Lecture Notes in Morphogenesis
Series Editor: Alessandro Sarti

Richard A. Clement

Mathematical Tools for Neuroscience

A Geometric Approach

Springer

Lecture Notes in Morphogenesis

Lecture Notes in Morphogenesis is an interdisciplinary book series which aims to face the questions of emergence, individuation and becoming of forms from several different points of view: those of pure and applied mathematics, of computational algorithms, of biology, of neurophysiology, of cognitive and social structures. The set of questions above concerns all the manifestations of Being, all the manifestations of Life. At the heart of contemporary embryogenesis lies an essential question: How can form emerge from the constant, chaotic flow? How can a sequence of purely informational elements — an a-signifying combination of chemical substances organized in the DNA molecule — evolve into the highly complex and structured forms of the living organism? A similar question can be asked when we deal with the morphogenesis of vision in neural systems and with the creation of evolving synthetic images, since digital technology makes possible the simulation of emergent processes both of living bodies and of visual forms. Finally the very idea that abstract structures of meaning could be captured in terms of morphodynamic evolution opens the door to new models of semiolinguistics, semiotic morphodynamics, and cognitive grammars. An entire heritage of ideas and concepts has to be reconsidered in order to face new and challenging problems: the theoretical framework opened by Goethe with the introduction of the word "Morphogenesis" is developed by D'Arcy Thompson in "On Growth and Form", it is reorganized with new theoretical insights by the classical structuralism of Levi-Strauss and formalized by the dynamical structuralism of René Thom. The introduction of the post-structuralists ideas of individuation (in Gilbert Simondon and Gilles Deleuze) and plasticity of structures builds a bridge to contemporary problems of morphogenesis at a physical, biological, social and transindividual level. The objective of this book series is to provide suitable theoretical and practical tools for describing evolutionary phenomena at the level of Free boundary problems in Mathematics, Embryogenesis, Image Evolution in Visual Perception, Visual Models of Morphogenesis, Neuromathematics, Autonomy and Self-Organization, Morphogenetic Emergence and Individuation, Theoretical Biology, Cognitive Morphodynamics, Cities Evolution, Semiotics, Subjectivation processes, Social movements as well as new frontiers of Aesthetics. To submit a proposal or request further information, please use the PDF Proposal Form or contact directly: *Dr. Thomas Ditzinger (thomas.ditzinger@springer.com)*

More information about this series at https://link.springer.com/bookseries/11247

Richard A. Clement

Mathematical Tools for Neuroscience

A Geometric Approach

Springer

Richard A. Clement
Deal, Kent, UK

ISSN 2195-1934 ISSN 2195-1942 (electronic)
Lecture Notes in Morphogenesis
ISBN 978-3-030-98494-6 ISBN 978-3-030-98495-3 (eBook)
https://doi.org/10.1007/978-3-030-98495-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book was written as a brief but self-contained introduction to the application of geometric techniques to understanding the brain. The approach is illustrated by characterising the data structures used to describe objects geometrically and, subsequently, showing how measurements provided by sensory neurons can be used to uncover the geometric descriptions of the objects present. A similar approach is applied to characterising the actions made by an animal in order to handle an object, and what this involves at the level of muscle contractions. Finally, it is shown that the geometrical approach leads to new and distinctive experimental techniques.

The approach of this book relies heavily on visual representation of the material. In this respect, I have been influenced by the book *Dynamics: The Geometry of Behaviour* by Abraham and Shaw [1] that gives a comprehensive introduction to the ideas of nonlinear dynamics entirely in pictures. They wrote their book to show that mathematical ideas can be communicated easily, and I think the material contained in it has not been taken up as widely as it might be by neuroscientists for two reasons. Firstly, there were no examples of applications to neuroscience in the book, and secondly, they did not include the additional stage of how to turn the intuitive visual ideas into equations that could be used to make testable predictions. This book aims to fill those gaps.

The target readers are those with a neuroscience background, but the content is also intended to be of interest to mathematicians, who will find that their brains do mathematics which their minds are unaware of; to physicists, who will find that brains have a few ideas of their own about how the world works; to computer vision scientists and roboticists, who will discover alternatives to the computational approaches with which they are familiar; and to neurologists, who will gain insight into the different ways that abnormal behaviours can arise.

[1] Abraham, R.H., Shaw, C.D. *Dynamics: The Geometry of Behaviour*. Second Edition. Addison-Wesley Publishing Company, Redwood City, California (1992).

The reader I have in mind is someone, maybe a school-leaver, maybe a university student or maybe even a researcher, who has been introduced to calculus in the past but is not aware of the insights generated by the geometric approach to the brain. I have assumed some knowledge of neuroscience and mathematics which I have reviewed in an appendix.

Deal, Kent, UK
February 2022

Richard A. Clement

Acknowledgements

All the simulation results, data analysis and figures were conveniently produced using the Mathematica software package.

I have twisted the arms of many people to give me feedback on numerous inchoate drafts of this book, and I wish to thank them all for their help: Richard Abadi, Ozgur Akman, Lucy Alderson, Silvia Barbina, Brendan Barrett, Christine Boylan, Sally Crighton, Alberto Gomez Gomez, Dmitri Laptev, Jane Lyle, Michael Marshall, Richard Montgomery, Eve Pascal and Thomas Wright. Richard and Ozgur in particular have helped straighten out my frequent misconceptions during our two decades of collaboration. Everything untoward is entirely down to me.

I been fortunate to benefit from the guidance of the editor Eva Hiripi. Her initial assessment of the book proposal coincided with a time when most academics were committed to updating their lectures for online presentation, and so finding reviewers was problematical. Nevertheless, she succeeded in finding reviewers who were adroit in pointing out topics that needed to be included, and the book is much more rounded because of their contribution. I am grateful to both the reviewers and Dr Hiripi. Thanks to production staff.

Special thanks to Christine for making the index and for being very patient with me during the aeons I have taken to write this book. “Now about this washing line you bought three years ago and still have not got out of the box”. Rumbled, I fear.

Many thanks also to Harini Devi and her team for the efficient and accommodating production of this book.

Contents

Chapter 1
Mind and Brain

What sort of things are brains good at dealing with? The argument put forward here is that they are adept at handling low-dimensional dynamical systems, primarily because these capture the behaviour of biologically relevant objects. This approach immediately raises the question of what are the distinctive brain mechanisms that enable them to be good at this task? If the behaviour of a dynamical system is described geometrically by the trajectory of a point corresponding to its state, then the changes in the firing rates of a population of neurons can be used emulate the changes in the coordinates of the state of the system as its behaviour changes. Whilst expanding on this argument, the opportunity is taken to introduce the basic terminology of dynamical systems theory and linear algebra that will recur throughout the book.

1.1 Objects in Mind

Imagine you are out for a walk on a winter evening. The trees rustle gently in the breeze. There is a clear sky with the stars just coming out. Suddenly a flying saucer appears from over the horizon. It is a sure thing you will notice it. Now why is that?

One difference between the saucer and the rest of the scene illustrated in Fig. 1.1 is that it can change in only a limited way. It is a rigid object which can alter its position along the left-right, up-down and near-far directions, and its orientation along the pitch, roll and yaw axes. This property of the saucer can be summarised by saying that it has six **degrees of freedom**. By comparison the trees have many more degrees of freedom because every branch and twig moves differently in the wind. Another difference between the saucer and the rest of the scene is that it moves in a steady trajectory from one position and orientation to the next, so over the short term at least, it is possible to make a **prediction** about where it will be in

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_1

Fig. 1.1 How a flying saucer might appear when you are out walking on a winter evening

the future whereas different parts of the trees move in different directions depending on the wind in an unpredictable manner.

Perhaps surprisingly, the behaviour of any object with a few degrees of freedom of movement and predictable behaviour can be understood by using a geometric description. The position and orientation of the flying saucer at any instant of time can be specified by a sextuple of numbers each of which depends on the variation along one of the degrees of freedom. Such a specification is referred to as the **state** of the object. The geometric approach exploits the idea that the changes in the state correspond to the trajectory of a point through some form of space even if it is not immediately obvious what such a space might look like. The collection of all trajectories taken by the object form a region of the space known as a **manifold**.

The description of the behaviour of an object in terms of its state is incomplete without an additional specification of how it changes from one state to the next. In geometric terms the state change corresponds to displacement of a point. Any such quantity which has both a length and a direction is referred to as a **vector**. So specification of how the state of the object changes is equivalent to assigning a vector to every point on the manifold. This assignment is known as a vector field. The combination of a manifold and an associated vector field is referred to as a **dynamical system**.

In practice it is predominantly inanimate objects that satisfy the requirements for a dynamical system in that they change from state to state according to a fixed rule. In the case of the spaceship, at this very moment there might be a couple of aliens wrestling over the controls to decide whether to land or go home, so the assumption of a fixed rule may not be valid.

