane will divide this group into 2 subgroups. Those in the positive half space will explain more variance as their dimension is higher. The next step, for the 3rd eigenvector will create 4 groups, two children; one child for every node whose neurons happen to be on the positive half space generated by this hyperplane. Geometrically, this encodes the way in which that hyperplane cuts the corresponding face of the previous subarrangement. The process of growing at each step is shown in figure 6.5.

Because being part of an ensemble is an equivalence relation, we can think of each node as representing a single, ideal neuron with a specific pattern in the firing rate. The relationship between the ensembles in the tree turns out to be very informative; it imposes a special kind of order in the ensembles: A semiorder. The basic properties of semiorders are presented in the chapter 3.

**Theorem 10** (The ensemble semiorder). *The difference tree define a semiorder or a decision pattern on the set of neurons.*

*Proof.* We have to show that there is a bijection between the set of semiorders and that of the trees generated by this procedure. Using the common representation of semiorders [98], the bijection, shown in figure 6.6 for all possible trees with  $n = 4$  neurons, is done as follows:

tree  $\implies$  semiorders:

1. Starting from the leftmost child, write the nodes as labeled until you reach a branching point.
2. If the branching point is the root, start a new, disjoint line and repeat the algorithm for the remaining nodes
3. Apply the algorithm recursively to the next child the branch and extend the initial line to overlap as many nodes as the leftmost branch of this sub tree has.
4. Repeat this process for all the branches to the right.
5. Write the common ancestor last, along with the ancestors immediately on top that are not branching points and connect them with a new line that overlaps as many nodes as children of the leftmost branch in the right most children of the branching point.

semiorders  $\implies$  trees:

- Draw a root node
- Draw as many children as non-overlapping lines there are
- For each set of overlapping set of lines draw a new child with as many branches as overlapping lines there are
- For each set of points completely covered by a line, create a chain of nodes.

This procedure creates a unique tree for each semiorder (“into”) and every tree will be created by a given semiorder (“onto”).  $\square$

The significance of the semiorder is thus that it *unfolds the dependencies* between the non-independent principal components. It provides also a natural clustering of the population into ensembles that fire together at some point of the measurement interval; we call them *life time ensembles*.

One of the interesting consequences of this bijection is that the principal components are blind to *Crossing relationships*, that is, a relationship of the form shown in figure 6.7. In this pattern,  $A$  and  $B$  belong to an ensemble as well as  $C$

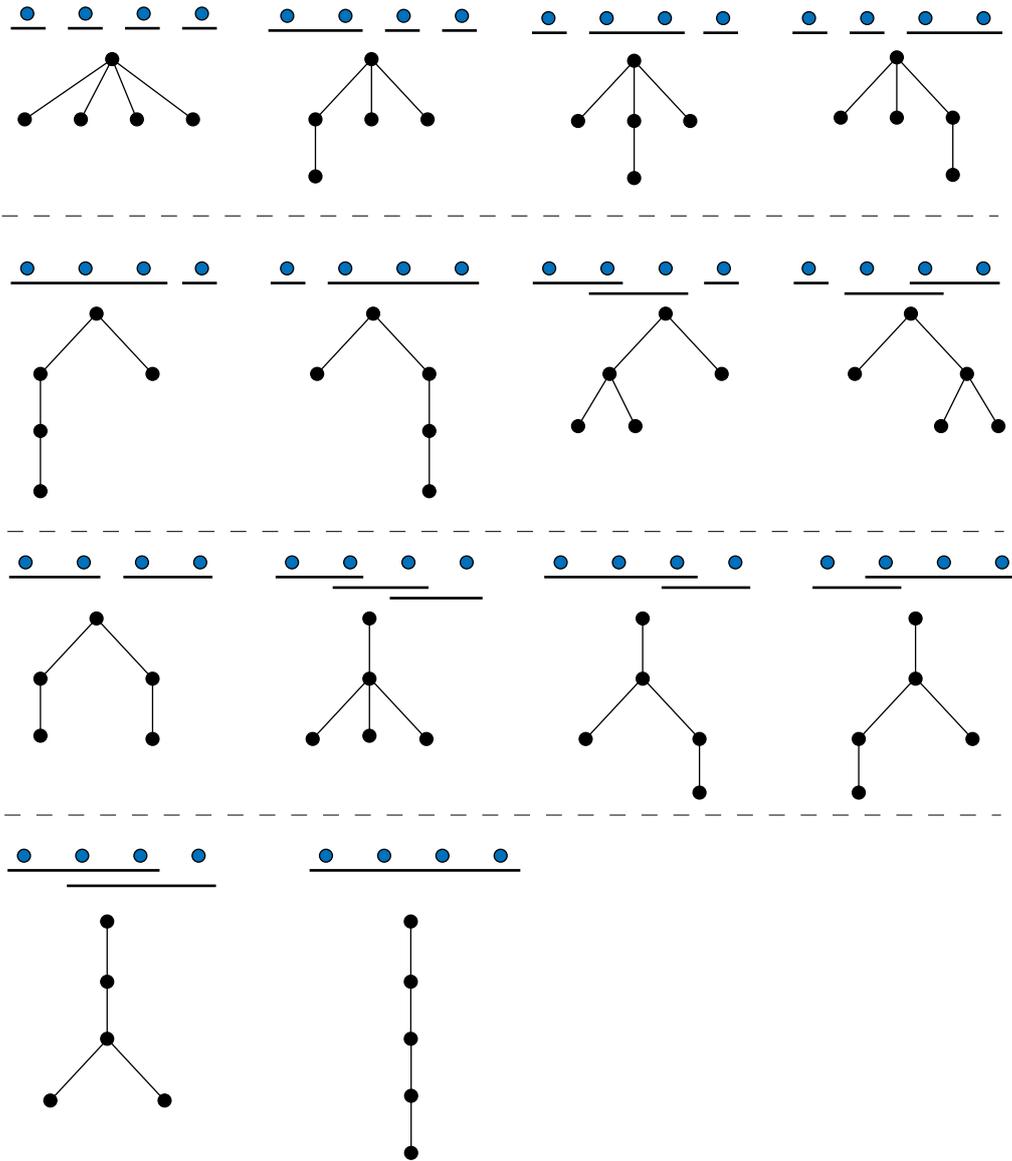


Figure 6.6: Semiorder bijection

and  $D$ , however, *part* of the first ensemble,  $A$ , also belongs to an ensemble with  $C$ . The same applies to  $B$  and  $D$  and *not* to  $B$  and  $C$  or  $A$  and  $D$ . The implicit logic in the trees forces us to conclude that all 4 neurons belong to a single ensemble, more details will be given when studying specific examples.

Another consequence of the bijection and the particular way in which we chose

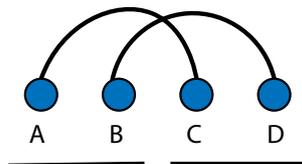


Figure 6.7: Crossing patterns are not represented in the semiorder

the rules is that the presented semiorder underlines two kinds of relationships. First the equivalence relation  $\sim$  in which  $A \sim B$  iff  $A$  and  $B$  belong to the same ensemble. Second, an ordering that shows how much variance each of the ensembles contributes to the total variance of the population. In order to understand the ordering and the proposed bijection better, consider the examples in figure 6.8.

Because the way the positive direction is chosen (see before), we can say that  $b > a$  in the first tree to the left. The corresponding semiorder or decision pattern is clear (there are only 2 options for  $n = 2$ ). For the others, following the fact that each descendant is greater than its parent with respect to the contribution and following the rules established in the bijection proof, it is easy to see how the ensembles are sorted according to their compound, overlapping contributions to the principal vectors.

We finish this part by underlying the fact that, incidentally, we have given a new proof of a well know fact about semiorders:

**Corollary 1.** *The number possible full ensemble semiorders for  $n$  neurons is the  $n$ th Catalan number  $C_n$*

Indeed, the trees just obtained are called *rooted trees* and the total number of rooted trees with  $n + 1$  nodes is known to be  $C_n$ [87]. Some examples of the trees and the corresponding semiorders are given in figure 6.9

### 6.5.2 The compound tree

The difference tree algorithm describes in fact a *family* of trees, one for each eigenvector we apply the given rules to. Each set of ensembles thus obtained is a valid ensemble in itself but it is also contained in the ensemble represented by the node it grew from; by preserving all the ensembles in each generation we obtain the *Compound tree*:

The compound tree can be seen as a *coarse graining* of the population activity and is related to the process of renormalization used in physics.

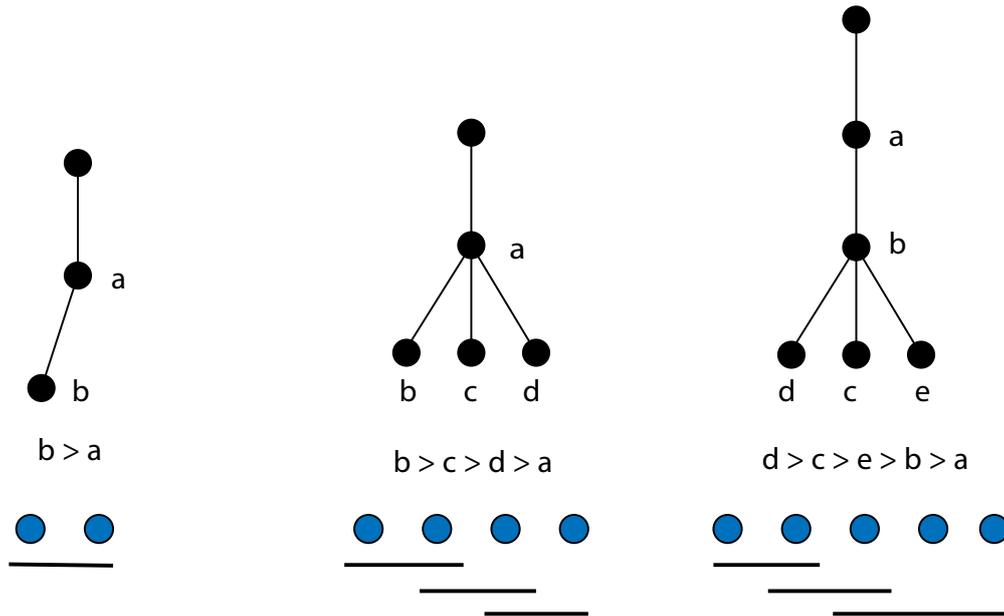


Figure 6.8: Examples of the ordering and the bijection

### 6.5.3 Coding tree

A final tree associated with the set of principal components is a binary tree called the *coding tree*. In this tree, instead of leaving behind the neurons in the closure of the negative half space as in the previous algorithms, those are also put in a new node; as a result, each step brings about two new children for each node. When a single neuron is reached, the rules can either stop growing (compressed tree) or continue adding redundant levels (full code).

Figure 6.10 illustrates the growing of the binary trees. It also shows how each node in the compressed tree has a natural binary code by following the path from

---

#### Algorithm 2 Compound tree

---

- 1:  $node_0 \leftarrow$  all neurons
  - 2:  $nodes \leftarrow node_0$
  - 3: **for**  $i = 0$  to  $k \leq n$  **do**
  - 4:     **for all**  $nodes$  **do**
  - 5:          $u \leftarrow$  neurons in *current node*
  - 6:          $h^+ \leftarrow j$  s.t.  $v_{i,j} > 0$  for  $j \in u$
  - 7:          $new\ node \leftarrow h^+$
  - 8:         add *new node* to *current node* as a child
-

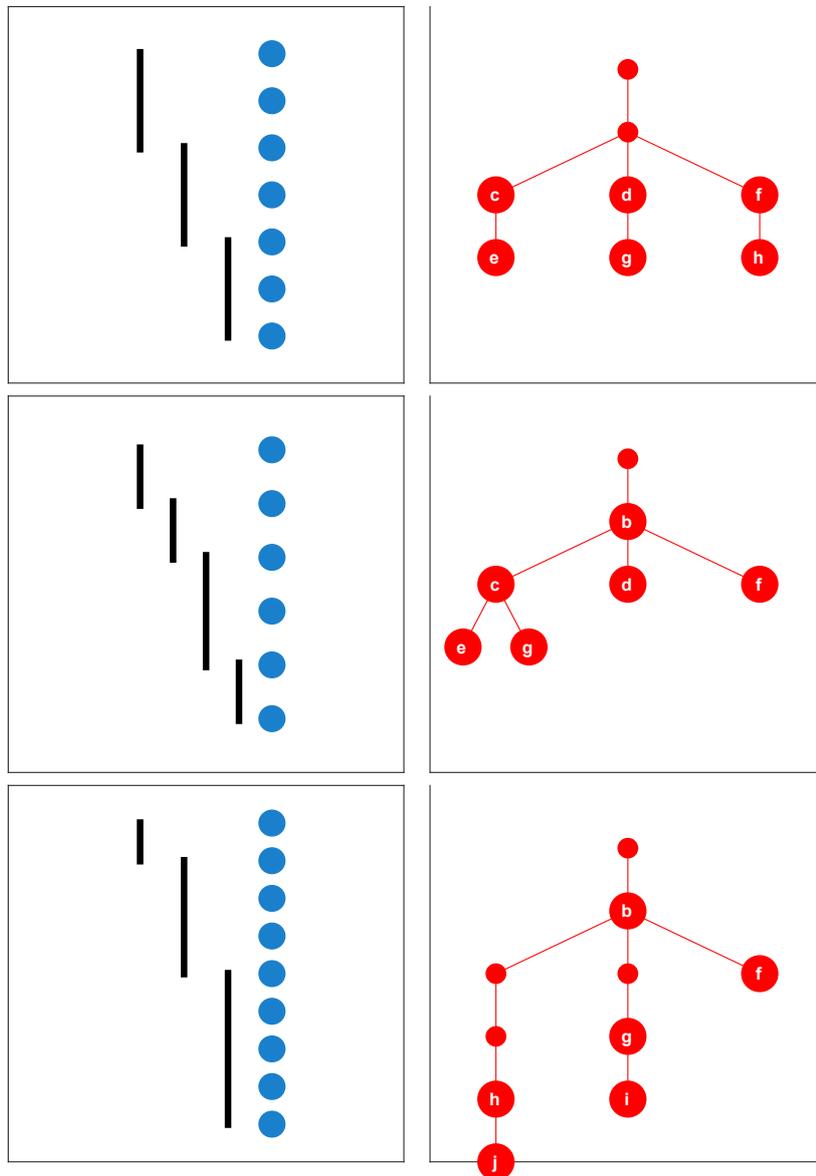


Figure 6.9: Example trees

the root and writing down 1 for right turns and 0 for left turns. The code, in the full tree, also points to the *chamber* (a chamber code) where the trajectory is expected if that ensemble is active.

Finally, it is important to add that there is a natural bijection of the rooted trees presented in before and these trees, more detailed treatment of these trees and their relationships can be found in [93]. An alternative bijection identifies

---

**Algorithm 3** Coding tree
 

---

```

1:  $node_0 \leftarrow$  all neurons
2:  $nodes \leftarrow node_0$ 
3: for  $i = 0$  to  $k \leq n$  do
4:   for all  $nodes$  do
5:      $u \leftarrow$  neurons in current node
6:      $h^+ \leftarrow j$  s.t.  $v_{i,j} > 0$  for  $j \in u$ 
7:      $h^- \leftarrow j$  s.t.  $v_{i,j} \leq 0$  for  $j \in u$ 
8:     new node  $R \leftarrow h^+$ 
9:     new node  $L \leftarrow h^-$ 
10:    add new node  $R$  to current node as a right child
11:    add new node  $L$  to current node as a left child
  
```

---

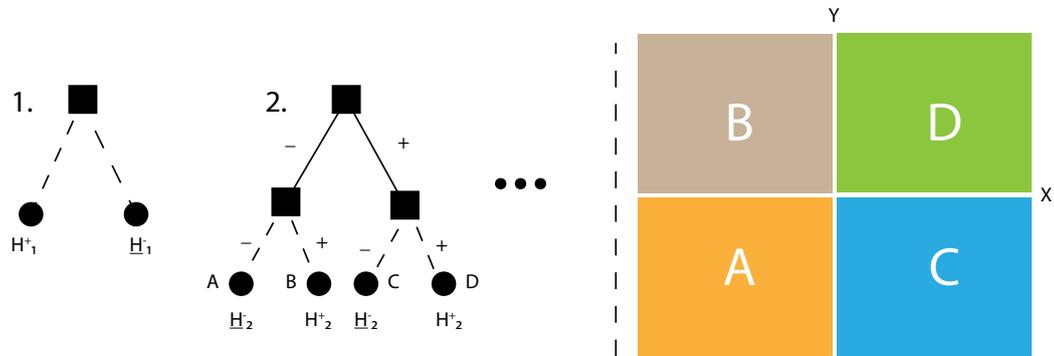


Figure 6.10: Growing of the coding trees

the each node with the corresponding one during the growth process of the tree (figure6.11)

## 6.6 The semiorder gives information about the embedding of neural low dimensional subspaces in the high dimensional space

Connecting back to the insights from section 6.3, the semiorder gives additional information about the positioning of the low dimensional subspace in the ambient space. As seen is the mathematical preliminaries, every semiorder has a symmetric and an antisymmetric part. The symmetric part can be interpreted as giving the different ways the data can be put in subspaces inside the high dimensional space. As an illustration, consider the case of 3 neurons. There are only 4 possibilities

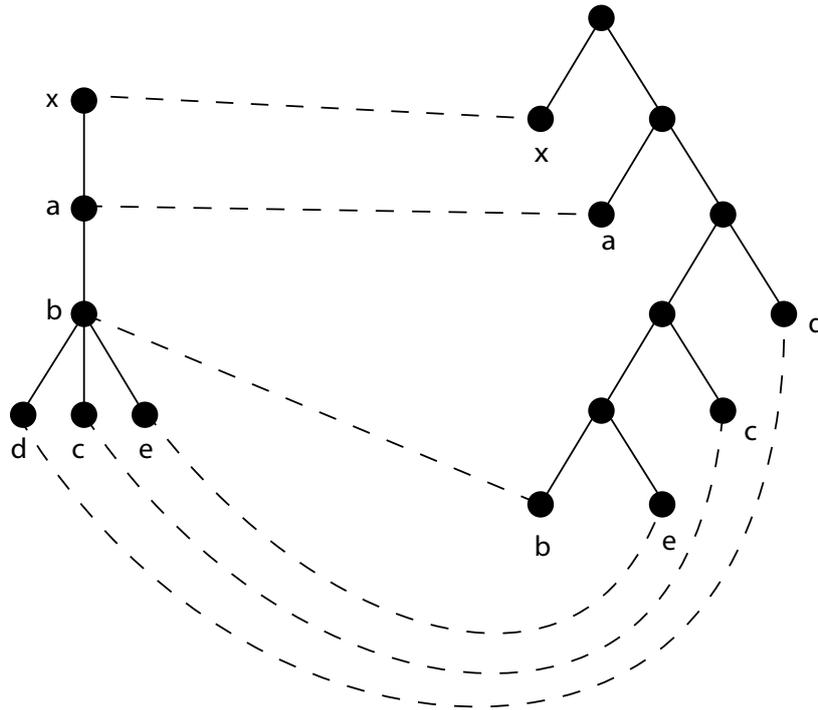


Figure 6.11: Bijection between the difference trees and the coding trees

(figure 6.12):

- $1 + 1 + 1$ : The 3 neurons are independent
- $1 + 2$ : Two neurons are correlated and one independent
- $1 + 1$ : The neurons lie in a proper subspace
- $1$ : The neurons visit all the space

The patterns can be continued for higher dimensions. All the patterns in figure 6.12 are symmetric and so correspond to the same semiorder, except  $2 + 1$ , which could also be  $1 + 2$ , giving a total of 5 possible patterns. In general, the number of patterns is, again, the  $n$ th Catalan number [87].

## 6.7 Mean subtraction and the center surround interpretation

Until now we have assumed that no mean subtraction is performed in the PCA but every analysis of data does, in general includes this pre-processing step. Algebraically, it is not important, after all, the resulting covariance matrix is only a

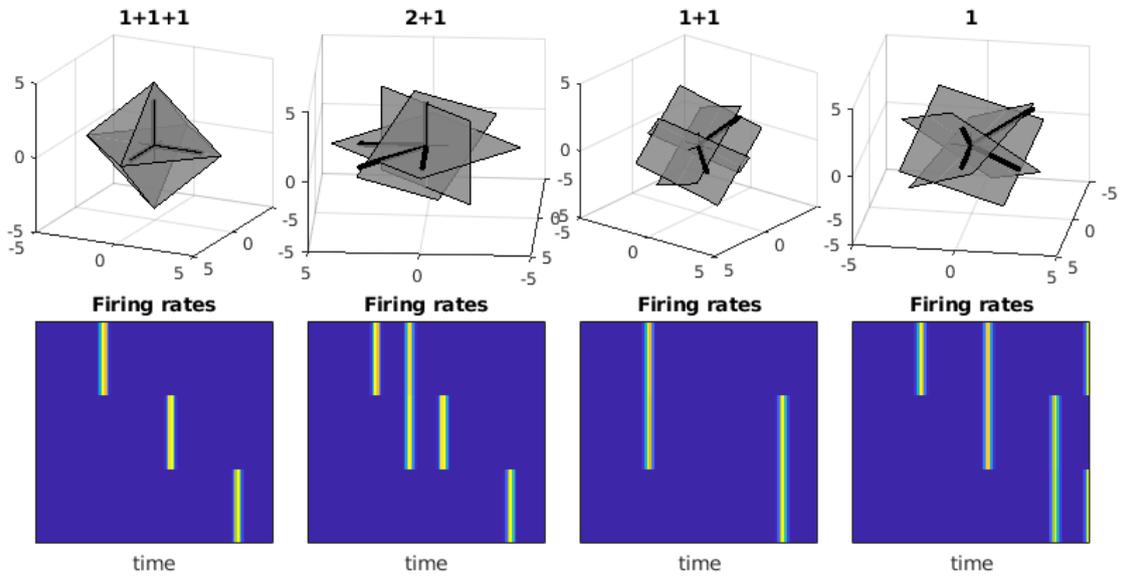


Figure 6.12: Possible ways in which the neuron activity can lie in the high dimensional space. From left to right, The three neurons can be disjoint meaning that we are ultimately embedding 3 one dimensional subspace in the 3 dimensional space. 2 neurons can be correlated but one uncorrelated; in this case we are embedding a 2 dimensional and a 1 dimensional subspaces. The three neurons can be linearly dependent with the 3 of them spanning a 2 dimensional subspace. All the neurons are correlated and span the whole space. Those are the only possibilities

rank one perturbation of the non mean subtracted one. However, mean subtraction has interesting yet subtle consequences that are the topic of this section.

Remember the differential formula for the eigenvectors given in chapter 3

$$\dot{\mathbf{v}}_i = \mathbf{v}_i^T \mathbf{x}(t) \sum_{i \neq j} \frac{\mathbf{v}_j^T \mathbf{x}(t)}{\lambda_i - \lambda_j} \mathbf{v}_j \quad (6.9)$$

This is a complicated differential equation that is coupled to the corresponding for the eigenvalues, however, it can be used to try to understand the location of the first eigenvector in the mean subtracted population. As seen in figure 6.13, when the mean is subtracted, the first eigenvector is positive, as expected for positive definite or semidefinite matrices. Mean subtraction positions this vector in an interesting position. Equation 6.9 suggest that neurons that display a predominantly negative firing rate will end up with negative coefficients in the first eigenvector while neurons with positive firing rates will have a positive coefficient.

The main aspects of the method remain valid, however, two consequences need

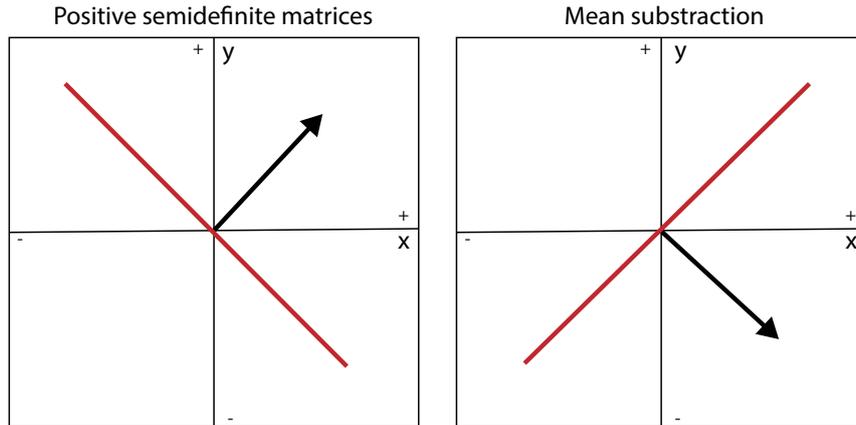


Figure 6.13: The effect of mean subtraction on the first eigenvector. In the right figure, the component corresponding to the neuron  $y$  indicates that this neuron has predominantly negative firing rates.

consideration. First, as illustrated in figure 6.14, in the developed algorithms the groups of positive and negative firing rates will be disjoint in the construction and therefore could be treated as separate groups. Second, and more importantly, if the population has *symmetric firing rates*, that is, one in which patterns of increases are mirrored by patterns of decreases, the two trees will capture this symmetries effectively separating the population in a group that we will refer to as the *center* and another that we will call the *surround*. The nature of the center (surround), whether it is the positive or the negative, depends of the physiological properties of the given neural population /region.

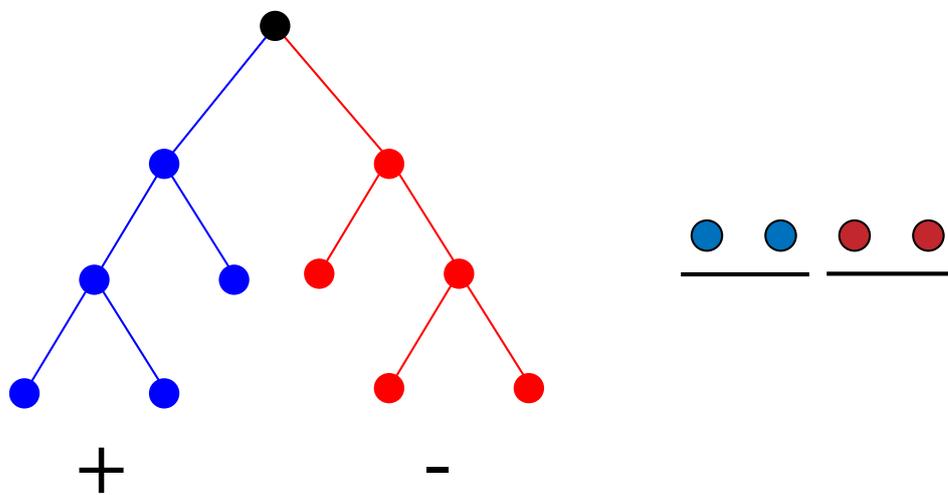


Figure 6.14: Paralell tree growing. Left and right branches are separated through the whole growing process, representing patterns that mainly increase or decrease the firing rates. To the right is the corresponding semiorder representation

## Chapter 7

# Neural ensembles and trajectory shape in the basal ganglia

## 7.1 Chapter overview

In this chapter we relate the shape of the trajectories as studied in chapter 5 to aspects of the population code in the basal ganglia, reflected in its ensemble structure. We attempt to answer the questions:

1. How does the population code changes in the different regions?
2. What aspects of the population activity generate the morphological features seen before in the principal trajectories?
3. What is the coding scheme at the output of the basal ganglia?

We start by showing, in section 7.2, that the activity at the output stage (SNr) is homogeneous and low dimensional. By means of a simple population vector approach, we study how similar is the code along the trajectory in the time interval around movement initiation. Moreover, the *intrinsic*<sup>1</sup> population features suggested by the so called *trajectory anatomy* are studied quantitatively by the development of 2 novel measures of heterogeneity and sparseness.

We then move forward to identify the specific patterns present in the population by means of the geometric ensembles approach developed in the previous chapter. We show that the population code is simplified along the processing pipeline of basal ganglia regions, with simpler, compressed signals at the output and complex signals at the input. The later as revealed by the presence of mixed selectivity neurons with complex responses to different stimuli. In terms of the geometry, we give interpretations to the sharp points, flat regions, vertices and irregularities observed in the different regions.

Overall, the results of this chapter confirm long standing hypotheses about the computational role of the basal ganglia in dimensionality reduction and compression of the signals/commands coming from the motor cortex.

## 7.2 The basal ganglia output behaves as a low dimensional gate for movement

Evidence of dimensionality reduction can be seen in the spectrum of the covariance matrix for each of the regions. In figure 7.1, the proportion of explained variance is shown for each of the studied regions. It is readily apparent that the SNr *explains* more variance with the first three eigenvalues (AOC = 9.5) than all the other regions. Even more, a large proportion of it is explained only by the first principal component.

---

<sup>1</sup>Intrinsic as opposed to extrinsic ones like selectivity to an external stimuli, see [73]

Among the subpopulations, the Arky subgroup of the GP has an equally large first principal component but a slow increase in the explained variance, making it higher dimensional than the Protos for the chosen (arbitrary) threshold. MSNs are lower dimensional than FSIs.

We now focus on the aspects of the population code that make regions like SNr and STN lower dimensional while others like Str and GP higher dimensional.

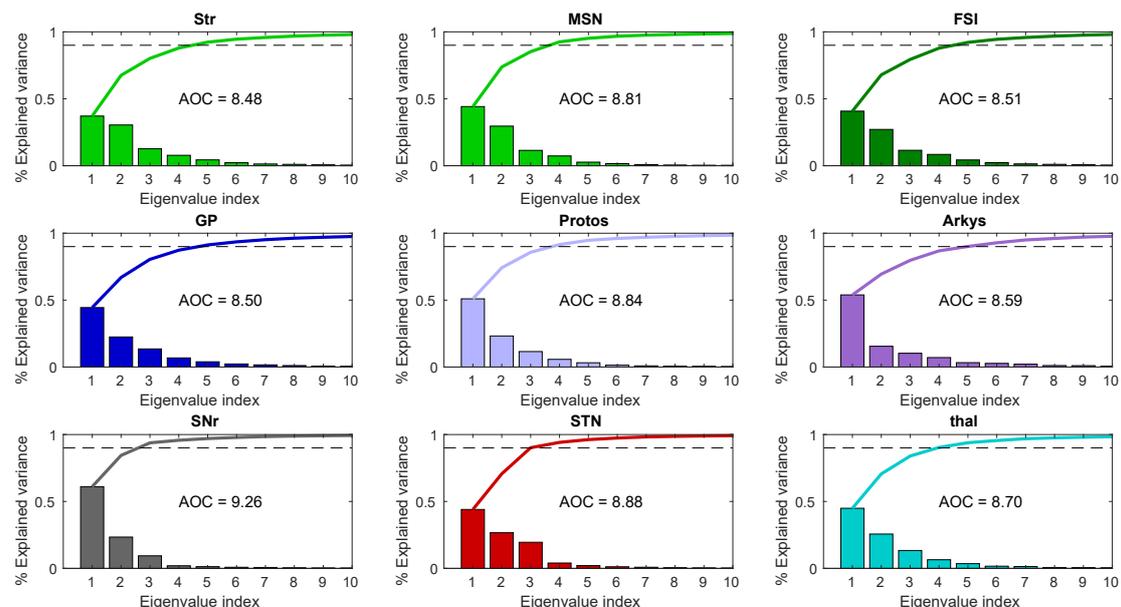


Figure 7.1: Explained variance in each of the basal ganglia regions studied including sub populations. The bars are the normalized eigenvalues of the covariance matrix. The solid lines are the cumulative spectra. The dashed line indicates an arbitrary threshold of 90% explained variance. SNr shows the fastest increase in explained variance per eigenvalue, followed by STN.

### 7.2.1 The anatomy of the principal trajectories suggests characteristic time resolved ensemble structures for each region

The anatomy of the trajectory is a methodology inspired by similar work in chemistry [51] and, more recently, in the study of the population code of the *C. elegans* [47]. It allows us to visualize how the population vector changes along the trajectory. Highly clustered, homogeneous and low dimensional trajectories are expected to show minimal changes in the population vector with time.

## Method

Figure 7.2 is constructed as follows. The insets show the similarities of the population vector at each point in the trajectory. First, the vector of firing rates at each time point  $x(t) \in \mathbb{R}^n$  is binarized based upon whether each neuron fires above a certain threshold or not. Then, the matrix of *hamming distances* between those vectors is built and used to produce a *multidimensional embedding* in 2 dimensions in such a way that close points in the plane will reflect patterns that are close in the hamming distance sense.

The insets are then the result of choosing equally spaced time points and plotting the corresponding points naming each pattern with a letter in alphabetical order. Those points could be considered different *states* of the population in the phase space and the final inset, a state diagram. The size of the circles is proportional to the number of active neurons at that point.

In the trajectory, the black point represents movement onset and the diamond is the starting point 500ms before movement. Three dimensional trajectories are used for reasons that will become apparent later in the chapter.

## Anatomy of the different regions around movement initiation

Str displays rather different patterns at each time point suggesting a high dimensional trajectory and that 3 dimensions are unlikely to capture all the details of the population activity. The equally spaced points starting at D mean that the ensembles are constantly changing but none is particularly different from the others. The trajectory before that is homogeneous due mostly to the effect of the filtering of the go cue distant in the past.

The other regions display 2 major branches with different ensembles separated by what can be interpreted as a “transition” point (around F in GP, E in STN, G in SNr and F again in Pf). In SNr and STN, the transition point is markedly sparse, as evidenced by the size of the corresponding circle. In those areas, the patterns before and after the transition point are markedly different.

Note that the transition points in SNr and STN, translate directly into sharp points of the trajectory (see later). The three patterns of the GP, on the other hand translate into a “polygonal” shape with 3 vertices (see chapter 5). Other aspects of the trajectory in relation to the trajectory will be studied later.

What are those patterns and where does the abrupt change before and after the transition come from? Are there any behavioral relevant events that can be associated with each “branch” of the trajectory? We start gaining more information about the firing patterns in the population by studying its ensemble structure in the next sections but first we want to quantify the observations made here in terms of number of patterns (heterogeneity) and number of active neurons (sparseness).