But now you notice that the spaceship is getting closer to the ground and it looks as though it is going to land. You make your way towards it and by the time you get to it the saucer has landed and a ramp has been lowered to the ground. A strange

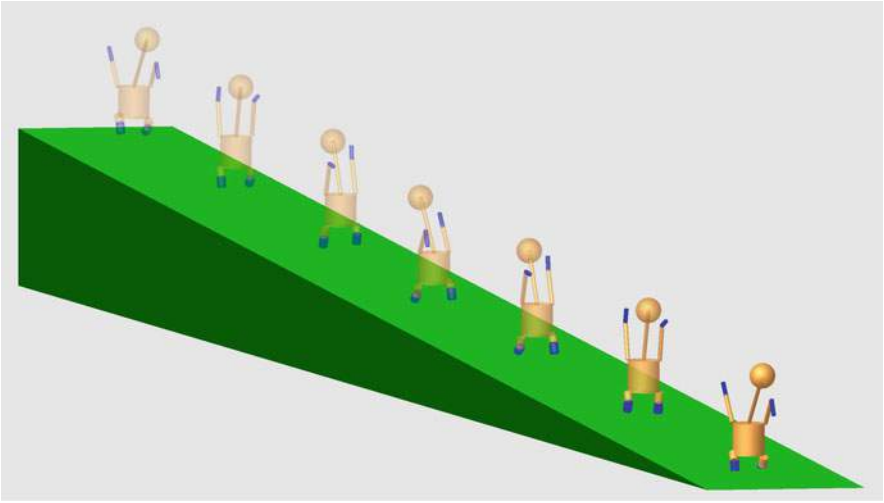


Fig. 1.2 Successive positions of an alien walking down a ramp. The images of the alien are faded to indicate that they represent positions further back in time

being comes down the ramp, perhaps like that shown in Fig. 1.2. You have never seen anything quite like it before but you can immediately tell that it has a lifelike way of walking.

Again, the point to note is that the alien has relatively few degrees of freedom. The movement patterns of animals are constrained by their skeletons and in the case of this simulated alien each of its arms moves in a plane and can only bend at the shoulder and elbow and so has only two degrees of freedom and similarly for the legs.

Maybe you have taken your dog with you for the walk. The dog shows no interest in the flying saucers and aliens but is engrossed by a plastic bag caught in a tree which for it may or may not be a cat. So it is all very well to argue that the brain can represent the behaviour of an object by the trajectory of a point on a manifold, but one has to go further and explain how the ability to decide that there is an object out there which corresponds to a particular dynamical system depends on the brain involved.

1.2 Objects on the Brain

Consider the simpler situation of a primitive videogame where an icon of rocket is shown moving across the screen. In order to ensure that the rocket does not simply disappear if it reaches the edge of the screen one can wrap around the screen so that when the spaceship reaches the edge of one side it continues travelling in the same direction from the same position on the opposite side. In terms of geometric

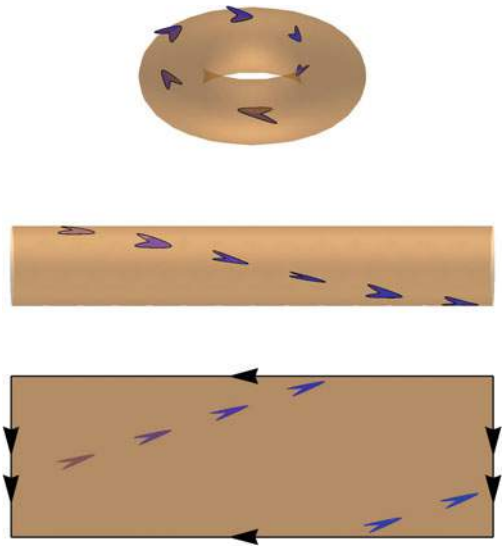


Fig. 1.3 Illustration of the plane diagram representation of a surface. **Upper Figure** Successive positions of a spaceship following a trajectory on a torus. **Middle Figure** An intermediate step in obtaining the plane representation of the torus. A tube is obtained by making a cut through the doughnut shape and then uncurling it to give a tube. **Lower Figure** The plane diagram is obtained by making a cut along the length of the tube and then unfurling it. In the plane diagram representation of a torus when the spaceship reaches the edge of the rectangle it jumps to the opposite side of the rectangle but continues travelling in the same direction. The edges of the plane diagram that should be understood as glued together are indicated by arrows

description, wrapping around the screen constrains the spaceship to move along the surface of a shape like a doughnut, as illustrated in Fig. 1.3.

The doughnut shape is an example of a torus and the unrolled representation of the torus is referred to as a **plane diagram**. In the plane diagram representation of a surface one begins with a polygon with 2n sides and then adds the additional requirement that pairs of edges are taken to be attached. The attached sides can be identified on the polygon by labelling each pair of edges with arrows to indicate how the edges should be joined together. Sides with the same number of arrowheads are taken to be glued together so that the arrowheads point in the same direction.

So here is a thing. The manifold with two degrees-of-freedom which is generated by the path of the spaceship is equivalent to a surface situated within everyday three-dimensional space. This observation raises the question of the space in which the manifold is contained.

Everyday three-dimensional space has some distinctive properties, which are simplest to explain with respect to a two-dimensional plane surface. Any point in the plane can be specified by a pair of coordinates (x_1, x_2). The point specified by the **coordinates** has a position given by a movement of an amount x_1 along the horizontal axis followed by an amount x_2 along the direction of the vertical axis,

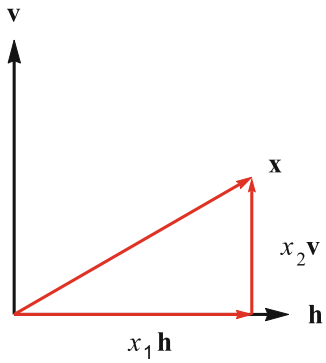


Fig. 1.4 Diagram illustrating how a point in the plane can specified by the coordinates (x_1, x_2) of a point on a graph or by a vector sum of horizontal and vertical displacements $x_1\mathbf{h} + x_2\mathbf{v}$

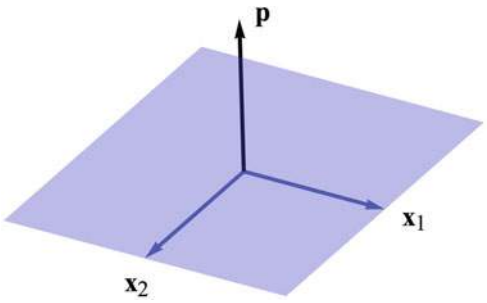


Fig. 1.5 Diagram illustrating a vector **p**, shown in black, that cannot be generated by a linear combination of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, shown in blue, that lie in the plane illustrated by a blue polygon

with the order in which the movements are made being unimportant. Using the vector concept, a point X in the plane which is specified by a pair of coordinates (x_1, x_2) can also be specified by a vector **x** which describes the change of position from the origin to the point by a sum of multiples of vectors in the horizontal (**h**) and vertical (**v**) directions, as illustrated in Fig. 1.4.

Given a pair of vectors, then their sum can be defined by attaching the head of one vector to the tail of the other. On can go on to generate many vectors by adding different multiples of two vectors together. Any such vector produced by adding multiples of a pair of vectors is referred to as a **linear combination** of the original set of vectors. If a vector cannot be expressed as a linear combination of the pair of vectors then it is referred to as linearly independent of the pair of vectors. Such a vector will lie outside the plane generated by linear combinations of the original vectors as illustrated in Fig. 1.5, except in the case where the original vectors have the same direction.

Although the linear combination concept can only be illustrated graphically for two or three vectors, it extends to any number of vectors. This generalisation is useful because many phenomena involve more than two or three variables and yet the geometric approach can still be used. The set of all the vectors which can be formed by linear combinations of a set of n linearly independent vectors constitutes a **vector space** and the set of n linearly independent vectors itself is referred to as the **basis** for the vector space which can be generated from it. The number of independent vectors n is referred to as the **dimension** of the vector space. So,

any vector $\mathbf{x}$ contained in a n-dimensional vector space can be written as a linear combination of the n basis vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ of the space:

$$\mathbf{x} = (x_1\mathbf{x}_1, x_2\mathbf{x}_2, \ldots, x_n\mathbf{x}_n) \tag{1.1}$$

where the n-tuple of numbers $(x_1, x_2, \ldots, x_n)$ are the coordinates of the vector.

Implicit in the explanation of how a pair of coordinates can be used to specify the position of a point X in the plane is the assumption that the amount of the corresponding vector $\mathbf{x}$ in the direction of the horizontal and vertical vectors can be calculated. The amount of a vector $\mathbf{x}$ in the direction of another vector $\mathbf{y}$ is referred to as the projection of $\mathbf{x}$ onto $\mathbf{y}$. The projection of a pair of vectors onto each other is given by the inner product of the vectors. This is an operation which takes two vectors $\mathbf{x}$ and $\mathbf{y}$ and gives a single number that is a measure of their projection. What is special about the plane is that it has an inner product, denote by a dot, that is obtained by multiplying corresponding coordinates and then summing the individual products:

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 \tag{1.2}$$

This inner product is referred to as the Euclidean inner product, or more succinctly as the **dot product** .