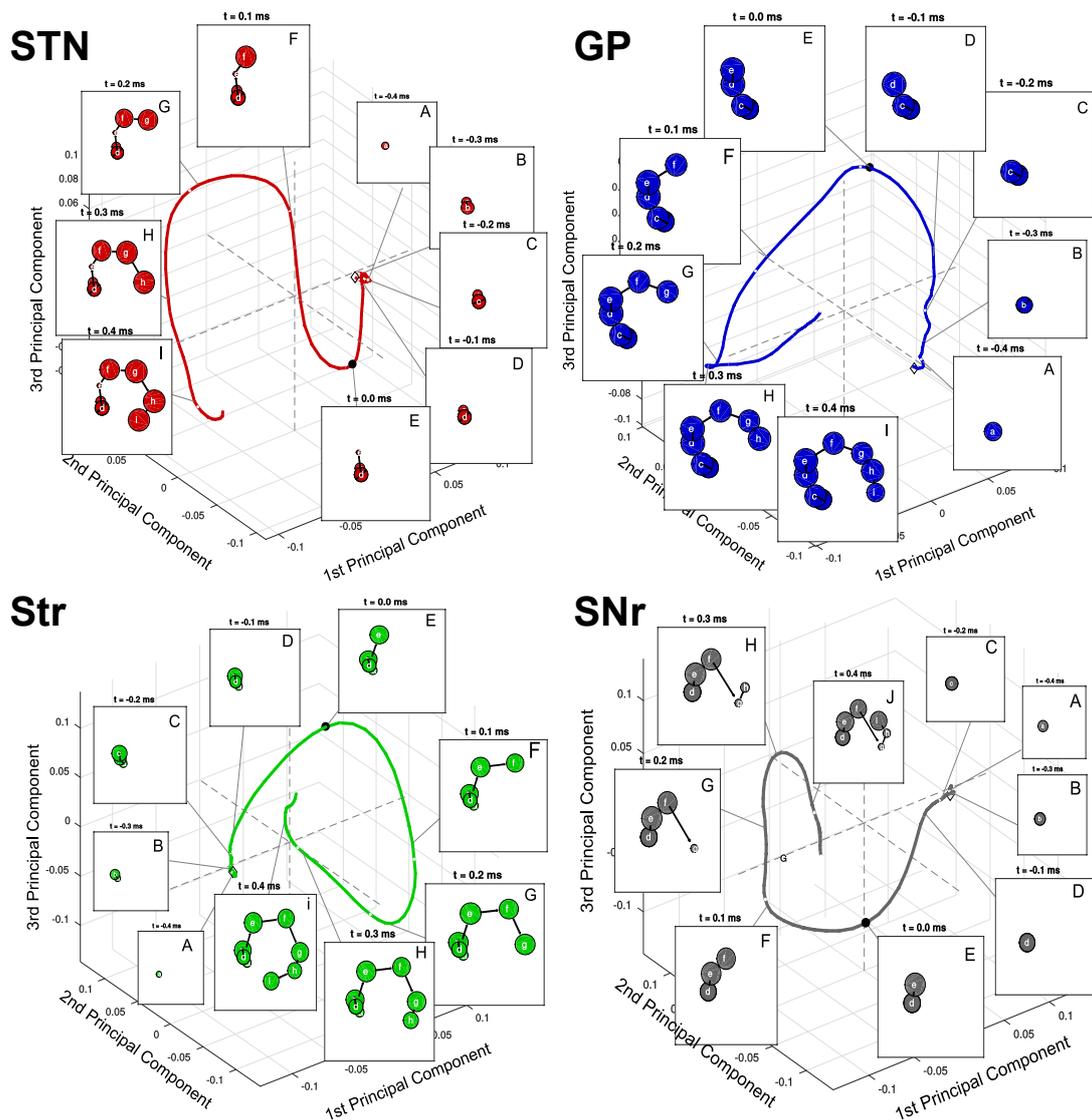


Figure 7.2: Anatomy of the trajectory for regions STN, GP, Str and SNr. The input region (Str) evolves traversing more or less equally sized ensembles, equally spaced in the phase space. This is in contrast with the sharp transition between two states in the output stage (SNr). The STN has two sections, one with homogeneous evolution and one with heterogeneous one similar to the Str. GP has three distinguishable ensembles separated by large changes (More detail in the main text)

## 7.2.2 The observed low dimensionality comes from separated but homogeneous clusters

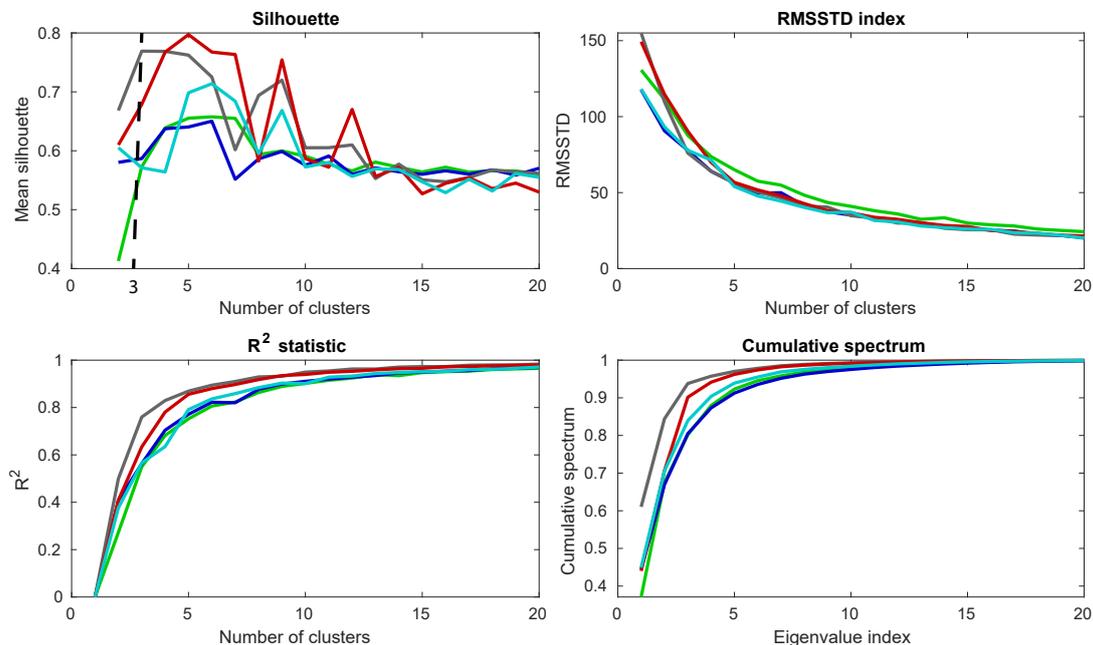


Figure 7.3: Evidence of dimensionality reduction in the basal ganglia. From top to bottom and left to right. The silhouette index reveals highly coherent clusters in the SNr and STN after for two clusters. The RMSSTD index suggest comparable distances between clusters but a sharp decrease is noticeable for the SNr after the second cluster. The  $R^2$  statistic and the PCA cumulative spectrum provide similar information about the quality of the clustering. Colors of the lines correspond to the color code used through the thesis for the different regions.

In order to quantify the observations made in the previous section we use two approaches. First, by means of a conventional clustering algorithm and measures of clustering performance [80], we show that the intuitions behind our analysis are sound. Then, we introduce new measures of the population properties derived directly from the PCA structure, reaching similar conclusions in a more self contained analysis.

Figure 7.3 shows the result of applying k-means clustering algorithm[29] to the set of population vectors (states) traversed by the trajectory. The silhouette index [79] peaks earlier for the SNr followed by STN, indicating the population patterns fit well into 3-4 clusters (high coherence). The separation between clusters, as measured by the Root Mean Square Standard Deviation (RMSSTD) remains comparable among regions but has a sharp drop after 3 clusters for the SNr and

is slightly higher for the Str, conforming to the observations made in the previous sections.

A common measure of the quality of clustering is the  $R^2$  statistic [80]. Again, it raises faster for the SNr and STN indicating cluster with higher internal coherence. Notice the similarity of the this measure with the cumulative spectrum from PCA (figure 7.3, bottom right). We conclude that the differences in dimensionality between regions arise as a consequence of separated homogeneous clusters in the state space.

### 7.2.3 Heterogeneity and sparseness measures support the dimensionality reduction view of the basal ganglia

Among the different features of a population code (see chapter 2), heterogeneity and sparseness are of particular interest from both the modeling and the experimental points of view. They are in direct relation with information theoretical concepts like the compression rate of traditional codes. They also constraint the dynamics, representations and computations performed by a given population [1]. It is interesting to see how these properties vary across different basal ganglia regions and how they relate to the observed shape and clustering features previously described.

#### Heterogeneity

Formulas (6.4) and (6.7) suggest a way of measuring how heterogeneous the firing patterns of the neurons are. Remember that they are related to how fit a set of points is to the given basis. If the points are very close to the coordinate axes, they will be, by definition, well explained by each PC and therefore, we say that the population is *homogeneous*. If they are far from the basis, they are called *heterogeneous*.

One caveat to the last interpretation is that, if we decide to use formula (6.7) as a measure of heterogeneity, the usual maximum of this convex function will not correspond to a population of maximal entropy. Indeed, the eigenspectrum of the covariance matrix gets constrained by the time correlations in a way that was treated analytically in chapter 5.

**Definition 20** (Life time Heterogeneity). *The lifetime heterogeneity of a matrix of firing rates is defined as the divergence[23] between the empirical spectrum and the theoretical one,  $\{\eta_i\}$ , derived in chapter 5 for an exponential or Gaussian null covariance matrix model. That is*

$$H(\mathbf{X}) = \sum_i \lambda_i \log \left( \frac{\eta_i}{\lambda_i} \right) \quad (7.1)$$

If the null covariance model is exponential, then

$$H(\mathbf{X}) = - \sum_k \lambda_k [(a - \log b_k) - \log \lambda_k], \quad (7.2)$$

with  $a = \log(2\sigma/L^2)$  and  $b_k = k^2\pi^2 - L^2\sigma^2$ . If the covariance model is Gaussian, then

$$H(\mathbf{X}) = \sum_k \lambda_k (c_k + \log \lambda_k), \quad (7.3)$$

with  $c_k = k^2\pi^2\sigma/L^2$ . In both cases,  $L$  is the length of the measurement interval and  $\sigma$  the width of the fitted null autocovariance.

Therefore, the life time heterogeneity gives a measure of how many different patterns there are in the *whole interval* around the alignment event. We propose also a time resolved version of that measure. As was seen in chapter 3, the time-covariance matrix can be seen as a function of time on the interval around the even. The rate of change of its eigenvalues is proportional to the square of the similarity between the corresponding eigenvector and the population vector at that time, then we have

**Definition 21** (Instantaneous heterogeneity). *The instantaneous heterogeneity is defined as*

$$\begin{aligned} \Delta H(t) &= \sum_i \dot{\lambda}_i(t) \log \left( \frac{1}{\dot{\lambda}_i(t)} \right) \\ &= \sum_i (v_i^T x(t))^2 \log \left( \frac{1}{(v_i^T x(t))^2} \right) \\ &= \sum_i z_i^2(t) \log \left( \frac{1}{z_i^2(t)} \right), \end{aligned} \quad (7.4)$$

where  $z_i(t)$  is the  $i$ th PC.

In each of the entropy expressions, the arguments must be appropriately normalized to be a probability distribution. Intuitively, the instantaneous heterogeneity tell us the dimensionality of the *face* that the trajectory is visiting at each time step. Note that each PC contributes to the instantaneous heterogeneity as long as the trajectory is far enough along that axis at the specific instant.

### Population sparseness

In the same guise, we can study the sparseness or sparsity of the neural activity around movement. In this case, do sparse ensembles contribute to the shape?

The 2 first PCs represent broad patterns that tend to hide sparse responses unless they are strong enough to capture enough variance. In any case, a measure that captures the intuitions behind counting the number of neurons contributing to the aligned event is

**Definition 22** (Population sparseness). *The population sparseness is defined as*

$$S(X) = - \sum_i \frac{\|x_k\|^2}{K} \log \frac{\|x_k\|^2}{K}, \quad (7.5)$$

where  $\|x_k\| = \int x_k^2(t)dt$  is the norm in the space of functions and  $K = \text{Tr } \mathbf{C}$  is the normalization constant.

The population sparseness, being an entropy (other entropy functions could be used like Rényi entropy), tells you the width of the distribution of activation of the neurons in the whole interval. A sparse population will have low entropy or, in other words, will be very thin. A dense population will have the activity spread across all neurons and, therefore, will be wide. In practice, all neurons tend to be activated at some point in the interval and therefore, have a total activity different from zero. A more useful measure of sparseness is the *instantaneous sparseness*.

**Definition 23** (Instantaneous sparseness). *The instantaneous sparseness is defined as*

$$\Delta S(t) = - \sum_i x_k^2(t) \log x_k^2(t), \quad (7.6)$$

which gives a measure of the width of the instantaneous population distribution of firing rates. Note that the last expression (equation 7.6) differs from equation 7.4 in that the former can be seen as measuring the distance of the data to the canonical basis, the canonical basis represents single neurons. Being very close to the canonical basis is a signature of sparseness. Equation 7.4, in contrast, measures the distance of the data to the PCA basis or to the best fitting hyperplane; the PCA axes represent ensembles as we have already seen.

## Heterogeneity and sparseness in the basal ganglia

In figure 7.4 we show the heterogeneity and sparseness of each of the studied regions in the heterogeneity sparseness plane on the left. On the right, we make sure that the observed differences are not due to the size of the corresponding populations.

The GP and their subpopulations are the most heterogeneous. As seen before, the subpopulations with sharp points consist of relatively few patterns of activation before, around and after movement in our measurement interval; they are therefore more homogeneous. As we saw before, these regions share the feature of a sharp

point that is due to disjoint ensembles related with different aspects of movement. Although this measure is theoretically well justified, it can be seen as an lower bound (due to Jensen's inequality) of the total time resolved heterogeneity that we investigate now.

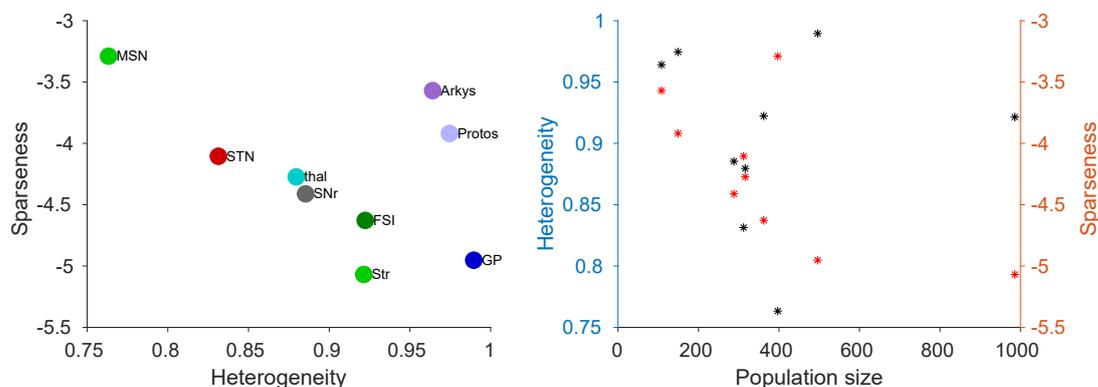


Figure 7.4: Basal ganglia regions in the Heterogeneity-sparseness plane (left). The (lack of) correlation of the measures with the size of the population(right). Neurons with sharp points are closed in this space. The MSNs are sparse as expected along with arkys and protos. No correlation between the size of the population and the measures' values is observed.

In figure 7.5(a), we show the instantaneous heterogeneity and sparseness for each region. Two aspects become apparent. In some regions (black arrows for example) the heterogeneity and sparseness are uncorrelated; the more neurons have different patterns (high heterogeneity), the more neurons become active for those patterns (low sparseness). On the contrary, in some segments of the SNr, STN and GP for example, increases in heterogeneity are co-occur with increases in sparseness (white arrows). That situation corresponds to the pauses (center) and increases (surround) that potential allow/inhibit movements [66].

Movement initiation is associated with a drop in sparseness meaning that most neurons respond. That response is markedly heterogeneous in subpopulations like the Arkys and FSIs and is homogeneous in the SNr and MSNs. After that the population response diversifies in ways that we will see in the next section.

Note that drop is early in the Str and later in the STN and Pf. This is related to the dimensionality reduction approach to PCA in that, while the input structures have high dimensionality (remember that heterogeneity tells indirectly the dimension of the face the population is visiting), the output ones visit lower dimensional subspaces (see later). Note that in the case of the SNr, the population behaves differently at the end of the interval.

Another way to visualize the compound heterogeneity and sparseness is to

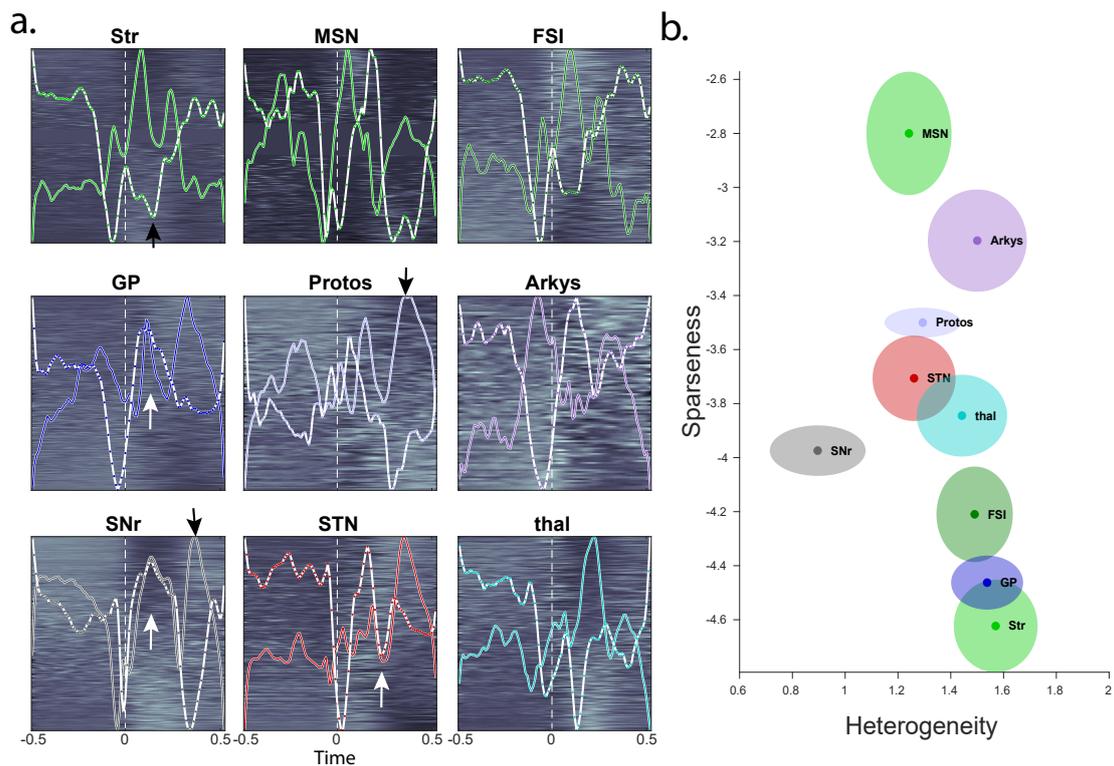


Figure 7.5: a. Time resolved Heterogeneity and sparseness. Heterogeneity: Solid colored line, Sparseness: White line with colored dots. b. Average life-time heterogeneity and sparseness. The ellipses around each point indicate the standard deviation for each of the measures.

study the mean and standard deviation of the instantaneous measures (figure 7.5 b). This paints a slightly different picture than the one given in figure 7.4. Here, the SNr is the more homogeneous one capturing the phenomenon found in the clustering. The ellipses show the variability smaller in SNr, protos and GP.

In conclusion, the measures of heterogeneity and sparseness provide an intrinsic description of the properties of the code. An extrinsic description [73] would require knowing the specific map of actions / event to codewords, this is unlikely to be achieved any time soon in the basal ganglia (i.e. which pattern encodes which action). The trees, along with those measures, are therefore a reasonable characterization of the coding scheme of each region.

## 7.3 Movement initiation occurs through parallel channels with specific functional roles at the basal ganglia output

We now delve more into the nature computation in the basal ganglia, the origin of the low dimensional activity at the output and the reason behind the geometric features observed. We use the geometry as a guiding principle in studying the ensemble structure of the population. One of the striking observations of this work is that sharp points do not always come from sharp changes in the underlying firing rates; more often they reveal aspects of the underlying population subspaces.

### 7.3.1 The singularities are associated with orthogonal subspaces of the population trajectory

Interest in the neural trajectories has come with an increase interest in the underlying geometrical subspaces [81]. Often, telling whether the involved subspaces are orthogonal or “overlapping” is of interest. Here we show that orthogonal subspaces are involved in the generation of movement. The main indication of orthogonality are singularities<sup>2</sup> in the low dimensional trajectory. There are three main populations with singularities (sharp points) in the data set: SNr, STN and Pf. Those sharp points are qualitatively different and reveal different population phenomena.

As it can be seen from the 3D trajectory (figure 7.6, left), the singularity of Pf is stable or high dimensional. This means that it is preserved under perturbations of the projection angle. This is likely and indication of sharp increase in the firing rate itself but it will not be studied further in this work. For SNr and STN however, the singularity is unstable; it reveals instead two salient features of those populations.

For the SNr, evolution around movement initiation happens in the plane spanned by PCs 1 and 3. Evidence of this is the one dimensional evolution, parallel to the line “L”, that is preserved under rotations around PC 2 (figure 7.6, right). For STN, evolution happens in two different subspace orthogonal to each other. Note that the projections are always parallel to the lines “L” and “M” around the singularity.

### 7.3.2 The population dynamics in the SNr is driven by minimally overlapping ensembles

Where does this orthogonality come from? Theoretically, subspaces derived from PCA can only be orthogonal if the underlying population vectors are *uncorre-*

---

<sup>2</sup>Remember that sharp points and irregular points can be indication of singularities.

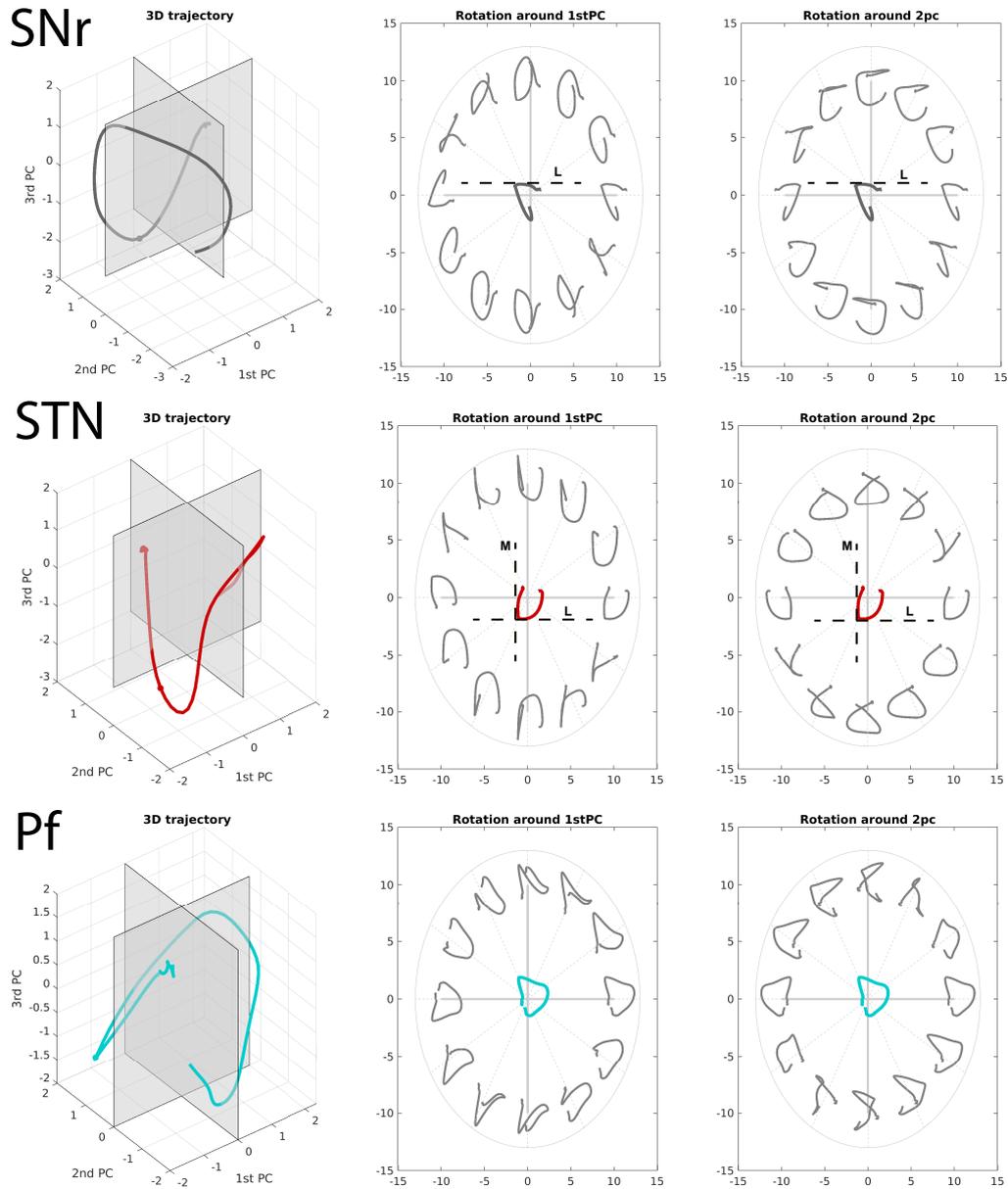


Figure 7.6: Stability of the singularities in the basal ganglia. Left. 3 Dimensional PCA trajectory showing evolution restricted to orthogonal subspaces in some stretches of the curve. Right. The stability of the singularities can be studied by rotating the trajectories and projecting them back to two dimensions. The only stable singularity is Pf's. For SNr and STN the orthogonal subspaces are revealed by the lines “L” and “M”

lated. We study now the patterns of firing in the population to determine if such independence is present at the level of the neuronal ensembles.

### Ensembles responsible for the evolution of the 2D trajectory

Figure 7.7 shows the Catalan ensembles for the SNr. As seen in the previous chapter, four ensembles contribute to the trajectory in PC1 - PC2 plane. The dots to the left indicate the pattern of overlaps in the form of a semiorder. The decreasing and increasing patterns are in different branches of the semiorder as expected. The distributions at the top indicate that late and early parts of the interval are associated with the Go cue (orange) and the sideIn movement (blue). The first two principal components capture therefore *broad patterns of firing* in the given interval.

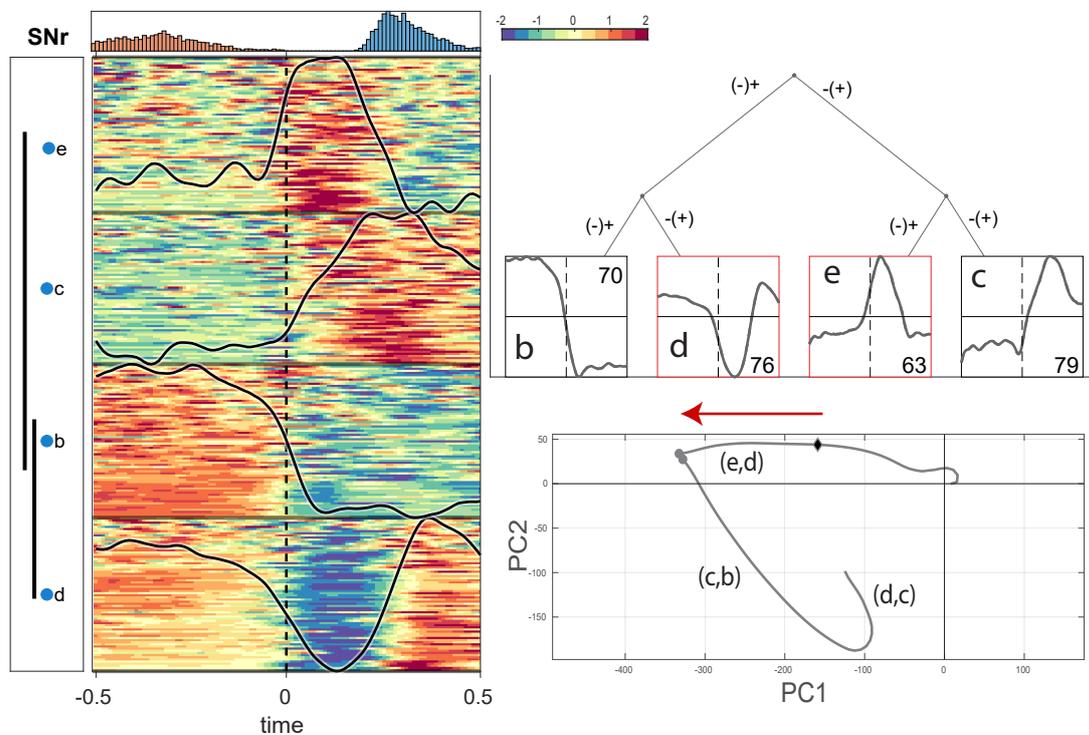


Figure 7.7: Level 2 ensembles in the SNr. Left. The corresponding semiorder. z-scores of the neurons are sorted and grouped by the Catalan ensembles. Distributions at the top show the previous and future events in the task. Right. The coding tree and corresponding trajectory with the ensemble contributions. The red arrow and border line indicates patterns related to movement initiation. Inside each leaf of the tree the mean z-score for the corresponding ensemble is shown. Numbers inside the panels are the number of neurons associated to each ensemble.

Remember that the principal components and, hence the trajectory, are linear combination of the original neurons' firing rates. Grouping those linear combinations by the contributions of the different ensembles gives:

$$\begin{aligned}
z_i(t) &= \sum_k v_k^i x_i(t) \\
&= \sum_{i_b} v_{i_b}^i x_{i_b}(t) + \sum_{i_c} v_{i_c}^i x_{i_c}(t) + \sum_{i_d} v_{i_d}^i x_{i_d}(t) + \sum_{i_e} v_{i_e}^i x_{i_e}(t) \\
&\approx \pm \bar{x}_{i_b}(t) + \pm \bar{x}_{i_c}(t) + \pm \bar{x}_{i_d}(t) + \pm \bar{x}_{i_e}(t),
\end{aligned} \tag{7.7}$$

so, the *position* vector of the trajectory is driven by 4 terms, one for each ensemble; moreover, it can be seen as approximately driven by the pattern (i.e. mean  $\pm \bar{x}$ ) in each ensemble in the direction given by its sign of the eigenvector. In the same vein, the *tangent* vector

$$\begin{aligned}
\dot{z}_i(t) &= \sum_k v_k^i \dot{x}_i(t) \\
&= \sum_{i_b} v_{i_b}^i \dot{x}_{i_b}(t) + \sum_{i_c} v_{i_c}^i \dot{x}_{i_c}(t) + \sum_{i_d} v_{i_d}^i \dot{x}_{i_d}(t) + \sum_{i_e} v_{i_e}^i \dot{x}_{i_e}(t) \\
&\approx \pm \dot{\bar{x}}_{i_b}(t) + \pm \dot{\bar{x}}_{i_c}(t) + \pm \dot{\bar{x}}_{i_d}(t) + \pm \dot{\bar{x}}_{i_e}(t),
\end{aligned} \tag{7.8}$$

is also decomposed in contributions from each ensemble. Using this information we can locate the ensembles contributing to the trajectory and ultimately to the points of interest, as follows.

With this in mind, we use the *coding tree* to find the ensembles responsible for the sharp point and irregularity in the SNr. Note that the sharp point in figure 7.7 (bottom right) is in the IV quadrant of the Cartesian plane. According to the chamber code 7.7 (top right), the two ensembles that are contributing to this point are e and d. Indeed, around movement, the tangent vector is driven by the components corresponding to those ensembles.

The first one can be characterized as an increase of firing rate that is persistent, on average, during approximately 200ms after movement (surround). Ensemble d is its approximate mirror image, that is, the corresponding movement related pause (center).

The trajectory can then be described as follows: It starts in quadrant I (++) driven mainly by the high firing rate in ensemble b and the low firing rate in ensemble c. It is not far along the first quadrant mainly because of d which is driving it towards (+-). The decrease in d along with the increase in e bring the trajectory to quadrant II where it stops. As it was shown in the previous chapter,

that stop lasts around 100ms and then the trajectory turns towards the (-) driven mainly by the increase in  $c$  and the decrease in  $b$ . At the end it makes the hook style turn as it starts being driven to (+-) by the increase in  $d$  and the decrease in  $e$ .

### Phenomenological origin of the sharp point in 2 dimensions

From the previous observations we can conclude that before movement initiation, the trajectory is barely moving as the velocity vector is close to zero. The evolution of the trajectory after movement is the result of three patterns of activity as illustrated in figure 7.8 (b). First a decrease in the firing rate and consequent rebound (bump no. 1), second, an increase of firing rate at the time of movement (bump no. 2) and third, a late increase potentially related with subsequent movements.

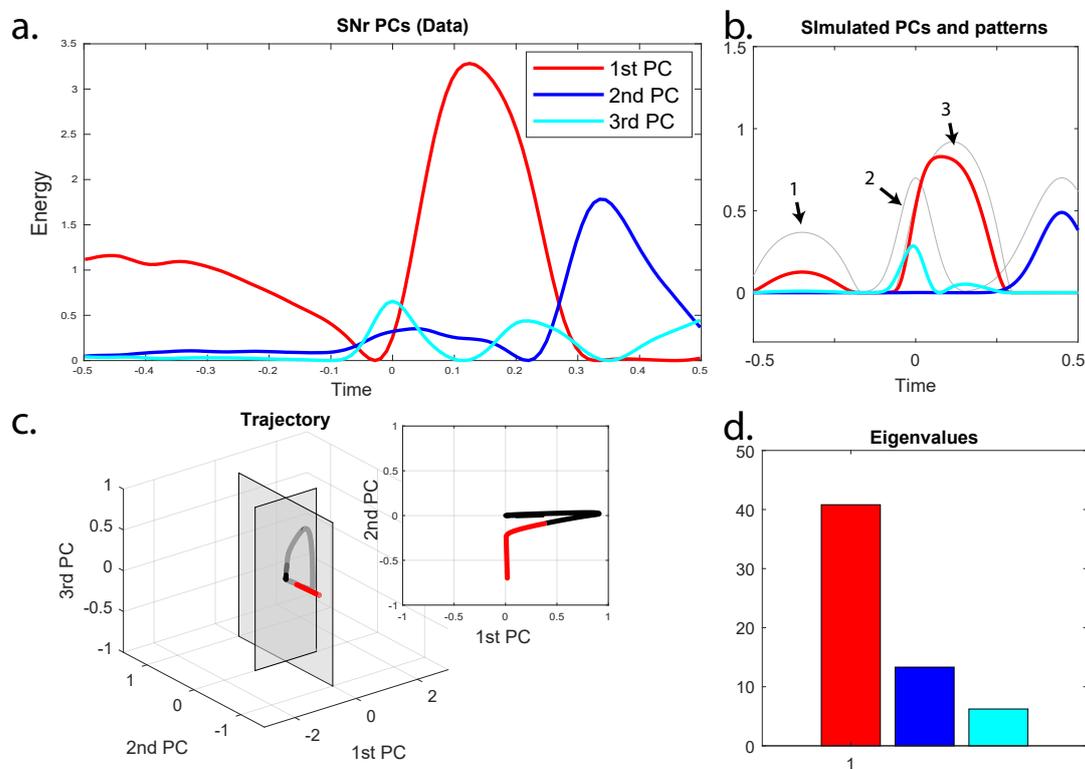


Figure 7.8: SNr phenomenological model. a. The *energy* (magnitude square) reflect the locations of the time interval they “care about”. A simple model of there firing rate bumps generate similar looking components (b.), trajectories (c.) and spectrum (d.).

From the figure it can be seen that such a simple descriptive model generate PCs

that are similar to the observed ones (figure 7.8(a)). The trajectory generated this way also displays constrained evolution to the PC1-PC3 plane and singularities. Finally, the eigenvalue spectrum as a large first eigenvalue similarly as with the biological population.

Note that both in the model and in the data, the second principal component is almost independent from the other two. In the region in which they overlap it only contributes a constant translation to the trajectory. That means that the second eigenvalue reflects mostly variance coming from this pattern of activity.

### 7.3.3 The third PC can be used to separate the population into different channels with specific computational roles

Is this the case in general? Are the ensembles in the basal ganglia parallel channels for different behavioral signals? To study the answer to this question we grow the corresponding Catalan tree one step more (figure 7.9). In doing so we reveal interesting patterns of activity in the population each one with a characteristic time evolution.

The super ensemble previously named “e” now separate in two patterns “i” and “e” (ie). The later is locked at the time of movement while the former happens earlier but before than the next movement recorded (sideIn), therefore its functional role is uncertain. Super ensemble “c” is now “c” and “g” (cg). “g” has activity that is clearly related to the sideIn event. “c” on the other hand has a ramping behaviour towards late aspects of the task. Finally, the super ensemble (bfdh) has cue related high firing rate and movement related decrease; the sub-ensemble (dh) has in addition and increase associated to late aspects of the task.

In this case we can also use the coding tree to locate the ensemble activity in the trajectory. We observe, for example that the irregular part of the singularity, the part parallel to PC3 is indeed related to the previously described undefined motor related activity of ensemble “i” and its corresponding mirror image. Overall, the location of the main firing rate changes allows to partition the trajectory in different behaviorally related segments (figure 7.10).

#### Looking through the movement glass

In a more exploratory way and with the aim of strengthening the observations made before, we perform different alignments to nearby events to study the behaviour of the Catalan ensembles under different events. Two situations are possible in this case. If the neurons display mixed selectivity [78], different ensembles should be randomly assigned to stimulus related computations (i.e. tone or go cue) or movements. If they are part of a system of parallel channels, each ensemble should preserve its identity without need to regroup.