The inner product can also be used to obtain a measure of the length of a vector, denoted $\|\mathbf{x}\|$, by taking the square root of the inner product of a vector with itself. The Euclidean inner product implies a measure of distance along a vector that is familiar from Pythagoras's theorem. If a vector $\mathbf{x}$ is taken to lie along the hypotenuse of the right angle triangle formed by the projections (x_1, x_2) along the horizontal and vertical directions respectively, as is the case for the vector illustrated in Fig. 1.4, then the dot product is equivalent to Pythagoras's theorem.

$$\|\mathbf{x}\|\|\mathbf{x}\| = \mathbf{x} \cdot \mathbf{x} = x_1 x_1 + x_2 x_2 \tag{1.3}$$

It is also evident from Fig. 1.4 that the measure of projection provided by the dot product is related to the angle between the two vectors. If the horizontal vector $\mathbf{h}$ is assumed to be of unit length then the projection of the vector $\mathbf{x}$ onto the horizontal vector is equal to the length of the vector $\mathbf{x}$ times the cosine of the angle between the vectors. So the definition of the dot product in terms of coordinates is equivalent to another formulation that is not so easy to calculate but is simpler to understand:

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|(\text{cosine of the angle between the vectors } \mathbf{x} \text{ and } \mathbf{y}) \tag{1.4}$$

The definition of the Euclidean inner product generalises to n-tuples of coordinates:

$$\mathbf{x} \cdot \mathbf{x} = x_1 x_1 + x_2 x_2 + \ldots + x_n x_n \tag{1.5}$$

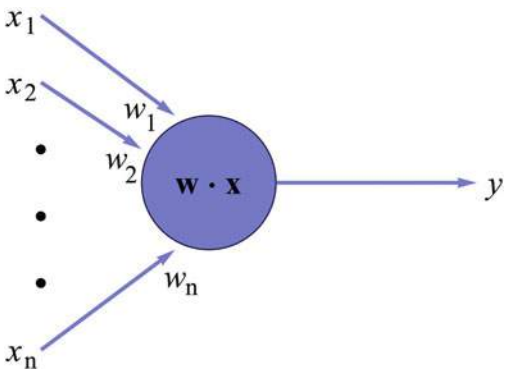


Fig. 1.6 Components of a simplified model of a neuron. Each input x_i is multiplied by an associated weight w_i and the output y is equal to the sum of the weighted inputs $\mathbf{w} \cdot \mathbf{x}$

and so one can define distance in an artificial space with n-dimensions, each associated with a different coordinate. The n-dimensional space of real numbers with a Euclidean inner product defined on it is referred to as n-dimensional **Euclidean space**. Generalising from the case of a doughnut-shaped surface contained within three-dimensional Euclidean space one can describe an m-dimensional manifold as a subset of n-dimensional Euclidean space, where the number n is greater than m. This description is referred to as an **embedding** of the manifold in Euclidean space. The advantage of this description is that every point on the embedded manifold then has a unique n-tuple of coordinates given by the position of the point in Euclidean space. Uniqueness of representation is a sine qua non for ensuring that the right movement is made when interacting with an object in a given state.

The definition of the n-dimensional Euclidean inner product leads directly to the reason that the geometric representation is suited to the capabilities of neurons. A simple model of a neuron has a number of inputs, which describe the effects of the synapses on the dendrites and cell body of the neuron, and a single output as shown in Fig. 1.6. The individual inputs are weighted to represent the differing synaptic strengths of the inputs and the sum of these weighted inputs corresponds to the potential across the membrane of the cell body. Let the levels of the n inputs be denoted by $(x_1, x_2, \ldots, x_n)$ and the levels of the output be denoted by y. Further, let the weights associated with each of the inputs be denoted by $(w_1, w_2, \ldots, w_n)$ and assume that the output depends on the sum of the inputs, then the output y of the neuron is equal to the sum of the products of each input and its associated weight.

$$y = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = \mathbf{w} \cdot \mathbf{x} \tag{1.6}$$

This model neuron behaves as though tuned to respond to an input vector $\mathbf{x}$ that is aligned with the direction of the weight vector $\mathbf{w}$ because its output depends on the cosine of the angle between the two vectors. As the output is directly proportional to the inputs, the model is also known as a **linear filter**, tuned to the input that matches the weight vector.

Because such neurons compute the Euclidean inner product it is easy to set up a neural representation of a two-dimensional Euclidean space by using a pair of neurons, each specifying the coordinate of the point in a Cartesian system of axes. One neuron has input weights that are a cosine function of the angle that the direction of the point makes with the x-axis. The output of the neuron then corresponds to the projection of the vector to the point onto the x-axis. The second neuron carries out a similar computation for the y-axis coordinate.

This model is inaccurate because biological neurons have a limited range of firing rates and only depolarisation of the membrane produces spikes so that a more realistic description of the output of the neuron is given by taking only the positive values of the output y^+:

$$y^+ = \begin{cases} y, \text{ if } \quad y \geq 0 \\ 0, \text{ otherwise} \end{cases} \tag{1.7}$$

Both these drawbacks can be overcome by using a population of p neurons to represent the projection of an input onto each basis vector. The use of many neurons to hold the value of a measured variable is referred to as a **population code**. For example, the instantaneous direction of the rocket on the screen can be represented by a population of p equally separated, unit length vectors, as illustrated in Fig. 1.7.

An estimate $\hat{\mathbf{x}}$ of the movement direction is given by the population vector, which is the vector sum of the preferred directions of the population of p neurons weighted by their responses:

$$\hat{\mathbf{x}} = y_1^+ \mathbf{w}_1 + y_2^+ \mathbf{w}_2 + \ldots + y_\mathrm{p}^+ \mathbf{w}_\mathrm{p} \tag{1.8}$$

where the vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\mathrm{p}$ specify the tuning directions of the population of neurons. Note, however, that the length of the estimated vector will be given by the sum of the responses of all the neurons responding to the input vector and so the estimate has to be rescaled to obtain an accurate estimate of the length of the input vector:

$$\hat{\mathbf{x}} = \frac{y_1^+ \mathbf{w}_1 + y_2^+ \mathbf{w}_2 + \ldots + y_\mathrm{p}^+ \mathbf{w}_\mathrm{p}}{y_1^+ + y_2^+ + \ldots + y_\mathrm{p}^+} \tag{1.9}$$

Reduction of the output of a neuron by divisive inhibition from a pool of surrounding neurons is widespread in nervous systems [3], and often leads to saturation of the output as divisive inhibition becomes far more effective as the output increases.

Although a model that responds to a weighted sum of its inputs is a good first approximation to the behaviour of a neuron, a more realistic model takes into account the time required for the membrane potential to return its resting level. Assuming that the rate of decay depends on the current value of the membrane potential the behaviour of a model that incorporates the membrane potential can be

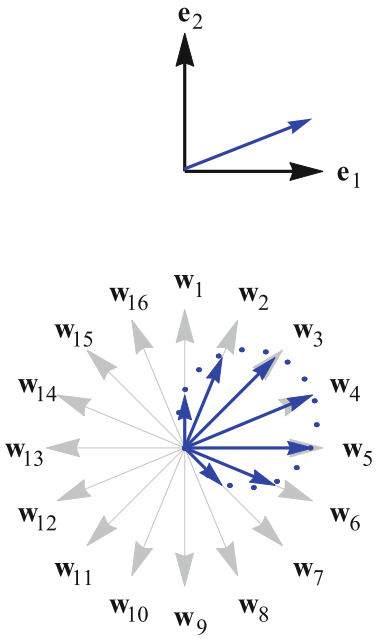


Fig. 1.7 Alternative descriptions of a two-dimensional vector. **Upper Figure** The vector is indicated in blue with respect to a pair of Cartesian axes described by the two basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. **Lower Figure** A population coding of the vector, based on 16 neurons with preferred tuning directions defined by the weight vectors $(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{16})$. These directions are plotted as grey arrows. The lengths of the blue vectors indicate the responses of the corresponding neurons, and the direction of the original vector can be reconstructed from the sum of these vectors. The blue dots are added to show how the response of a neuron with cosine tuning varies with the angle between its preferred direction and the target direction

described by the differential equation:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{y}{\tau} + \mathbf{w} \cdot \mathbf{x} \tag{1.10}$$

where τ is a constant referred to as the **time constant** τ.

The behaviour described by Eq. 1.10 can be understood by assuming that initially the neuron has a constant input. If the input suddenly drops to zero then the membrane potential will also change until the terms on the right hand side of the equation sum to zero. The response of the neuron decreases exponentially, as illustrated in Fig. 1.8 with a steepness that is determined by the time constant.