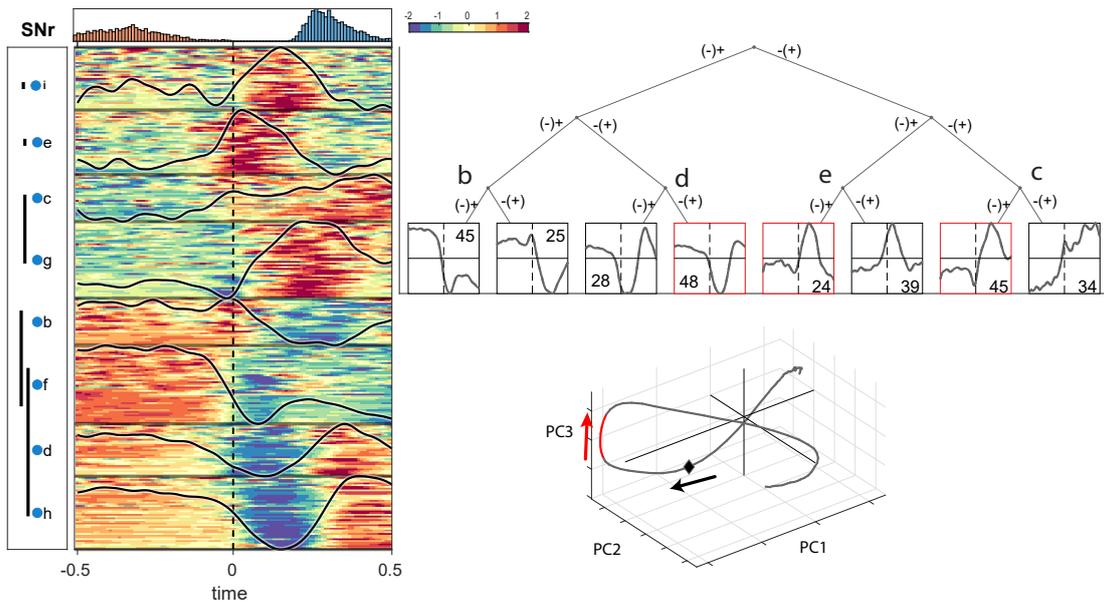


Figure 7.9: Level 3 Catalan tree for the SNr. The layout is the same as for figure 7.7. The red arrow and red squares are the portion of the trajectory, and contributed ensembles, that evolves parallel to PC3.

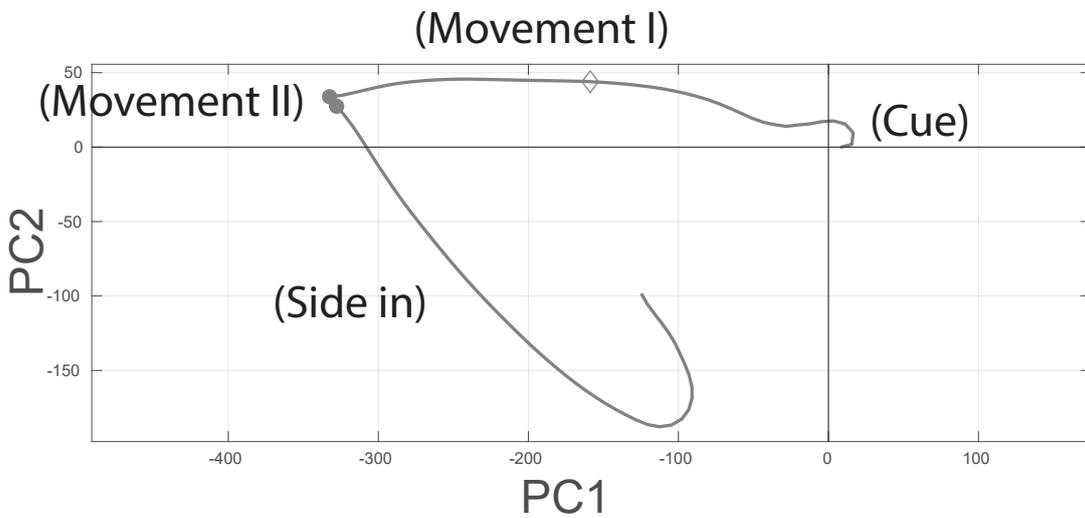


Figure 7.10: Summary of the population code in the SNr. Each segment of the trajectory is associated with the putative signal conveyed by each ensemble.

As seen in figure 7.11, the later seems to be the case. Specifically, some of the ensembles have specific roles and seem to contribute to different aspects of the

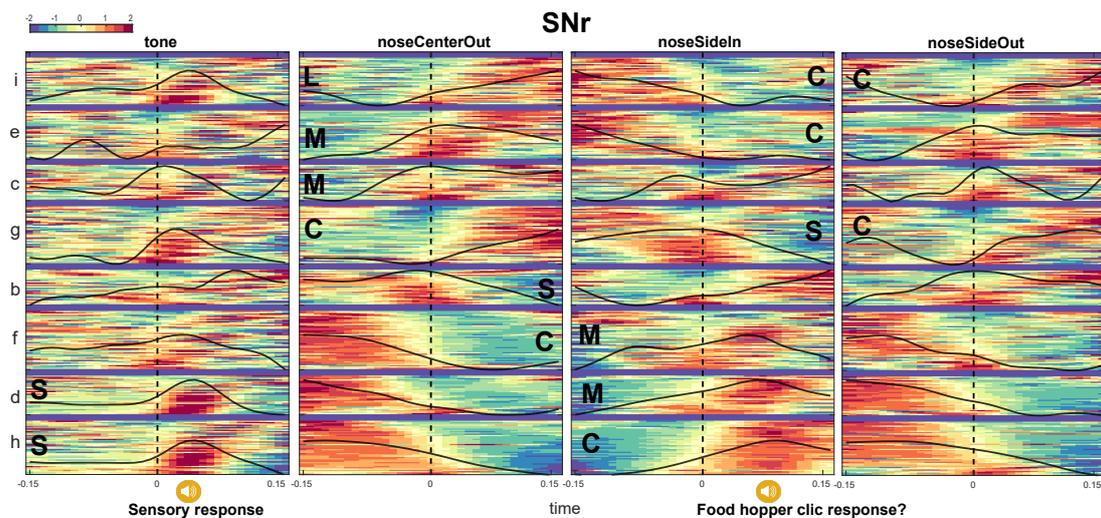


Figure 7.11: Catalan ensemble peri-event activity for go cue (tone), noseCenterOut (original movement studied), noseSideIn and noseSideOut. The ensemble activity is set up as in the previous figures. S: surround ensemble, C: Center ensemble, M: Mixed ensemble. L: Late movement component.

behavior. For example, ensemble “e” is a motor ensemble. It has low stimulus related modulation. Moreover, it forms part of the surround (excitation) for the “noseCenterOut” event but is part of the “center” for sideIn. Note the similarity of the firing patterns between noseCenterOut and noseSideOut.

Ensembles “d” and “h” are interesting because they have a very reliable modulation that comes after a sensory stimulation. That signal seems to be evoked again for the foodHopperClic event (reward delivery). Ensemble “g” is in the center for the noseCenterOut event but part of the surround for the noseSideIn. The ensemble “i” seems to be related entirely to the mysterious late component between the two recorded events, coming potentially from unrecorded but stereotypical movements. Finally, note that the ensembles are not completely stable across events. Sometimes they split, meaning that some neurons can be dynamically reassigned to different channels.

If this categorization is correct, we might be able to make some predictions. For example, the activity in (dh) should be independent of different aspects of movement (direction, speed or stopping). Also, differences in the speed and the outcome of the movement should be reflected in the dynamics of ensembles like (bf) or (ec). The fact that they are mostly disjoint, as reflected by the sharp turn in the trajectory, suggests that there should not be much cross talk between those computations.

In order to give a glimpse as to whether those predictions are correct, we studied the

activity of the ensembles during slow vs. fast trials and correct stop vs. failed stop conditions (figure 7.12). As expected, the ensembles showing major modulations with motor aspects of the behavior were (bf) and (ec). As seen from the correlation matrix, the patterns of activation of those ensembles are more correlated for slow go and correct stop trials, and between fast go and failed stop trials. This confirms and adds detail to the physiological similarities observed between those trial types in previous work [83].

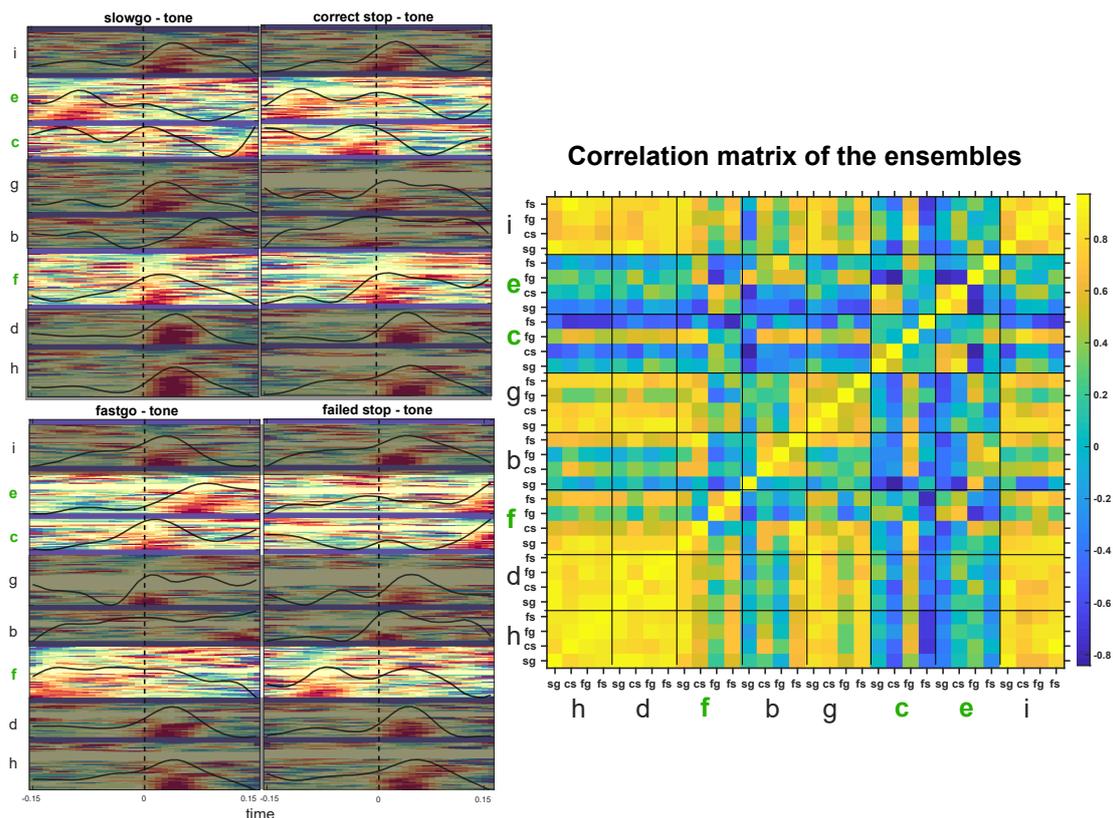


Figure 7.12: Comparison of the ensemble activity during slow go (sg), fast go (fg), correct stop (cs) and failed stop (fs) trials. The ensembles highlighted are the ones showing high modulations between conditions. The correlation matrix shows the par-wise correlation between the mean firing rates of the ensembles for the different trial types.

These observations allow us to classify, broadly, the response of the neurons in the substantia nigra in the following categories:

- Go process: Activity in (dh) is classified as a go-process channel because it

seems to carry sensory information but it starts late after other neurons have receive the actual sensory input.

- Sensory-motor: Ensembles (f) (g) and (i)
- Motor related - Ensembles (ec), (b).

### 7.3.4 The STN is mostly excitation with low dimensional patterns that reflect mixed selectivity

We now follow a similar procedure to study the nature of STN's sharp point. In order to understand this shape we locate the ensembles that are likely to be contributing by their increases or their decreases. The corresponding ensemble semiorde and coding trees are shown in figure 7.13. Following the chamber code we discover that the initial one dimensional evolution is driven almost exclusively by the increase (decrease) of the ensemble b (e). When the sharp point is reached the population changes then due to an increase (decrease) in c (d), generating a one dimensional evolution in an orthogonal direction towards (-+). Finally, after b (e) decreases (increases) below(above) zero, the trajectory make a slow turn towards (-).

This observations allows us to propose a phenomenological model as with the SNr 7.14. The main feature in this case is the ensemble (mixed) that have a wide response pattern that covers different movements (noseCenterOut) and (noseSideIn). This unspecific ensemble is enough to generate the main features of the trajectory and principal components for the STN. The fact that this ensemble shows nonspecific responses is termed *mixed selectivity*.

#### Functional significance of the two branches of the trajectories

The sharp point in the STN is also an irregular one. This is an indication of some unexplained activity happening in an additional dimension. In this case, the trajectory is split into two branches (figure 7.15). To understand this behavior, we extend our trees one more generation and plot the surrounding events' distributions. We split our analysis in two, one for each branch of the trajectory.

#### First branch

The first branch is composed of movement related increases in (bf) and decreases in (ie). The ensemble (e) is also responsive to the cue and behaves similarly to the ensemble (d) in the SNr.

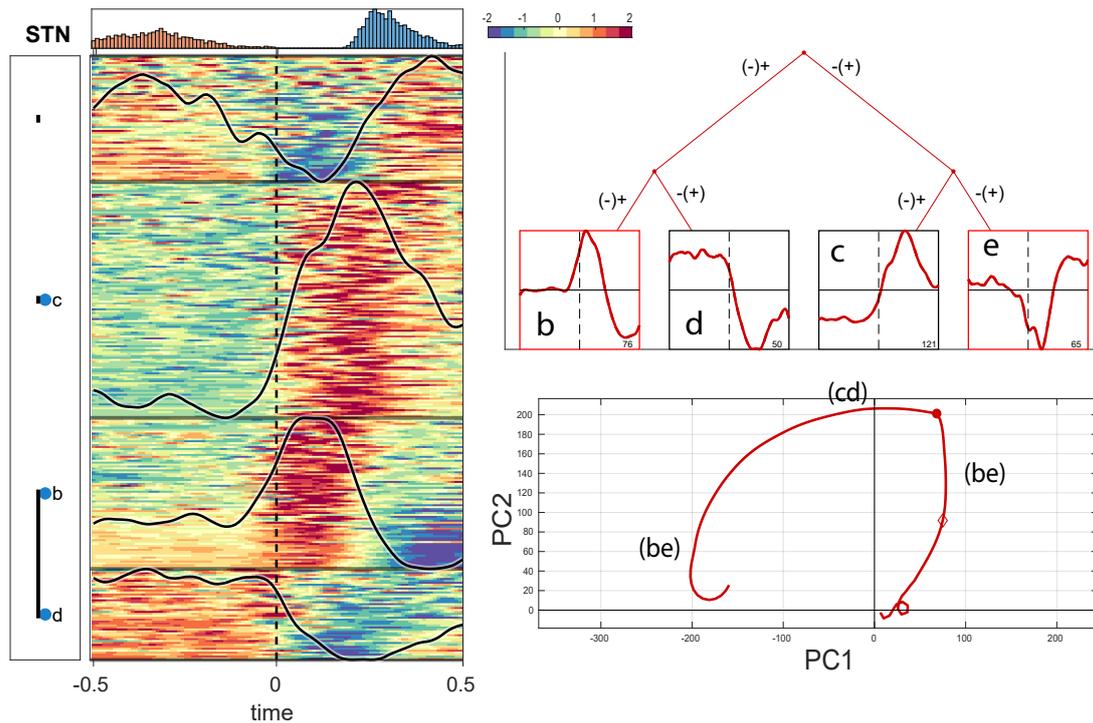


Figure 7.13: The shape of the STN and its quadrant location. The ensembles, located using the coding tree, are shown as before.

### Second branch

The second branch is characterized by increases (cg) and decreases (dh) after movement initiation. The increases in (c) are unspecific with respect to the action (noseCenterOut or noseSideIn). We hypothesize that this ensemble conveys signal that are independent of the identity of the action to be executed. For that reason we call that pattern as representing *vigor*.

Based upon these observations, we can categorize the activity of the STN as follows:

- Movement related: Increases, (bf) for noseCenterOut and decreases (ie).
- Cue related: Ensembles (dh)
- Vigour or unspecific responses (cg)

We conclude that the geometric features of this trajectory have different underlying population causes than the ones of the SNr. Nevertheless, the ensemble activity seems to be composed also of simple signals traveling through channels whose functional significance needs to be further studied.

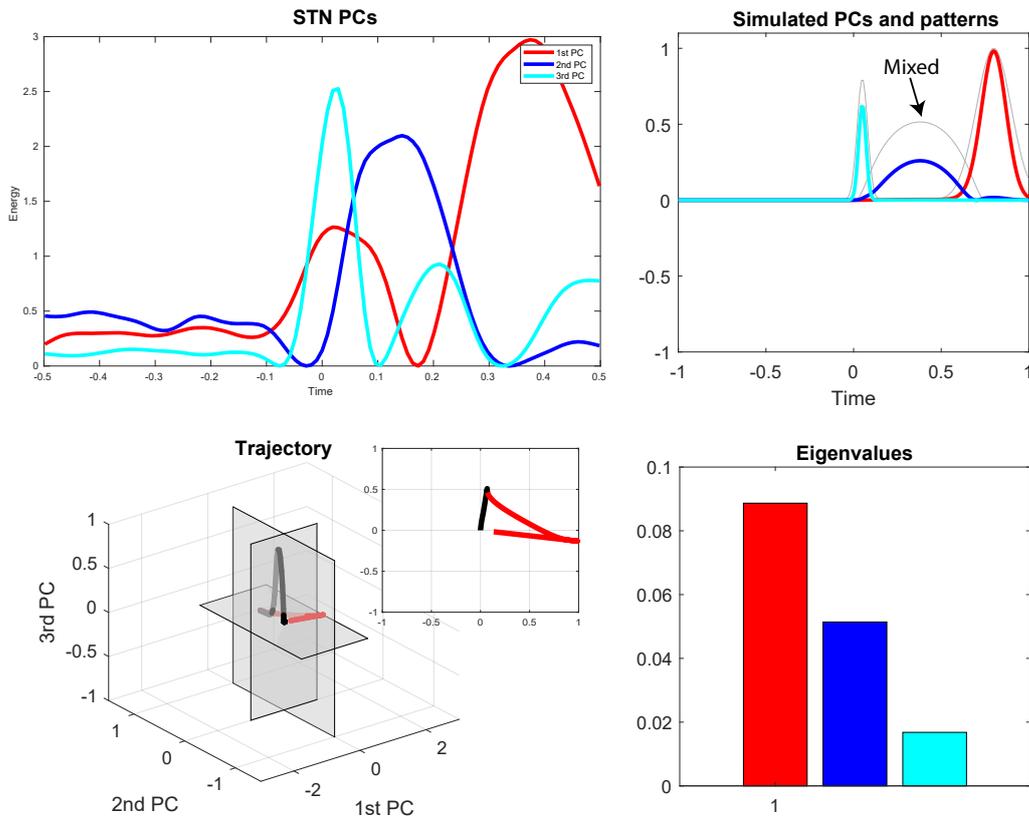


Figure 7.14: Phenomenological model for the STN. Three bumps are sufficient to generate a trajectory that evolves in orthogonal 2D planes. The layout is the same as in figure 7.8

## 7.4 Other regions and features

Each region of the basal ganglia has characteristic ensembles along with some of the observed patterns to different degrees. A detailed study of each of the population structures is beyond this work but its understanding is of paramount importance in order to explain the different geometric features of the population dynamics and their relation to behaviour.