Although the membrane time constant of a single neuron is typically only a few tens of milliseconds, an advantage of using population vector coding is that this time constant can be increased by feedback connections within the population [2]. Both excitatory feedback connection of a neuron to itself and reciprocal inhibitory connections between different neurons result in lengthening of the effective time

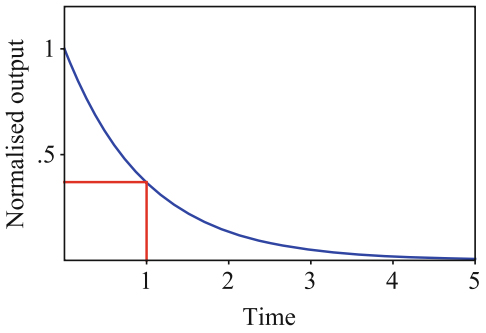


Fig. 1.8 Plot of the response of specified by Eq. 1.10 to an abrupt removal of an input. The response decreases exponentially at a rate that is determined by the time constant. In this case the time constant was 1. The red lines illustrate the property of the exponential function that when the amount of elapsed time t is equal to the time constant the output will have dropped to approximately 37% of its initial level

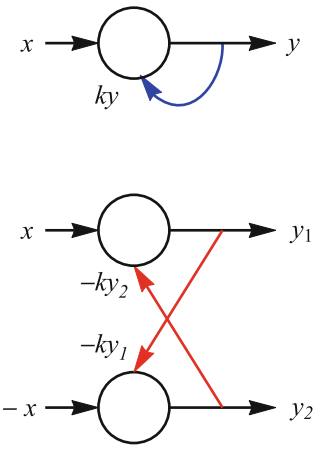


Fig. 1.9 Use of feedback to lengthen the time constant of a neuron. **Upper Figure** A positive feedback connection indicated by a blue arrow. **Lower Figure** Reciprocal inhibitory feedback connections, indicated by red arrows

constant so that a population of neurons can maintain a response even if an object is temporarily hidden.

In the case of excitatory feedback illustrated in Fig. 1.9, the equation describing the neuron can be put in the same form as Eq. 1.10 to show that the effective time constant of the neuron τ^* has been lengthened by comparison with the original time constant τ. Let a proportion k of the output be fed back then:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{y}{\tau} + \mathrm{k}y + x = -\frac{(1-\tau\mathrm{k})y}{\tau} + x = -\frac{y}{\tau^*} + x \tag{1.11}$$

where

$$\tau^* = -\frac{\tau}{1 - \tau \mathrm{k}} \tag{1.12}$$

The effect of inhibitory feedback connections is easiest to grasp in neural networks in which as many neurons respond to a signal change in one direction as respond to a signal change in the opposite direction. A simple two neuron network with this property is shown is also shown in Fig. 1.9. The behaviour of the neurons in this circuit is described by the equations:

$$\begin{aligned} \frac{\mathrm{d}y_1}{\mathrm{d}t} &= -\frac{y_1}{\tau} - \mathrm{k}y_2 + x \\ \frac{\mathrm{d}y_2}{\mathrm{d}t} &= -\frac{y_2}{\tau} - \mathrm{k}y_1 - x \end{aligned} \tag{1.13}$$

Given that the inputs to the two neurons are equal and opposite, from the symmetry of the network it follows that the outputs will be equal and opposite, provided that the outputs of both neurons are initially zero. So one can let $y = y_1 = -y_2$, and substitute for y_1 and $-y_2$ in the pair of Eqs. 1.13 for the circuit. In both cases, the result in an equation identical to the Eq. 1.11 for positive feedback, and the time constant is correspondingly lengthened.

This outline of the properties of a simple model neural network indicates how a brain could embody an object such as the spaceship on the game screen. A brain with sufficient neurons can embed the manifold of the object in an n-dimensional Euclidean space and a population of neurons can be used to keep track of the trajectory of the object. Furthermore, given that the accuracy of the vector reconstructed from the population of neurons depends on the number of neurons [6], it follows that different brains will vary in their capability for interacting with an object depending on the number of neurons they can throw at the problem. But animal behaviour is not that simple. In the everyday world there are usually many objects present and to achieve a goal the brain has to discern the relevant objects.

1.3 Using Your Grey Matter

The ability to identify and select the appropriate objects to interact with is perhaps the most distinctive aspect of animal behaviour. What is the mechanism underlying this ability? A working assumption is that the brain measures the physical properties of the objects and uses this data to identify features of the objects, the idea being that the objects can then be distinguished by their different combinations of features.

Given a feature one can arrange all the possible instances of the feature as points on a manifold to obtain a **feature space**. The simplest features are directly related to sensory measurements. For example, the opponent colour features of light/dark, red/green and blue/yellow can be defined approximately in terms of the outputs

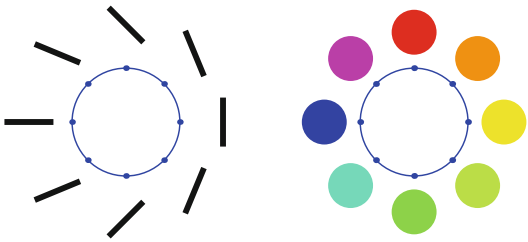


Fig. 1.10 Correspondence of the orientation and hue feature spaces to the projective line

of the long (L), medium (M) and short (S) wavelength cones in the retina by the expressions L+M+S, L-M and L+M-S respectively and these triplets of numbers can be used as coordinates of points in three-dimensional Euclidean space. In passing it is important to note that just because the arrangement of features follows that of points in Euclidean space does not mean that subjective sensations are separated by the Euclidean distance between them, because all subjective sensations are reconstructed from the outputs of sensory neurons and are interpreted within the context of all the ongoing neural activity.

Other features, especially spatial ones, can be specified by an equivalence relationship between points in Euclidean space. For example the real projective space is formed by taking an n-dimensional Euclidean space with the point at the origin removed and applying the relation that point $\mathbf{x}$ is equivalent to point $\mathbf{y}$ if and only if $\mathbf{x} = \lambda\mathbf{y}$ for some decimal number λ. The simplest example of a projective space is the projective line. Each point of the projective line is equivalent to the collection of points on a line through the origin of a plane, excluding the origin. So one can represent each point in the feature space by an icon consisting of an orientated line and these can be arranged in a circle as illustrated in Fig. 1.10.

Another example of a real projective line can be created from the colour features. By ignoring the light/dark coordinates, all the points in opponent colour feature space are projected into the plane. One can then apply the projective line equivalence relationship to obtain a circle where each point corresponds to a different hue. Going anticlockwise from red to blue around the circumference of the circle the hue becomes reddish-blue (magenta) and going further from blue to green the hue becomes bluish-green (cyan) and so on. However when going from magenta to cyan there is a hue (blue) which is neither reddish or greenish.

The equivalence relationship that defines an m- dimensional projective space equates every point on a line through the origin in an $(m + 1)$-dimensional Euclidean space to a single point, with the proviso that the origin of the Euclidean space is excluded. In the case of the **projective plane** this means that one can take a sphere and form the projective plane by joining points on the sphere that lie at opposite ends of a line through the origin of the sphere, as illustrated in Fig. 1.11. Almost all the pairs of antipodal points can be represented by a point on a hemisphere, but not those on the equator. Points along the two halves of the equator have to be joined together with a twist, and this can only by done by having an extra dimension which enables one to make the twist without the surface intersecting itself. So the real

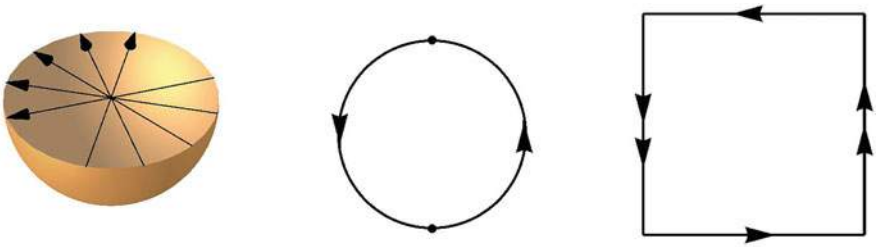


Fig. 1.11 Hemisphere model of the real projective plane. Diametrically opposite points on the equator are linked together as illustrated by the arrows in the middle diagram and the corresponding plane diagram is shown on the right

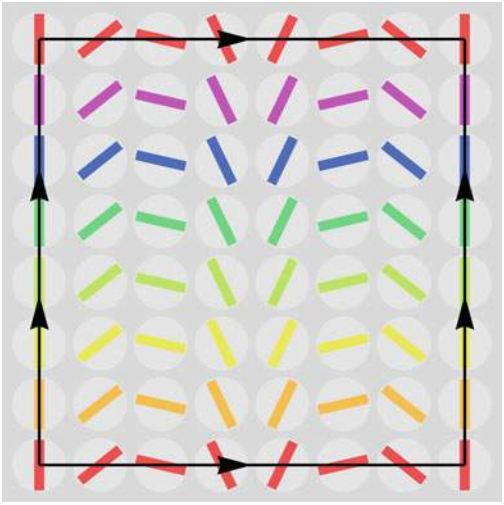


Fig. 1.12 Plane diagram of a torus formed of a product space of orientation and hue features

projective plane is defined with respect to three dimensional Euclidean space but cannot be embedded in a Euclidean space with less than four dimensions.