For completeness, we provide a gallery of the ensemble structure for all the regions studied in this work.

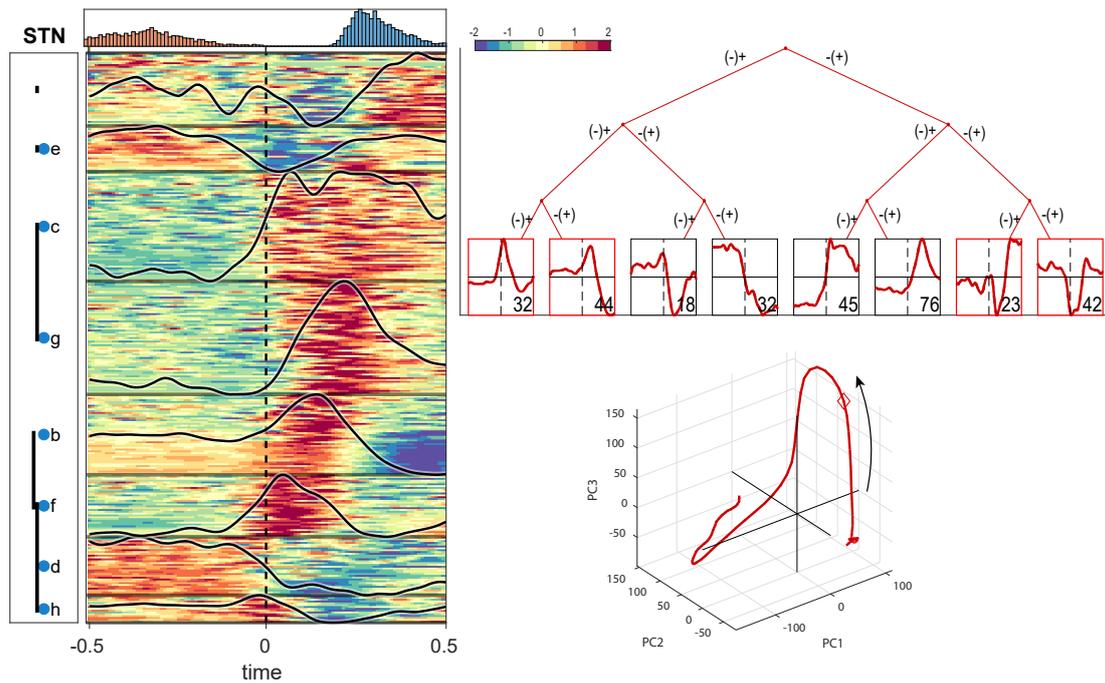


Figure 7.15: Level 3 ensemble semiorder for the STN. Same as with the SNr.

## 7.5 Appendix: The growing of the trees

The difference tree structure is mostly saturated for most regions

The trees associated to each region are shown in figure 7.16. These trees have been grown for 10 eigenvectors, as you can see, there is already a lot of structure and branching at this point. Remember that for each eigenvector, children are born at each node if there is still “unexplained variance” in that node, the fact that all the trees look similar reveals that there is enough structure in that population. Later ensembles (rightmost and deep) start showing differences.

It can be appreciated that, for diverse enough data, the trees are fractal. Figures 7.17 and 7.18; the more diverse, the more self-similar will be those structures as a consequence of the rules of growing. Those trees are therefore not as useful in providing information about region-specific broad patterns but as providing “boxes” in which to put neurons with similar patterns of firing rate. These boxes, as explained before are faces of the PCA hyperplane arrangement that the trajectory visits for some time.

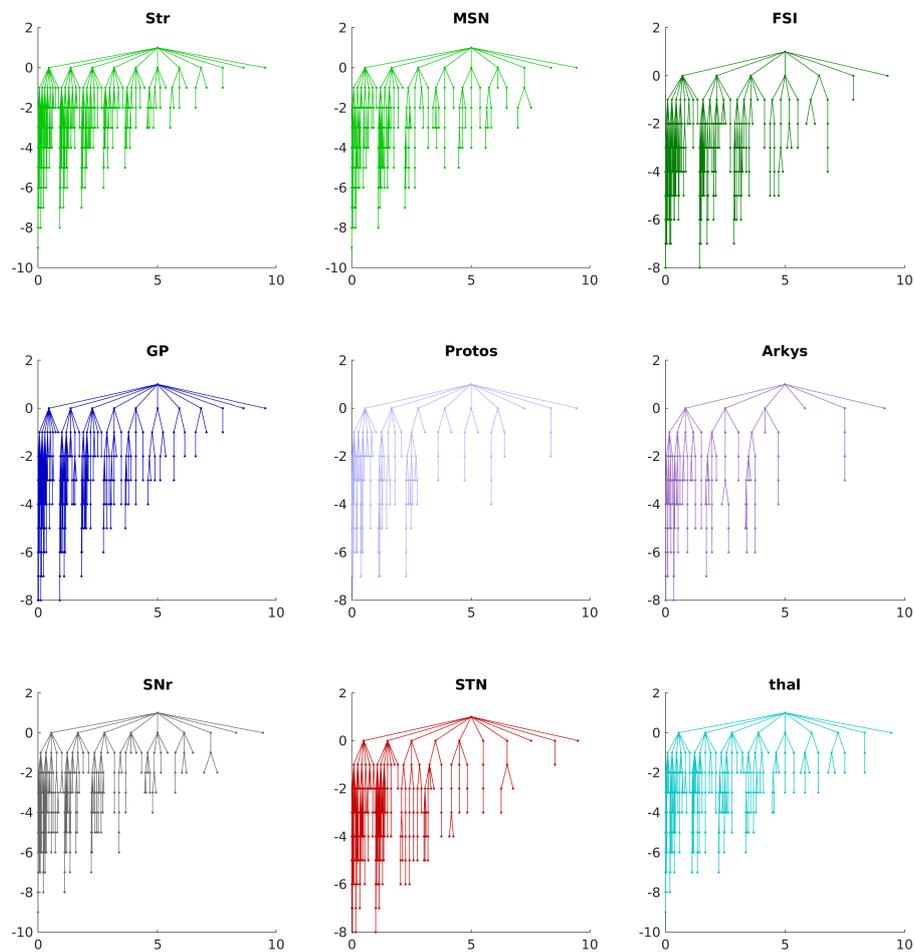


Figure 7.16: Trees for the first 10 eigenvectors (Level 10)

The tree growing reveals similar patterns in most regions related with the timing and breadth of the response

In non mean subtracted data, neurons with very low firing rates, which we call *null neurons*, are automatically left behind in the first step of growing. Geometrically, neurons with very low or zero firing rate lie in the *Frobenius hyperplane* or the hyperplane corresponding to the *all positive* first eigenvector. In mean subtracted data, the null neurons will be left in the ensemble with least explained variance which, as a matter of fact, will be always be the top most in the subsequent figures.

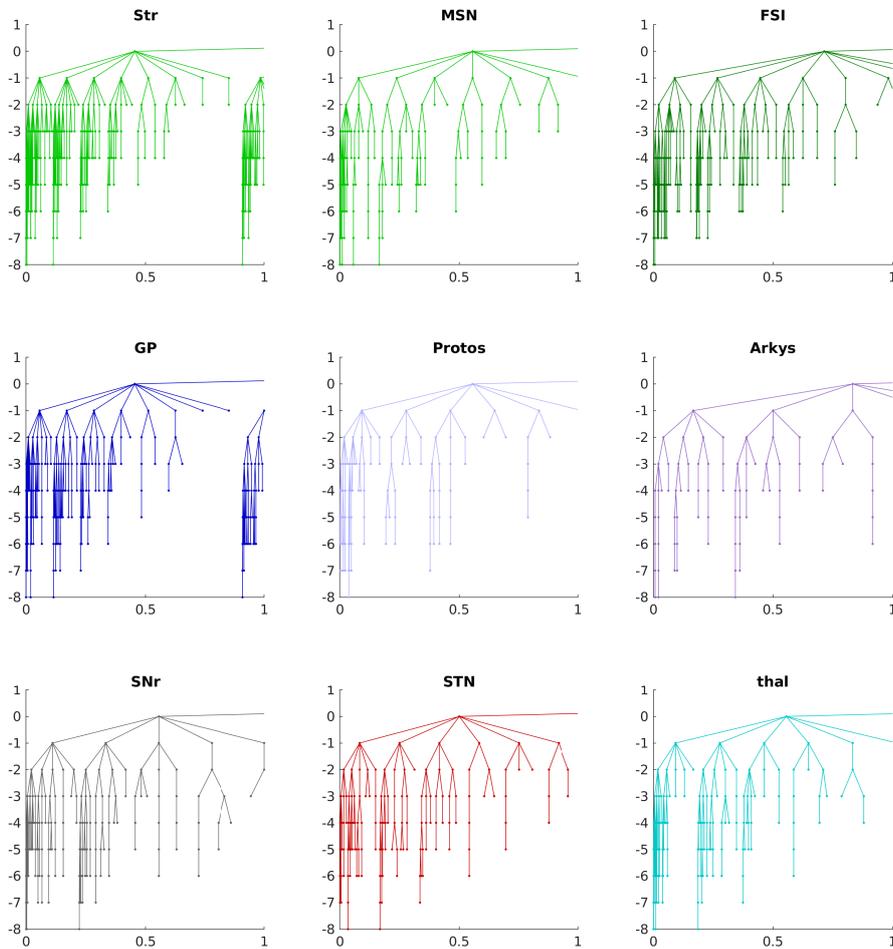


Figure 7.17: Same as figure 7.16 but summed in the leftmost branch

This feature is illustrated, for example, in figure 7.20.

This figure has a layout that will be repeated during this section. To the right we show the z-scored population activity, which we will always use from now on, as a heat map with one row per unit and grouped by ensembles (nodes of the tree). The mean firing rate of each ensemble is overlaid on top of the corresponding population activity. The time of movement initiation is marked by a dotted line in the middle and the interval around the alignment event is  $[-0.5, 0.5]$ .

Figures 7.19 to 7.27 show the evolution of the semiorder procedure introduced in the last section as it grows, fine graining the population's patterns. Taking the

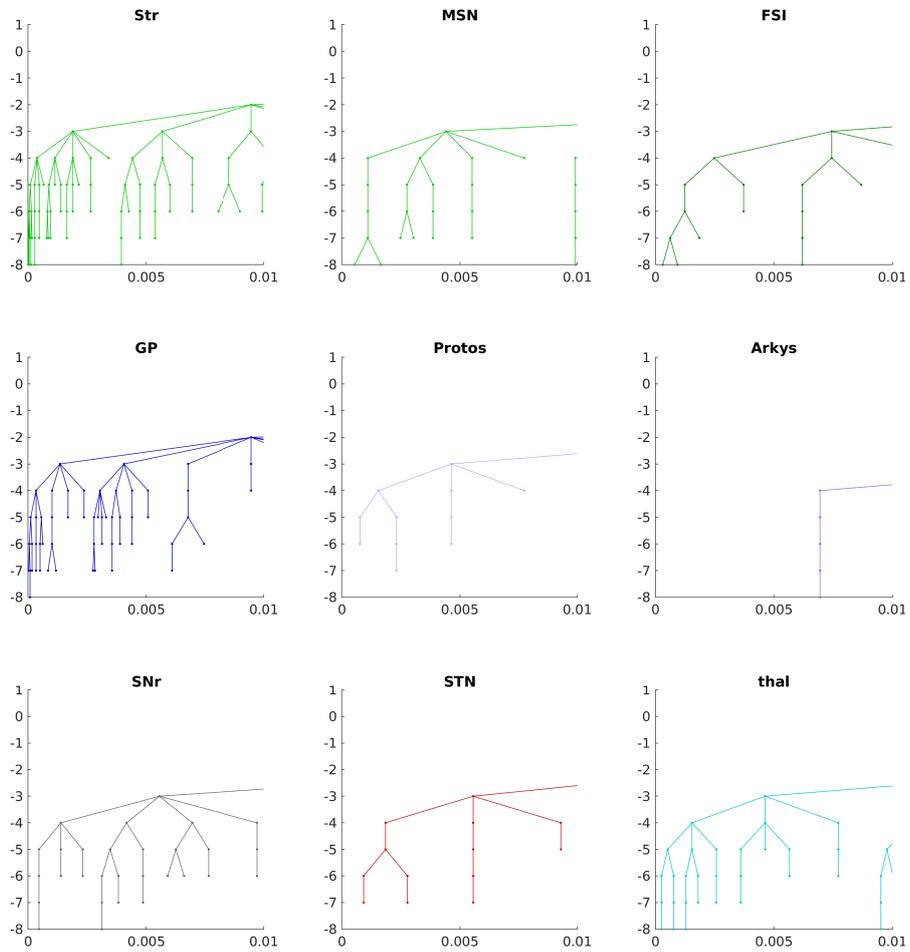


Figure 7.18: Same as figure 7.17 but summed in the leftmost branch

SNr as an example, the process of tree building structures the population in its natural ensembles based on life time covariance (co-firing in specific parts of the interval). Each step of the growing adds detail into the clustering. Also apparent from some of the figures, is the symmetry of regions like the SNr as revealed by the clusters. Patterns of increases are roughly followed by patterns of decreases.

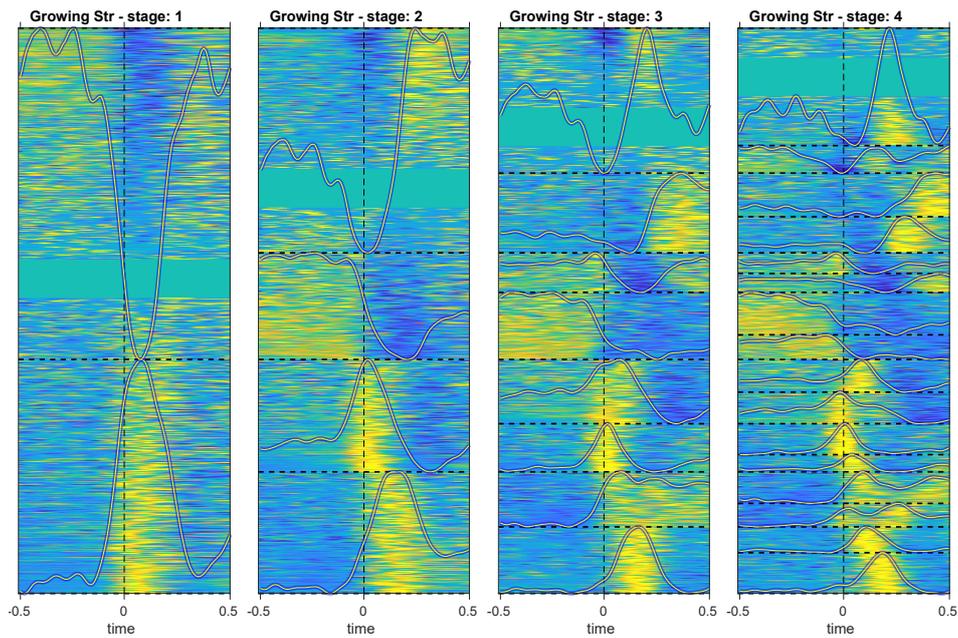


Figure 7.19: Structuring the Str population by growing the semiorder. Patterns with higher variance are in the bottom of the figure. The first step of growing and therefore the first principal component keeps track of broad increases and decreases in the firing rate. Those are the parallel subtrees mentioned before for mean subtracted data.