Another technique for defining new features is to form combinations of existing ones. Given two manifolds X and Y then one can attach a point x from manifold X to a point y from manifold Y to make a single point in a new manifold. The manifold obtained by generating all the possible pairs is known the **product space** and is denoted by $X \times Y$. For example a torus can be considered to be the product space of two circles, so once can use the product of the orientation and hue feature spaces to form a new feature space with the form of a torus. The distribution of features in this space can be illustrated by plotting the graphical feature associated with a point at a corresponding point on the plane diagram of the space, as illustrated in Fig. 1.12.

Given a feature space then any trajectory across the manifold will generate a sequence of features and such sequences can be used as stimuli to test the

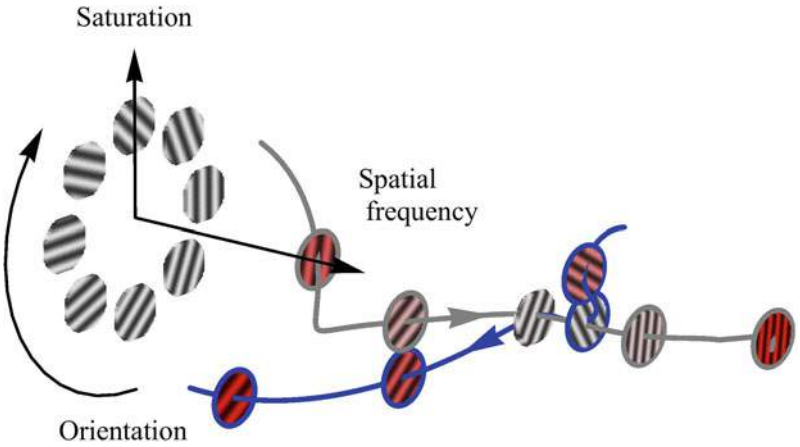


Fig. 1.13 Trajectories of the target (plotted as a blue line) and distractor (plotted as a grey line) in the feature space used in the tracking task. The corresponding visual stimuli at regular time intervals are illustrated by the grating icons, and the trajectories intersect at the third icon along. Note that the feature space coordinates do not include time, so that the samples taken at regular time intervals are not regularly spaced in the plot. Instead their separation depends on how rapidly the features are changing

capabilities of the brain. For example, one of the experiments that demonstrated that attention can select for specific visual objects on the basis of features, as opposed to simply being in different locations, consists of a tracking task in which the subject is presented with a target object and a distractor, both consisting of a grating with changing orientation, colour saturation and spatial frequency but in the same spatial location, and at the end of the display period has to report the final orientation of the target.

The product space of orientation, saturation and spatial frequency of the tracking task can be embedded in three-dimensional Euclidean space. The simplest way to arrange the features is to use a cylindrical system of coordinates in which spatial frequency is plotted along the long axis of the cylinder, saturation along the radius of the cylinder and orientation around its circular base. An example of two trajectories, one corresponding to the target and the other to the distractor, is illustrated in Fig. 1.13 and a sample of the associated image sequence as it appears to the subject of the experiment is shown in Fig. 1.14. Despite the changing form of both target and distractor, subjects are able to track the target successfully [1].

A possible interpretation of the results of the tracking task is that equivalence relationships between the measurements provided by the sensory neurons are used to group the measurements into more complicated features which ultimately relate to the type of objects which an organism has to interact with. An illustration of where this approach leads is that it should be possible to define a relationship which identifies the pairs of arms and legs which belong to the same person [4]. Then a

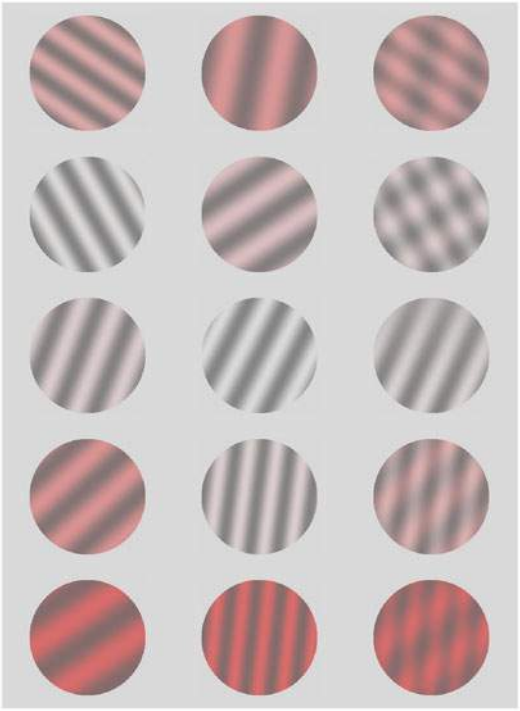


Fig. 1.14 The leftmost two columns are comprised of the samples of the target and distractor that are illustrated by icons in Fig. 1.13 and the column on the right shows the combined stimulus which is what the subject sees

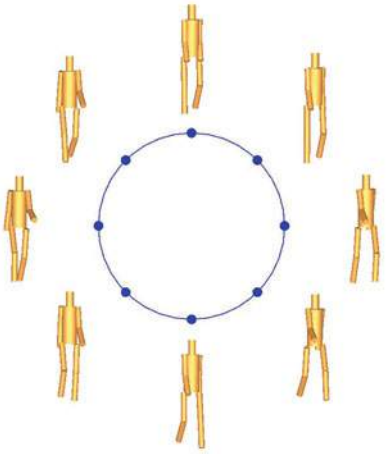


Fig. 1.15 Illustration of how an example of complicated behaviour such as walking can be simplified if it is possible to match the relative positions of corresponding body parts (in this case head, torso, arms and legs) with points on a one-dimensional manifold

cyclical behaviour pattern such as walking, which might appear complicated, has a simple description as a point moving around a circle, as illustrated in Fig. 1.15.

Although the argument that objects can be identified by their characteristic features sounds plausible, no one has yet succeeded in actually defining the features used by the brain to identify an object such as a person walking. To be clear, the

definition of a feature has to specify an equivalence relationship, and how it can be computed, so that data which comply with the relation can be considered to have the feature. This was possible for the opponent colour feature, where the colour is defined by a specific cone ratio, and for orientation and hue, where a projective relationship is used. But most features are not well-defined, so that the term feature is possibly the second most overused term in neuroscience (the most overused term is undoubtedly information). It may be that we need to move away from investigating collections of features to more complicated relationships such as the constraints underlying possible movements [5], some examples of which are given in the next chapter.

1.4 About This Book

The argument of this book is that the type of data structures processed by the brain can be understood by analysing them geometrically and its content is designed to build on this geometrical approach by introducing examples of the geometric structures that the brain has to work with and then describing how these data structures are handled by neurons.

Depending on whether you are a visual or a symbolic thinker, a positive feature of the geometric approach is that the ideas about how the brain works can be expressed visually. For example, if one considers a discrete item of behaviour, such as moving one's gaze from one object to the next, then one can work out an underlying mechanism using only geometric arguments [7]. The argument begins with the assumption that there are three components to the behaviour: a stable equilibrium, a threshold that has to be crossed for the action to begin and a trajectory that returns the state of the mechanism to the stable equilibrium.

Given the required behaviour, one can develop a geometric description of the underlying dynamical system as illustrated in Fig. 1.16. The manifold defined by the collection of possible trajectories of a dynamical system is referred to as a **state space** and an equilibrium state is referred to as a **fixed point** of the state space because an equilibrium state is one that does not change. Using this terminology, one can begin the description with a one-dimensional state space that has a vector field directed towards a fixed point. This is then extended to a two dimensional state space state space, because the trigger property requires a vector field transverse to the one-dimensional system. It also requires a fold in the line representing the one dimensional-system so that if the state of the system is displaced beyond the fold the transverse vector field carries the state away from the stable equilibrium. Finally, the return to the stable equilibrium can be ensured by further folding the line of the one-dimensional system into an S-shape so that the state is guided back to the equilibrium.