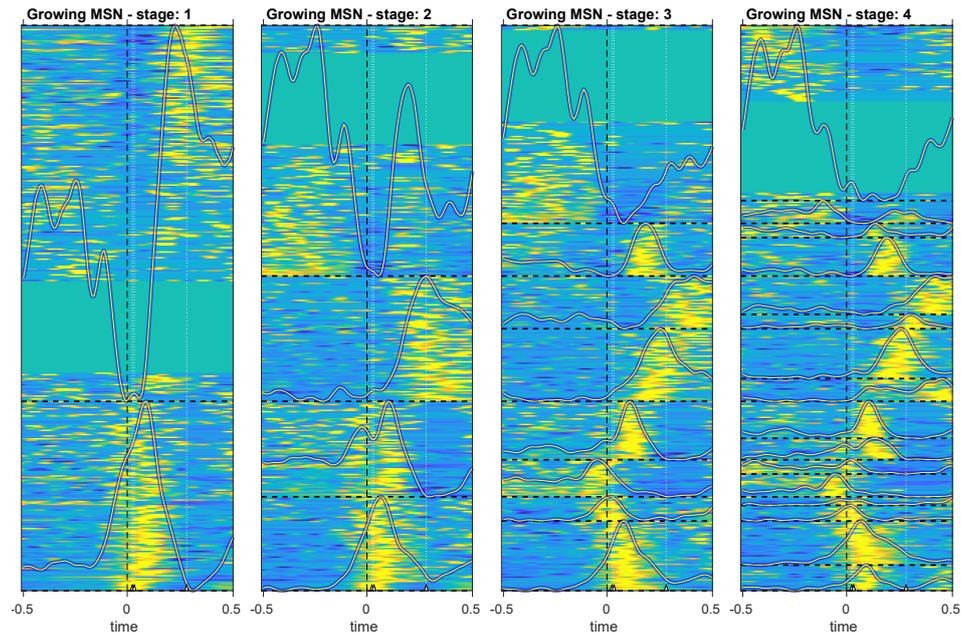


Figure 7.20: Structuring the MSN population by growing the semiorder. Again, patterns with higher variance are in the bottom of the figure. Note that the silent neurons are exclusively of this subpopulation (compared to the FSIs)

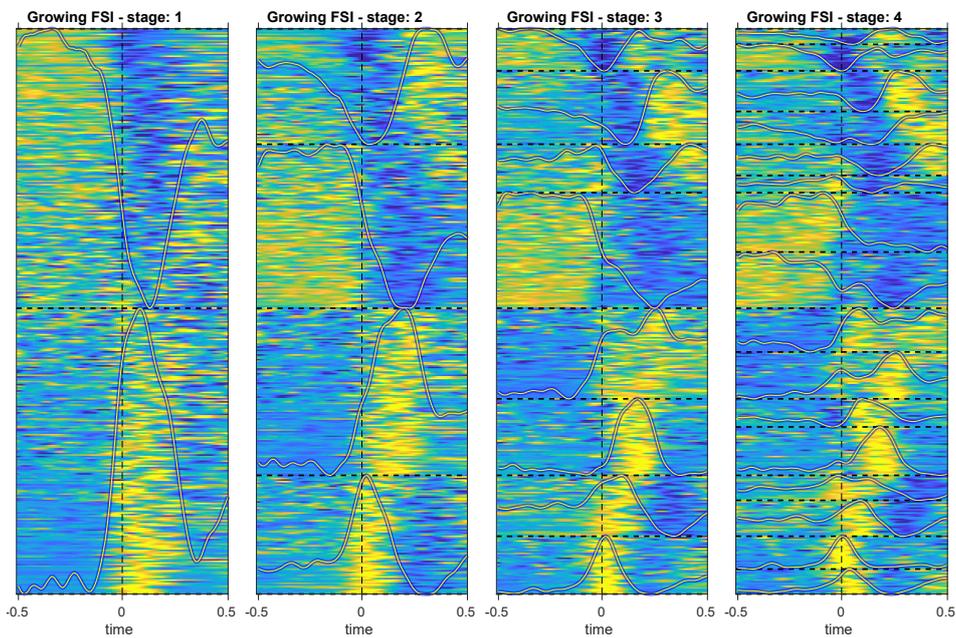


Figure 7.21: Structuring the FSI population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

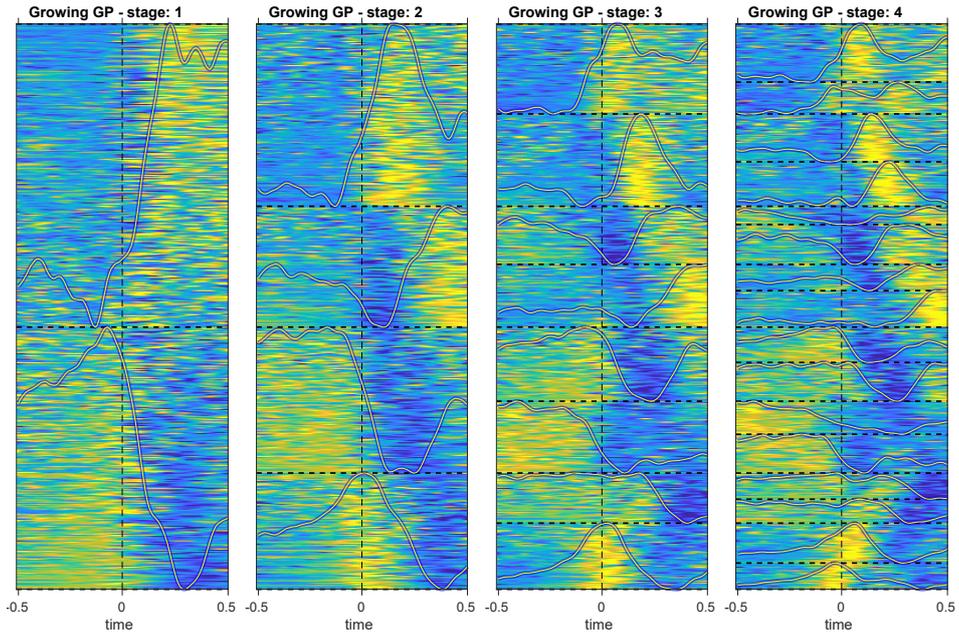


Figure 7.22: Structuring the GP population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

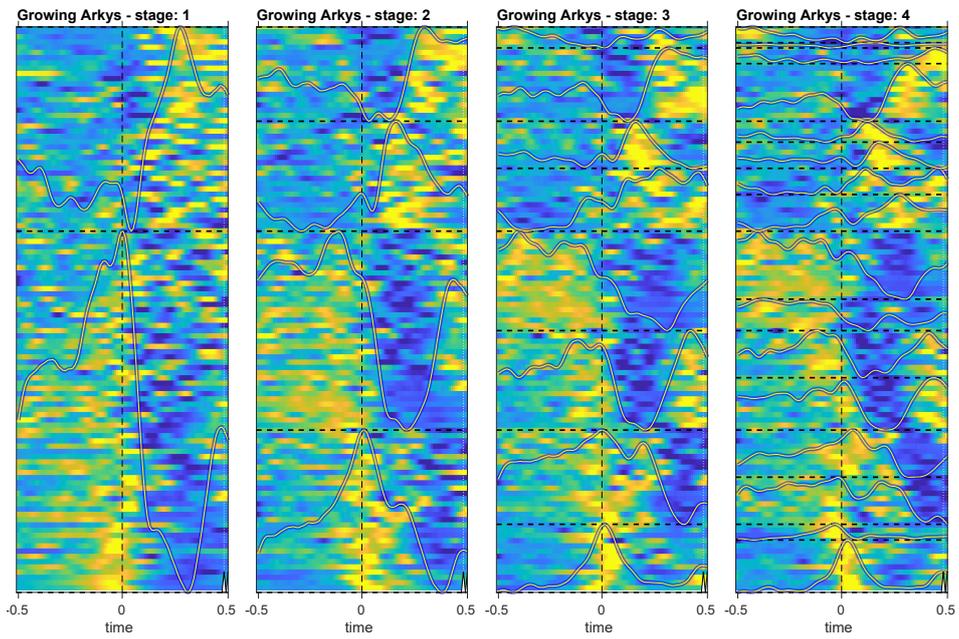


Figure 7.23: Structuring the Arky population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

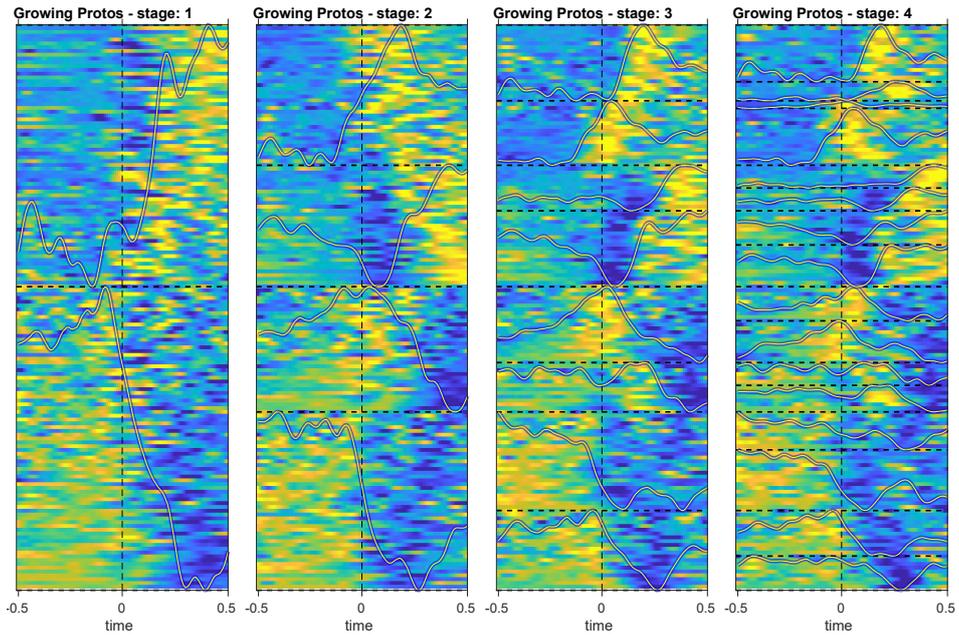


Figure 7.24: Structuring the Proto population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

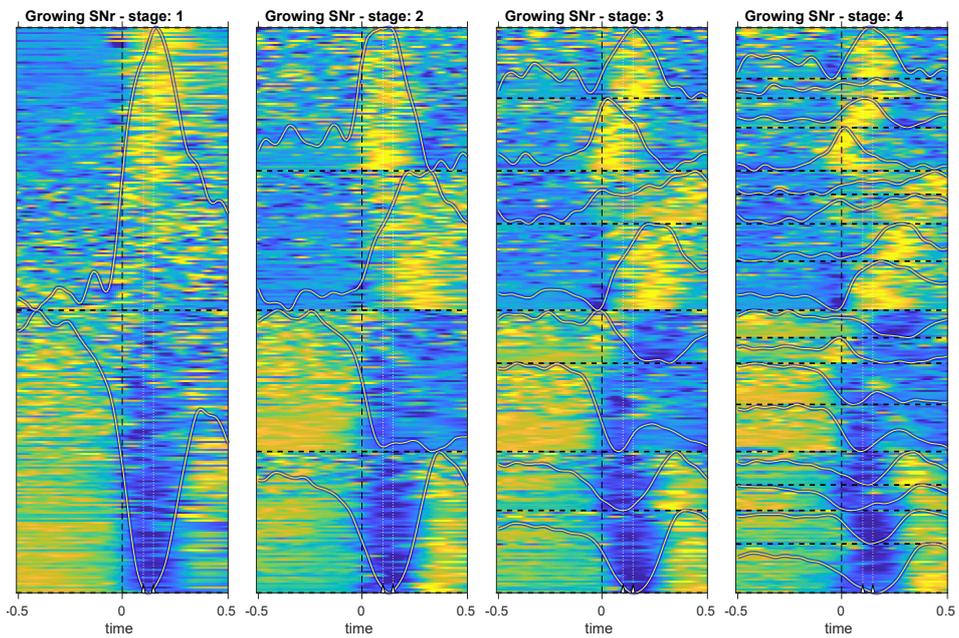


Figure 7.25: Structuring the SNr population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

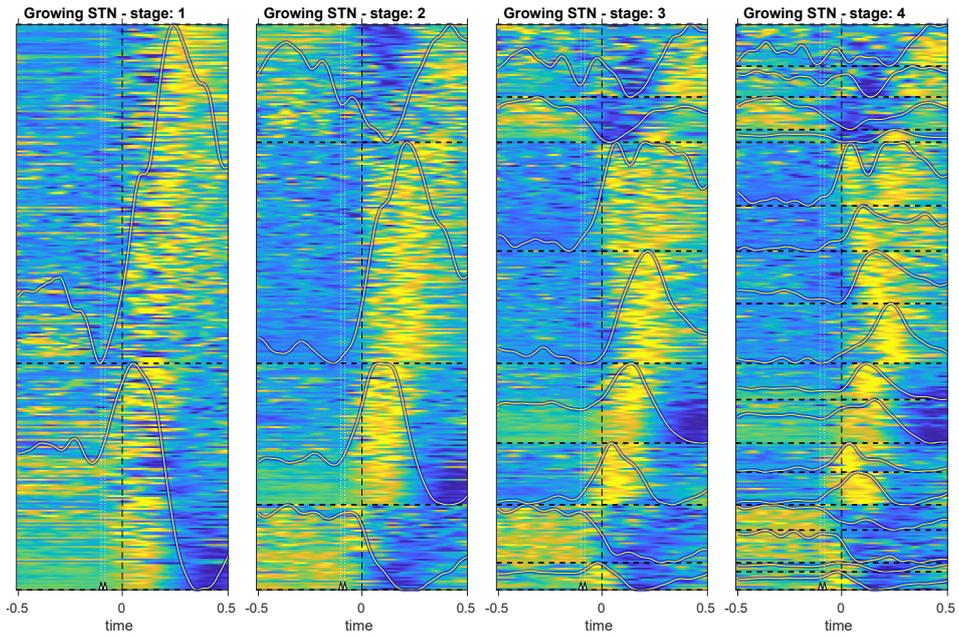


Figure 7.26: Structuring the STN population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

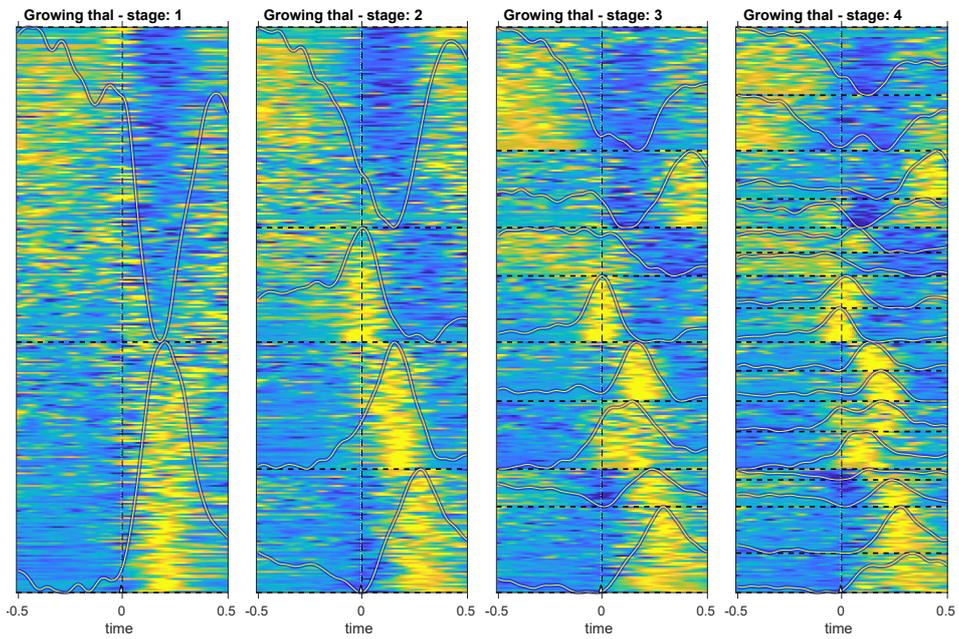


Figure 7.27: Structuring the Pf population by growing the semiorder. Patterns with higher variance are in the bottom of the figure.

# Chapter 8

## Discussion

In this work we developed a systematic approach to the study of the low dimensional neural trajectories obtained from dimensionality reduction by PCA. The aspect of shape was approached rigorously from two perspectives. The invariance approach focused on the problem of comparison of shapes; the singularity one in the specific features of each shape like sharp points, vertices and flat regions. We performed a novel study of the curvature, the most important parameter of a curve, in terms of the coordinates that helps to understand how the overlaps of covariance between the single coordinates generates different curvature profiles. Finally we developed for the first time to the knowledge of the authors a rigorous account on the restrictions imposed by PCA on the shape.

The analysis of the “form” put on a rigorous ground observations about the shapes of the trajectories associated to each basal ganglia regions. For example, the fact that some regions have a rounder form while other have polygonal or hook like shapes with sharp turns. It was found that, in a particular shape space devised in that section, the trajectories cluster in two large groups that are related to the presence of absence of sharp points. In the construction of the shape space, one important aspect was factored out without a fully exposition of its origin: reflections.

It is well known that the principal components are ambiguous in sign, for that reason, all possible reflections of a trajectory represent the same shape. This aspect of PCA has a deeper meaning that the author unfortunately had to postpone for future work. An observation about the behavior of the reflections sheds light about their possible causes. If the neurons are randomly permuted in the population, the trajectory is reflected in one of the many axes in the high dimensional space. It turns out that the specific reflection coming from a particular permutation can be understood by the action of the symmetric group on the *induced hyperplane arrangement* [11]. A rigorous treatment of the reflections will appear somewhere else.

It is possible that the shape space developed in chapter 5 could be improved in 2 directions. Firstly, as seen in figure 8.1, the trajectory can be thought as a curve in the space of distributions of firing rates or activity in the neurons. Under this model, the geometry becomes an information geometry. Some strides in that direction were taken by the author but a full development requires more work. Secondly, the natural space in which the PCA acts is called a flag manifold[30]. The study of the trajectories in the flag manifold is also an ongoing work.

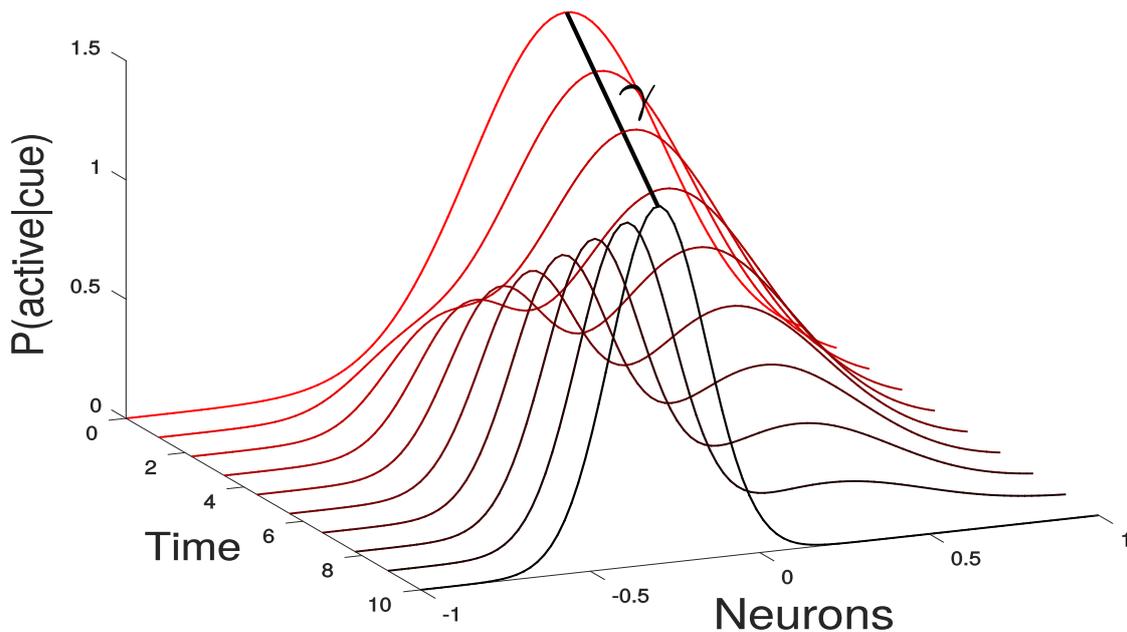


Figure 8.1: Information geometric space for the population trajectories. The trajectory  $\gamma$  can be thought as the parametrization of a curve in the space of distributions of activity in the neurons

The ensemble approach was developed with the natural geometry of the flag manifold in mind. Even if it was not complete, the results of this work are a step in the right direction for a complete understanding of the geometry of the trajectories in those spaces. As was shown, the ensembles are the main drives of the shape; indeed, replacing the population by the averages of the ensembles from the tree preserves the shape as expected. A big part of the effort was put in developing the method, and understanding and making clear the intended relationships between the components of the eigenvectors and the Catalan objects.