The features of the qualitative geometric model can be used to guide the development of a corresponding quantitative model. For example by using a cubic curve to describe the s-shaped one-dimensional system one can obtain a differential

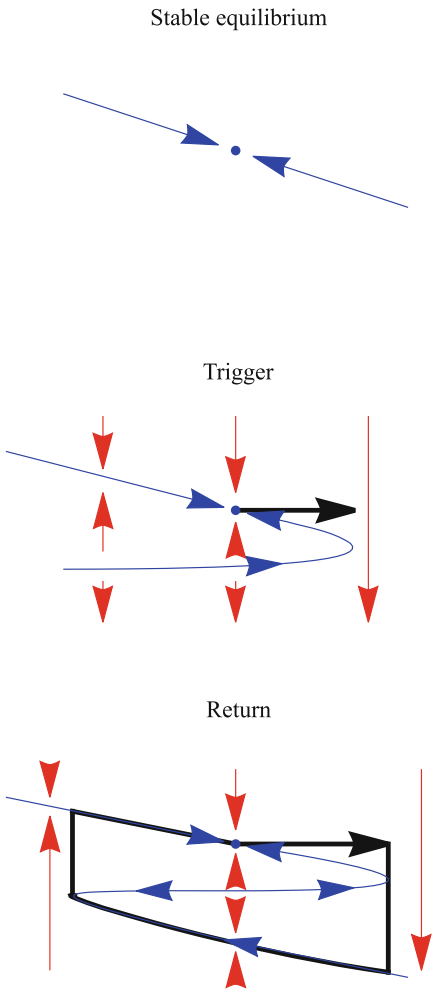


Fig. 1.16 Geometric construction of a mechanism for carrying out an action. **Upper Figure** The first step is to create a system with a stable equilibrium. The simplest case consists of a one-dimensional system. The equilibrium state is represented by a blue dot and the arrows indicate that elsewhere the state of the system moves towards the equilibrium state. **Middle Figure** The next step is to incorporate a trigger for initiating the action. This can be done by making a fold in the one-dimensional system and extending it to two dimension by adding a transverse vector field which is plotted in red. The black arrow indicates the displacement of the state from its stable equilibrium position that results in the state of the system being carried away from its equilibrium position. **Lower Figure** An additional fold results in the state of the state of the system returning to equilibrium after it has been displaced from equilbrium. The black line illustrates the trajectory associated with a complete action

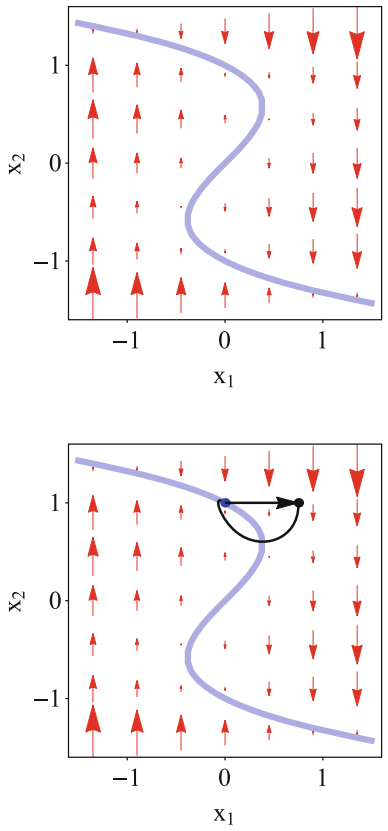


Fig. 1.17 Behaviour of the quantitative model. **Upper Figure** Transverse vector field described by Eq. 1.14. The blue line indicates the position in the state space of the one-dimensional system in which the equilibrium state will be placed to complete the model. **Lower Figure** Vector field of the completed model defined by Eq. 1.15. The blue dot indicates the stable equilibrium position. Although the state of the system spirals back to equilibrium after a displacement indicated by the black arrow, by comparison with Fig. 1.16 the sudden switch from one portion of the one-dimensional system to the other is absent

equation that specifies the transverse vector field illustrated in Fig. 1.17:

$$\frac{dx_2}{dt} = -(x_2^3 - x_2 + x_1) \tag{1.14}$$

The model is completed by adding a second equation that ensures that the state moves along the folded line towards an equilibrium position at the point where x_1 equals zero and x_2 equals 1:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 - 1 \\ \frac{dx_2}{dt} &= -(x_2^3 - x_2 + x_1) \end{aligned} \tag{1.15}$$

Plots of how the two variables change over time once the state has passed the trigger point are given in Fig. 1.18. These plots show that the two variables could provide the signals necessary to control an action. But the trajectory of the model does not have the abrupt threshold of the geometrical design illustrated in Fig. 1.16, and to achieve this some additional constraints have to be defined mathematically and these

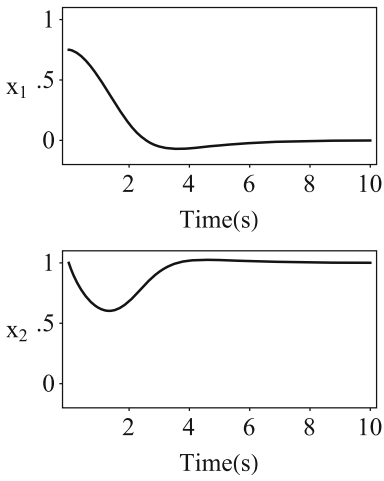


Fig. 1.18 Plots of the value of the model variables x_1 and x_2 against time for the trajectory illustrated by the black line in Fig. 1.17. The first variable could be used to track the progress of the action and the displacement of the second variable from its resting level could be used as a velocity command, reaching its peak midway through the action

are covered in Chaps. 5 and 6. But even with a completed model one still has to test if it matches the behaviour of an actual system and the geometrical approach leads to new ways of characterising behaviour which are covered in Chap. 7.

So a second goal of this book is to provide a self-contained introduction to the mathematical techniques used in the geometrical approach in order that the underlying ideas are made clear. This is especially important as the apparently abstract conceptualisations such as that of the geometrical description of an action turn out to have concrete embodiments in the brain, as described in the text in the context of the examples.

To summarise, the aim of this book is to give examples of applications of geometrical techniques relevant to mainstream neuroscience in order to motivate the approach and to introduce sufficient background to the techniques to enable the reader to understand what is involved in their application.

References

1. Blaser, E., Plyshyn, Z., Holcombe, A.: Tracking an object through feature space. Nature **408**, 196–199 (2000). https://doi.org/10.1038/35041567
2. Cannon, S.C., Robinson, D.A., Shamma, S.C.: A proposed neural network for the integrator of the oculomotor system. Biol. Cybern. **49**, 127–136 (1983). https://doi.org/10.1007/BF00320393
3. Caradini, M., Heeger, D.J.: Normalization as a canonical neural computation. Nat. Rev. Neurosci. **13**, 51–62 (2012). https://doi.org/10.1038/nrn3136
4. Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psychophys. **11**, 201–221 (1973). https://doi.org/10.3758/BF03212378
5. Öğmen, H., Herzog, M.H.: The geometry of visual perception: retinotopic and nonretinotopic representations in the human visual system. Proc. IEEE **98**, 479–492 (2010). https://doi.org/10.1109/JPROC.2009.2039028

6. Salinas, E., Abbott, L.F.: Vector reconstruction from firing rates. J. Comput. Neurosci. **1**, 89–107 (1994). https://doi.org/10.1007/BF00962720
7. Zeeman, E.C.: Differential equations for the heartbeat and nerve impulse. In: Waddington, C.H. (eds.) Biological Processes in Living Systems, pp. 8–67. Routledge, New York (1972). https://doi.org/10.4324/9781351297165

Chapter 2
Biological Objects

Brains are adept at interacting with biological objects. How to characterise such objects? Following the geometric approach, every possible combination of position and orientation of an object is described by a point in a configuration space and the instantaneous change in the position and orientation of the object is characterised by a vector in the tangent space to the configuration space at that point. This approach is extended to include specification of the length of a trajectory through the configuration space and movements of the eye, arm and falling cat are shown to follow the shortest path. In particular, the righting reflex of the falling cat shows that effective behaviour depends on the animal following the appropriate trajectory throughout the entire movement. The neural basis of trajectory following is introduced in the context of an animal navigating a path towards a required destination.

2.1 Configuration Spaces

The geometric approach can be extended to goal-directed behaviour, but to do so requires some additional concepts. To begin, if every configuration of a mechanical system is represented by a point then the collection of all the points can be used to define a manifold referred to as the **configuration space**. More often than not, mechanical constraints ensure that collections of data points generated by configuration changes correspond to smooth manifolds.

In the case of the eye, it is approximately spherical and is embedded within connective tissue in the orbit so that movements of the eye consist mainly of rotations, and the configuration space is that of rotations in three-dimensional Euclidean space. But for gaze directed at a distant object with the head upright the eyes only make a subset of the rotations which are mechanically possible.

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes
in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_2

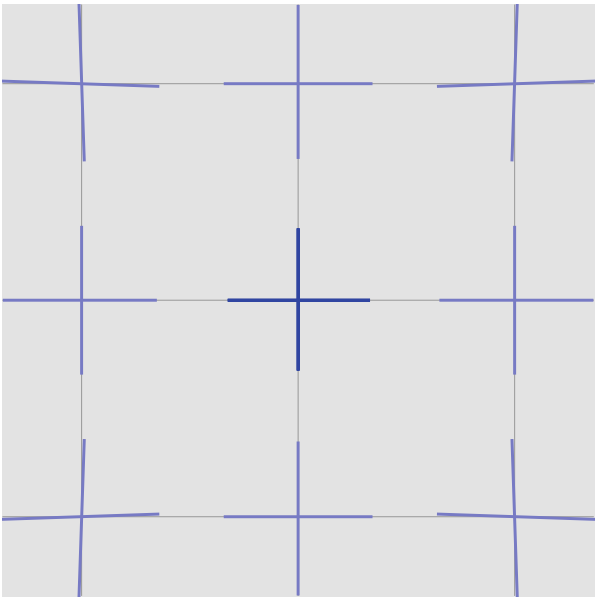


Fig. 2.1 Use of afterimages to investigate the orientation of the eye. An afterimage is formed of a cross on a wall facing the observer and the position of the afterimage is recorded with various directions of gaze

This subset was initially investigated by using afterimages, which have a fixed location on the retina and so move with the eye. A straightforward procedure involves forming an afterimage of a cross, with the eyes in the straight ahead position looking at a far point on the horizon with the head erect, and then measuring the apparent orientation of the cross whilst looking at a wall perpendicular to the primary position. This procedure is illustrated in Fig. 2.1.

The cross is not distorted when the point of fixation lies on the horizontal or vertical meridians. In general, an afterimage of an oblique line does not appear to alter its orientation at points along the associated oblique meridian, so the axis of rotation must be perpendicular to the oblique meridian. Let the line passing through the centre of rotation of the eye and the fovea be called the line of fixation, then this finding is summarised by **Listing's law**. Listing's law states that if the eye moves about a centre O so that the line of fixation moves away from the straight ahead position OA to another position OB, then the displacement of the eyeball is equivalent to rotating it around an axis perpendicular to the plane AOB [4].

If each rotation of the eye away from the straight ahead position made in accordance with Listing's law is represented by a point then a surface is defined that corresponds to the projective plane [12]. To visualise why this is so represent

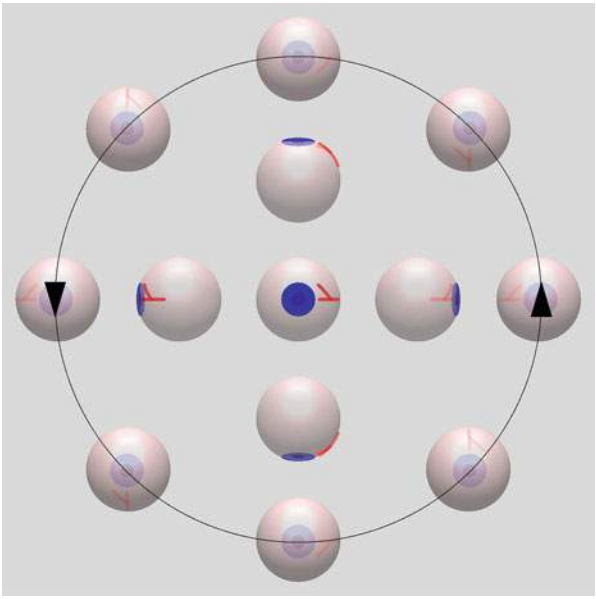


Fig. 2.2 Plane diagram of the configuration space of an eye moved in accordance with Listing's law. A prominent blood vessel has been used to indicate the orientation of the eye in a selection of positions. The outermost positions correspond to half turns of the eye and every pair at opposite ends of a diagonal are the same position, arrived at by rotations in opposite directions. In the plane diagram of the configuration space these positions are taken to be glued together, as indicated by the black circle, giving the plane diagram of a projective plane. It is clear that only a small portion of the configuration space is actually used by the eye

a rotation away from the straight ahead position by a vector that has a direction equal to the axis of rotation and a length equal to the size of the rotation. The maximum length of the vector is the half-turn π, and rotations through angles of plus or minus π are equivalent. If the positions in the axis-angle diagram corresponding to equivalent rotations are attached together then the plane diagram for the real projective plane is obtained, as illustrated in Fig. 2.2.

For another example of a configuration space take the possible configurations of a two-joint arm that moves only in the horizontal plane. These configurations could be represented by points on a torus, with the position of each point determined by the angles at the shoulder and elbow joints. But it is the position of the hand that is biologically important, so the configuration space consists of the points that the hand can reach. These points form an annulus, with the outer boundary of the annulus set by the maximum extension of the arm and the inner boundary determined by the difference in lengths between the upper part of the arm and the forearm. If it is assumed that these distances are the same, as is approximately the case in humans, then the configuration manifold corresponds to a disk as illustrated in Fig. 2.3.

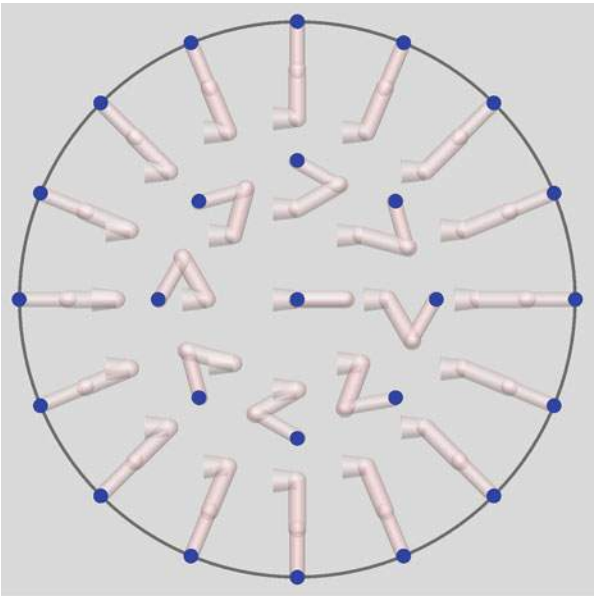


Fig. 2.3 Plane diagram of the configuration manifold associated with arm movements in the horizontal plane. The target hand positions located in the horizontal plane are represented by blue dots, and the relative positions of the shoulder, upper arm and forearm in each position are illustrated by icons. In this case no sides of the diagram are glued together and the diagram represents a portion of a plane

As the human arm cannot be bent backwards at the elbow there is a unique arm configuration for each hand position.

The arm example is different from the eye example in that the configuration changes involve shape changes. It illustrates how such configuration spaces can be formed from combinations of shape changes and transformations of a given shape in space. In the case of the horizontal movements of the arm, the shape is specified by the angle between the upper arm and the forearm and the transformation consists of a rotation about the shoulder joint.

In terms of manifolds, the configuration space can be constituted by taking a **base** or **shape** manifold and attaching another manifold to every point of it. In this context the attached manifold is referred to as a **fibre** and the manifold formed by this construction is known as a **fibre bundle**. A fibre bundle is not necessarily the same as the product of the base manifold with another manifold as can be appreciated from the fibre bundle of a base manifold consisting of a circle with a fibre consisting of a line. The product space consists of a cylinder, but the fibre bundle can be either a cylinder or a Möbius strip, as illustrated in Fig. 2.4. A fibre bundle which is equivalent to the product space is referred to as a trivial bundle.

The fibre component of the configuration manifold can often by identified because it matches a pattern of symmetry of the configuration space. For example,

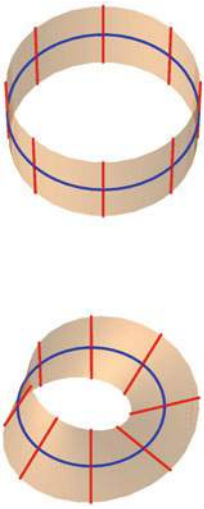


Fig. 2.4 Illustration of the possible fibre bundles comprised of a circle for a base and a line for a fibre. **Upper Figure** The trivial bundle is formed by the product space of the base and fibre manifolds and corresponds to a cylinder. **Lower Figure** It is also possible to form a non-trivial bundle by attaching the fibres with a twist, giving a Möbius band

the possible orientations of a fixed shape of the arm can be described by a rotation. Such transformations are examples of **symmetry operations**, in that they leave unchanged some property of the entity to which they are applied. The points of a configuration space, together with a symmetry operation, form what is referred to as a group structure.

In the case of rigid body transformations the relative positions of the points in the space are unchanged. The collection of translations, rotations and reflections associated with rigid body transformations of n-dimensional Euclidean space is called the Special Euclidean group. The collection of just the rotations and reflections is known as the Orthogonal group and that formed by just the rotations is referred to as the **Special Orthogonal group** and is denoted by **SO(n)** where n is the number of dimensions of the space. Each symmetry operation can be represented by a point on a manifold so these collections of transformations are also manifolds. A fibre bundle for which it also holds that the fibre is a group is referred to as a principal bundle.

The distinction between base and fibre manifolds turns out to be very useful in applying geometric ideas to the overall movement of an object which is produced only by changes in the shape of the object such as the righting reflex of a falling cat. In the case of the cat the configuration space is comprised of the possible positions and orientations of the limbs and body of the cat in space. The manifold is equivalent to a fibre bundle made up of a base consisting of the possible shapes of the cat and a fibre formed by the possible positions and orientations of the cat in everyday three-dimensional space [9].