The bijection and the hyperplane approach bring about theoretical benefits that can be exploited in future works. It is conceivable that such approach will help in the understanding of how the trajectory is expected to change by adding more neurons or extending the time interval around movement. It also puts the

problem in a combinatorial setting and allows the application of important theorem developed in that field. For example, each semiorder is in one to one relation with a high dimensional polytope called the associahedron that encodes the possible ways in which parenthesis can be put in a sequence of symbols; the implications of such relationships have yet to be developed.

The ensemble structure can be used to generate “region -like” activity around movement to feed computational models input. Conversely, it interesting to see if existing computational models of the basal ganglia generate a similar gamut of patterns and trajectories.

An obvious next step for this work is the analysis of the shapes related to other events in the experimental setup. Of particular interest is the activity around stopping. What would be the population signal related to stopping? Would it be restricted to a single ensemble or is it a concerted population activity. In this sense, correlates of the go and stop processes could be found in the evolution of the trajectories and the elusive “point of no return” might identified by such analysis.

With respect to the current analysis of the population activity during movement onset, the change in sparseness/heterogeneity around movement observed in the SNr and STN bears some resemblance with the quenching of variability around stimulus onset observed in the cortex [20]. It is interesting that such phenomenon is not observed in all the regions. It might be related to the selection of the action and suppression of undesired ones. In this sense, the “dense” striatum is a bit paradoxical; in general, this regions is expected to be sparse, however, the multiplicity of variables represented in such region (see chapter 2) justifies the presence of such a variety of patterns.

Indeed, the observed pattern derived from the sharp points fits Nambu’s center surround model presented in chapter 2. The unspecified signal found in the STN comes earlier than the two branches of the SNr and presents two peaks that could be related to the two step computations predicted by this model. Moreover, the independence of the patterns that gives origin to the sharp point in the SNr can be interpreted as a form of decorrelation or compression of the high dimensional activity observed in the striatum fitting in turn the dimensionality reduction model of computation in the basal ganglia.

The conceptualization of activity in regions, in the basal ganglia in general and in the SNr and STN in particular, as coming through different regions is not new [66]. Indeed, signals like go/stop/pause [82], urgency [91] or vigor [92] are part of the observed features of this network. In the task concerning this study, in which a highly trained, stereotypical behavior is performed, it makes sense to have different ensembles with orthogonal patterns of activation dedicated to each step in the sequence. If that was not the case, activity in the shared neurons in the ensemble could be wrongly activate undesired movements in down stream regions.

This methodology has shown to be instrumental in the understanding of the coding schemes in the basal ganglia. Further analysis are guaranteed to shed light into the details of the code and how such a network can have such parallel, low dimensional channels embedded in its dynamics and architecture.

# Chapter 9

## Code

All the code is on github under the following projects:

### 9.1 Population api

The code for generating the trees and ensemble structure from the PCA is made public under the repository:

<https://github.com/aljiro/PopulationAPI>

### 9.2 Curve api

The code for the analysis of the trajectories, reparametrization and differential geometric operation is made public under the repository:

<https://github.com/aljiro/CurveAPI>

### 9.3 TameD library

During this thesis an ongoing project for the analysis the neural data was started to support the performed analyses. The code can be found and followed at:

<https://github.com/aljiro/TameD>

### 9.4 Analyses

The analyses and the code for each of the figures can be found at:

<https://github.com/aljiro/AnalysesPhDThesis>

# Bibliography

- [1] Larry F Abbott and Peter Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.
- [2] Marcelo Aguiar and Swapneel Mahajan. *Topics in Hyperplane Arrangements*, volume 226. American Mathematical Soc., 2017.
- [3] Xavier Álvarez-Vizoso, Robert Arn, Bruce Draper, Michael Kirby, and Chris Peterson. Geometry of curves in  $\mathbb{R}^n$ , singular value decomposition, and hankel determinants. *arXiv preprint arXiv:1511.05008*, 2015.
- [4] Andrea Amadei, Antonius BM Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412–425, 1993.
- [5] Bruno B Averbeck and Daeyeol Lee. Coding and transmission of information by neural ensembles. *Trends in neurosciences*, 27(4):225–230, 2004.
- [6] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.
- [7] Izhar Bar-Gad, Genela Morris, and Hagai Bergman. Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in neurobiology*, 71(6):439–473, 2003.
- [8] Horace Barlow. -cell assemblies versus single cells. In *Brain Theory*, pages 261–267. Elsevier, 1996.
- [9] Daniel Beylkin and Vladimir Rokhlin. Fitting a bandlimited curve to points in a plane. *SIAM Journal on Scientific Computing*, 36(3):A1048–A1070, 2014.
- [10] James R Booth, Lydia Wood, Dong Lu, James C Houk, and Tali Bitan. The role of the basal ganglia and cerebellum in language processing. *Brain research*, 1133:136–144, 2007.

- [11] Alexandre V Borovik and Anna Borovik. *Mirrors and reflections*. Springer, 2010.
- [12] Hervé Bourlard and Yves Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics*, 59(4-5):291–294, 1988.
- [13] Kevin L Briggman, Henry DI Abarbanel, and William B Kristan. Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901, 2005.
- [14] James William Bruce, James William Bruce, and PJ Giblin. *Curves and Singularities: a geometrical introduction to singularity theory*. Cambridge university press, 1992.
- [15] Daniel A Butts and Mark S Goldman. Tuning curves, neuronal variability, and sensory coding. *PLoS biology*, 4(4):e92, 2006.
- [16] György Buzsáki. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68(3):362–385, 2010.
- [17] John K Chapin and Miguel AL Nicolelis. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *Journal of neuroscience methods*, 94(1):121–140, 1999.
- [18] Anindya Chatterjee. An introduction to the proper orthogonal decomposition. *Current science*, pages 808–817, 2000.
- [19] G Chevalier and JM Deniau. Disinhibition as a basic process in the expression of striatal functions. *Trends in neurosciences*, 13(7):277–280, 1990.
- [20] Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369, 2010.
- [21] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487(7405):51, 2012.
- [22] Ronald R Coifman and M Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on information theory*, 38(2):713–718, 1992.

- [23] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
- [24] John P Cunningham and M Yu Byron. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, 17(11):1500, 2014.
- [25] Carina Curto, Vladimir Itskov, Katherine Morrison, Zachary Roth, and Judy L Walker. Combinatorial neural codes from a mathematical coding theory perspective. *Neural computation*, 25(7):1891–1925, 2013.
- [26] Marc Deffains, Liliya Iskhakova, Shiran Katabi, Suzanne N Haber, Zvi Israel, and Hagai Bergman. Subthalamic, not striatal, activity correlates with basal ganglia downstream activity in normal and parkinsonian monkeys. *Elife*, 5:e16443, 2016.
- [27] Chris Ding and Xiaofeng He. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, page 29. ACM, 2004.
- [28] David L Donoho et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000):32, 2000.
- [29] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.
- [30] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [31] Gamaleldin F Elsayed and John P Cunningham. Structure in neural population recordings: an expected byproduct of simpler phenomena? *Nature neuroscience*, 20(9):1310, 2017.
- [32] Richard P Feynman, Robert B Leighton, and Matthew Sands. *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, volume 1. Basic books, 2011.
- [33] Gregory J Gage, Colin R Stoetzner, Alexander B Wiltschko, and Joshua D Berke. Selective activation of striatal fast-spiking interneurons during choice execution. *Neuron*, 67(3):466–479, 2010.
- [34] Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155, 2015.

- [35] Apostolos P Georgopoulos, Andrew B Schwartz, and Ronald E Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–1419, 1986.
- [36] Ann M Graybiel. The basal ganglia and chunking of action repertoires. *Neurobiology of learning and memory*, 70(1-2):119–136, 1998.
- [37] Brian Guenter and Richard Parent. Motion control: computing the arc length of parametric curves. *IEEE Computer Graphics and Applications*, (3):72–78, 1990.
- [38] HW Guggenheimer. Differential geometry. 1963, 1977.
- [39] Kevin Gurney, Tony J Prescott, and Peter Redgrave. A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, 84(6):401–410, 2001.
- [40] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.
- [41] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
- [42] Xin Jin, Fatuel Tecuapetla, and Rui M Costa. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. *Nature neuroscience*, 17(3):423, 2014.
- [43] Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.
- [44] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- [45] Joseph K Jun, Paul Miller, Adrián Hernández, Antonio Zainos, Luis Lemus, Carlos D Brody, and Ranulfo Romo. Heterogenous population coding of a short-term memory and decision task. *Journal of Neuroscience*, 30(3):916–929, 2010.
- [46] Robert E Kass, Valérie Ventura, and Can Cai. Statistical smoothing of neuronal data. *Network-Computation in Neural Systems*, 14(1):5–16, 2003.
- [47] Saul Kato, Harris S Kaplan, Tina Schrödel, Susanne Skora, Theodore H Lindsay, Eviatar Yemini, Shawn Lockery, and Manuel Zimmer. Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell*, 163(3):656–669, 2015.

- [48] Matthew T Kaufman, Mark M Churchland, Stephen I Ryu, and Krishna V Shenoy. Cortical activity in the null space: permitting preparation without movement. *Nature neuroscience*, 17(3):440, 2014.
- [49] Andreas Klaus, Gabriela J Martins, Vitor B Paixao, Pengcheng Zhou, Liam Paninski, and Rui M Costa. The spatiotemporal organization of the striatum encodes action space. *Neuron*, 95(5):1171–1180, 2017.
- [50] Dmitry Kobak, Wieland Brendel, Christos Constantinidis, Claudia E Feierstein, Adam Kepecs, Zachary F Mainen, Xue-Lian Qi, Ranulfo Romo, Naoshige Uchida, and Christian K Machens. Demixed principal component analysis of neural population data. *Elife*, 5:e10989, 2016.
- [51] Grigory Kolesov, Dmitry Vinichenko, Georgios A Tritsaridis, Cynthia M Friend, and Efthimios Kaxiras. Anatomy of the photochemical reaction: Excited-state dynamics reveals the c–h acidity mechanism of methoxy photo-oxidation on titania. *The journal of physical chemistry letters*, 6(9):1624–1627, 2015.
- [52] Ann B Lee and Boaz Nadler. Treelets— a tool for dimensionality reduction and multi-scale analysis of unstructured data. In *Artificial Intelligence and Statistics*, pages 259–266, 2007.
- [53] Daniel K Leventhal, Gregory J Gage, Robert Schmidt, Jeffrey R Pettibone, Alaina C Case, and Joshua D Berke. Basal ganglia beta oscillations accompany cue utilization. *Neuron*, 73(3):523–536, 2012.
- [54] Chung-Chuan Lo and Xiao-Jing Wang. Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature neuroscience*, 9(7):956, 2006.
- [55] Vítor Lopes-dos Santos, Sergio Conde-Ocazonez, Miguel AL Nicolelis, Sidarta T Ribeiro, and Adriano BL Tort. Neuronal assembly detection and cell membership specification by principal component analysis. *PloS one*, 6(6):e20996, 2011.
- [56] Mel MacMahon and Diego Garlaschelli. Community detection for correlation matrices. *arXiv preprint arXiv:1311.1924*, 2013.
- [57] Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.
- [58] Ofer Mazor and Gilles Laurent. Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, 48(4):661–673, 2005.

- [59] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.
- [60] Jonathan W Mink. The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, 50(4):381–425, 1996.
- [61] Jonathan W Mink. The basal ganglia and involuntary movements: impaired inhibition of competing motor patterns. *Archives of neurology*, 60(10):1365–1368, 2003.
- [62] Amin Mirzaei, Arvind Kumar, Daniel Leventhal, Nicolas Mallet, Ad Aertsen, Joshua Berke, and Robert Schmidt. Sensorimotor processing in the basal ganglia leads to transient beta oscillations during behavior. *Journal of Neuroscience*, pages 1289–17, 2017.
- [63] Jack G Modell, James M Mountz, George C Curtis, and John F Greden. Neuropsychologic dysfunction in basal ganglia/limbic striatal and thalamocortical circuits as a pathogenetic mechanism of obsessive-compulsive disorder. *The Journal of neuropsychiatry and clinical neurosciences*, 1989.
- [64] Genela Morris, Alon Nevet, and Hagai Bergman. Anatomical funneling, sparse connectivity and redundancy reduction in the neural networks of the basal ganglia. *Journal of Physiology-Paris*, 97(4-6):581–589, 2003.
- [65] Fionn Murtagh. The remarkable simplicity of very high dimensional data: application of model-based clustering. *Journal of classification*, 26(3):249–277, 2009.
- [66] Atsushi Nambu, Hironobu Tokuno, and Masahiko Takada. Functional significance of the cortico–subthalamo–pallidal ‘hyperdirect’ pathway. *Neuroscience research*, 43(2):111–117, 2002.
- [67] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607, 1996.
- [68] Norman H Packard, James P Crutchfield, J Doyne Farmer, and Robert S Shaw. Geometry from a time series. *Physical review letters*, 45(9):712, 1980.
- [69] GP Patil and Charles Taillie. Diversity as a concept and its measurement. *Journal of the American statistical Association*, 77(379):548–561, 1982.
- [70] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

- [71] Cengiz Pehlevan and Haim Sompolinsky. Selectivity and sparseness in randomly connected balanced networks. *PLoS One*, 9(2):e89992, 2014.
- [72] Adam Ponzi and Jeff Wickens. Sequentially switching cell assemblies in random inhibitory networks of spiking neurons in the striatum. *Journal of Neuroscience*, 30(17):5894–5911, 2010.
- [73] Rodrigo Quian Quiroga and Stefano Panzeri. *Principles of neural coding*. CRC Press, 2013.
- [74] C Radhakrishna Rao. Convexity properties of entropy functions and analysis of diversity. *Lecture Notes-Monograph Series*, pages 68–77, 1984.
- [75] P Redgrave, N Vautrelle, and JNJ Reynolds. Functional properties of the basal ganglia’s re-entrant loop architecture: selection and reinforcement. *Neuroscience*, 198:138–151, 2011.
- [76] Peter Redgrave, Tony J Prescott, and Kevin Gurney. The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4):1009–1023, 1999.
- [77] Peter Redgrave, Manuel Rodriguez, Yoland Smith, Maria C Rodriguez-Oroz, Stephane Lehericy, Hagai Bergman, Yves Agid, Mahlon R DeLong, and Jose A Obeso. Goal-directed and habitual control in the basal ganglia: implications for parkinson’s disease. *Nature Reviews Neuroscience*, 11(11):760, 2010.
- [78] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585, 2013.
- [79] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [80] Piyatida Rujasiri and Boonorm Chomtee. Comparison of clustering techniques for cluster analysis. *Nat. Sci*, 43:378–388, 2009.
- [81] Shreya Saxena and John P Cunningham. Towards the neural population doctrine. *Current opinion in neurobiology*, 55:103–111, 2019.
- [82] Robert Schmidt and Joshua D Berke. A pause-then-cancel model of stopping: evidence from basal ganglia neurophysiology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1718):20160202, 2017.

- [83] Robert Schmidt, Daniel K Leventhal, Nicolas Mallet, Fujun Chen, and Joshua D Berke. Canceling actions involves a race between basal ganglia pathways. *Nature neuroscience*, 16(8):1118, 2013.
- [84] Maoz Shamir and Haim Sompolinsky. Implications of neuronal diversity on population coding. *Neural computation*, 18(8):1951–1986, 2006.
- [85] Gordon M Shepherd and Sten Grillner. *Handbook of brain microcircuits*. Oxford University Press, 2018.
- [86] Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*. Springer, 2016.
- [87] Richard P Stanley. *Catalan numbers*. Cambridge University Press, 2015.
- [88] Tetsuya Takahashi. Principal component analysis is a group action of  $so(n)$  which minimizes an entropy function. In *Neural Networks, 1993. IJCNN'93-Nagoya. Proceedings of 1993 International Joint Conference on*, volume 1, pages 355–358. IEEE, 1993.
- [89] Tetsuya Takahashi. Two statistical methods for analyzing multiple neuronal data. *International journal of neuroscience*, 88(1-2):11–26, 1996.
- [90] René Thom. *Structural stability and morphogenesis*. CRC Press, 2018.
- [91] David Thura and Paul Cisek. The basal ganglia do not select reach targets but control the urgency of commitment. *Neuron*, 95(5):1160–1170, 2017.
- [92] Robert S Turner and Michel Desmurget. Basal ganglia contributions to motor control: a vigorous tutor. *Current opinion in neurobiology*, 20(6):704–716, 2010.
- [93] DEKN UTH. The art of computer programming, vol 1: Fundamental algorithms, 1973.
- [94] Frederick Verbruggen and Gordon D Logan. Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience & Biobehavioral Reviews*, 33(5):647–661, 2009.
- [95] Alex H Williams, Tony Hyun Kim, Forea Wang, Saurabh Vyas, Stephen I Ryu, Krishna V Shenoy, Mark Schnitzer, Tamara G Kolda, and Surya Ganguli. Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, 2018.

- [96] Benjamin Willmore and David J Tolhurst. Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, 12(3):255–270, 2001.
- [97] Charles J Wilson. Active decorrelation in the basal ganglia. *Neuroscience*, 250:467–482, 2013.
- [98] RL Wine and John E Freund. On the enumeration of decision patterns involving  $n$  means. *The Annals of Mathematical Statistics*, 28(1):256–259, 1957.
- [99] Henry H Yin and Barbara J Knowlton. The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6):464, 2006.