A simple model of the spine of the cat can be made by assuming that the front and back halves of the cat move as though they are joined together by a ball and socket joint. The shape of the cat can then be specified by an angle ψ, which specifies its backbend and an angle θ which specifies its roll, as illustrated in Fig. 2.5. The angle

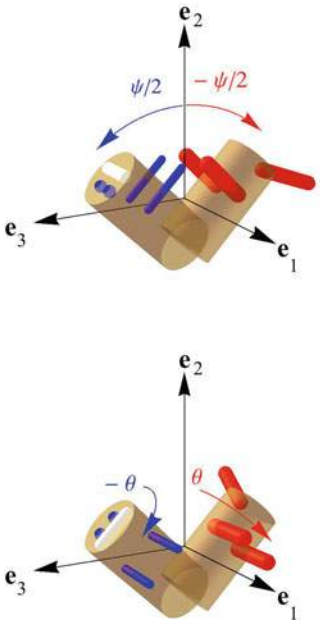


Fig. 2.5 The shape of the two cylinders model cat can be specified by a pair of angles that are defined with respect to the plane determined by the axes of the front and back halves. The system of axes has its origin located at the centre of mass of the cat. The $\mathbf{e}_2$ direction is aligned with the bisector of the two halves and points in the up direction. The $\mathbf{e}_3$ direction is perpendicular to $\mathbf{e}_2$ in the plane spanned by the two halves. The appendages to the cylinders are for illustration only and are not part of the actual model. **Upper Figure** The angle between the two halves is denoted by ψ. **Lower Figure** The tilt of the feet with respect to the vertical is denoted by θ

ψ corresponds to the angle between the two halves and so can vary between 0 and π. When the two halves are parallel ψ equals 0 and when they are in line ψ equals π. In both these limiting cases the axes of the two cylindrical halves have the same direction and cannot be used to define a unique plane passing through them.

With this model of the spine the base manifold consists of the product SO(3) $\times$ SO(3) corresponding to the possible orientations of the front and back halves. For the purposes of analysing the righting-reflex of the cat its position above the ground can be ignored. Also, of the possible orientations SO(3) of the cat in space, only the roll SO(1) is relevant for the righting. So the configuration space consists of a principal fibre bundle comprised of a base manifold SO(3) $\times$ SO(3) and a fibre SO(1).

From observations of the falling cat it can be seen that the body of the cat bends but does not twist. This limited flexibillity can be incorporated into the model by assuming a 'no-twist' condition that requires that the angle θ is equal and opposite for both front and back halves [6]. The arrangement of possible shapes of this model cat corresponds to that of the points on a projective plane as illustrated in Fig. 2.6.

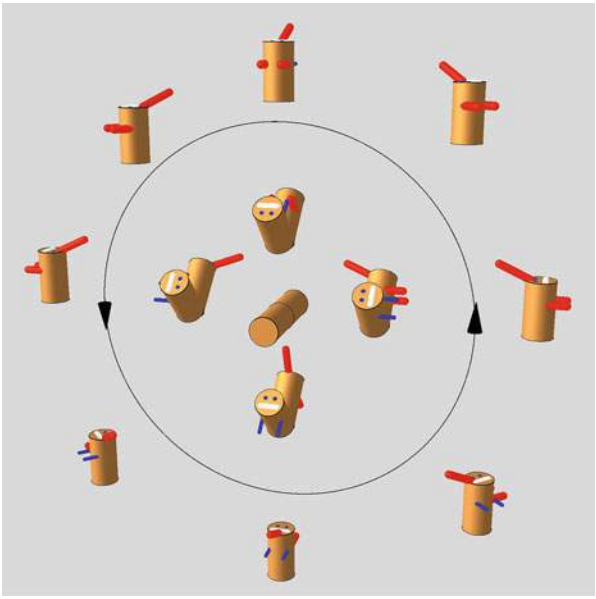


Fig. 2.6 Illustration that the arrangement of possible shapes of the falling cat corresponds to a projective plane, because the diametrically opposite shapes with zero angle between the halves are identical

Hence the configuration manifold can be simplified to the product of a rotation of a point, which gives a circle, and the real projective plane.

2.2 Tangent Spaces

Given that a smooth surface can be generated by range of configurations of a mechanical system, a change in the configuration of the mechanism will correspond to a trajectory over the surface. The instantaneous changes in the position of a point on the trajectory can be characterised by a vector that specifies the direction and distance moved by the point. At any given position on the manifold the vector describing the movement will lie in the tangent plane to the position, and being able to specify the tangent plane is a first step in calculating the distance moved along the trajectory.

The process of calculating the tangent plane is simplified by using the coordinates of n-dimensional Euclidean space. The system of coordinates can be set up by embedding the configuration space in Euclidean space. The embedding can be specified by a set of parametric equations in which a pair of parameters, corresponding to the two degrees of freedom for moving around the surface, determines the coordinates. For example, the position of a point on a unit radius sphere can be

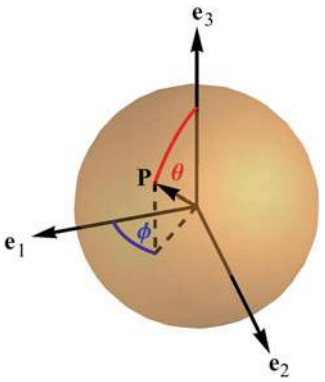


Fig. 2.7 Specification of the position of a point P on a sphere by the angles of latitude θ and longitude ϕ. Given these parameters, the coordinates of the point can be calculated by using Eq. 2.1

identified by the angles of latitude θ and longitude ϕ, as illustrated in Fig. 2.7, and a set of equations that give the corresponding coordinates in 3-dimensional Euclidean space is:

$$\begin{aligned} x_1 &= \sin(\theta)\cos(\phi) \\ x_2 &= \sin(\theta)\sin(\phi) \\ x_3 &= \cos(\theta) \end{aligned} \tag{2.1}$$

where the angle of latitude θ can vary between 0 and π and the angle of longitude ϕ can vary between -π and π.

The rate of change of a function $f(x_1, x_2, \ldots, x_\mathrm{n})$ of n variables with respect to one of the variables x_i is referred to as a **partial derivative** of f with respect to x_i and is denoted by $\partial f(x_1, x_2, \ldots, x_\mathrm{n})/\partial x_\mathrm{i}$. It can be computed in the same way as a derivative of one variable, simply by holding all the other variables constant when forming the limit.

Let (θ, ϕ) be a set of parameters for a surface embedded in n-dimensional Euclidean space then the partial derivatives $\partial x_\mathrm{i}/\partial\theta$ and $\partial x_\mathrm{i}/\partial\phi$ of the ith coordinate of a point with parameters (θ, ϕ) specify the rate of change of the coordinate with respect to each of the parameters and are given by the limits:

$$\begin{aligned} \frac{\partial x_i}{\partial\theta} &= \frac{x_i(\theta+h,\phi)-x_i(\theta,\phi)}{h} \\ \frac{\partial x_i}{\partial\phi} &= \lim_{h\to 0}\frac{x_i(\theta,\phi+h)-x_i(\theta,\phi)}{h} \end{aligned} \tag{2.2}$$

So by making a small change in each parameter one can obtain a pair of vectors:

$$\begin{aligned} \frac{\partial\mathbf{x}}{\partial\theta} &= \left(\frac{\partial x_1}{\partial\theta}, \frac{\partial x_2}{\partial\theta}, \ldots, \frac{\partial x_\mathrm{n}}{\partial\theta}\right) \\ \frac{\partial\mathbf{x}}{\partial\phi} &= \left(\frac{\partial x_1}{\partial\phi}, \frac{\partial x_2}{\partial\phi}, \ldots, \frac{\partial x_\mathrm{n}}{\partial\phi}\right) \end{aligned} \tag{2.3}$$

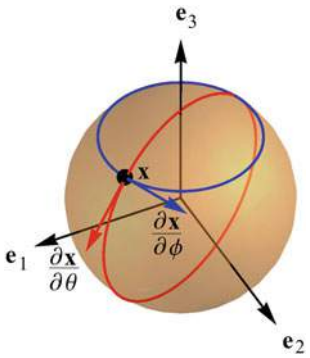


Fig. 2.8 Illustration of how tangent lines specified by partial derivatives of parameters can be used to span a tangent plane at a given point on a sphere. The partial derivative with respect to the latitude parameter θ is shown in red. This is tangent to the set of points, also shown in red, for which the longitude parameter is fixed. Similarly, the partial derivative with respect to the longitude parameter ϕ is shown in blue, as is the set of points for which the latitude parameter is fixed

Each of these n-tuples specify the direction of a tangent line through the point with parameter values (θ, ϕ) and the pair of tangent lines span the tangent plane to the surface at (θ, ϕ). This is illustrated for the sphere in Fig. 2.8. The procedure can be generalised to n-dimensional manifolds to define a **tangent space** at any point on the manifold, which has the same dimensionality as the manifold.

Given the tangent plane at every point on a surface, the length of a path between two points on a surface can be computed by using straight lines tangent to the path to approximate short segments of it and then adding the lengths of these line segments together. But to calculate the length exactly the segments have to be made infinitesimally small. For such displacements $(\mathrm{d}\theta, \mathrm{d}\phi)$ one can calculate the small changes $\mathrm{d}\mathbf{s}$ in the Euclidean coordinates by using the partial derivatives of the coordinates with respect to the parameters:

$$\mathrm{d}\mathbf{s} = \frac{\partial \mathbf{x}}{\partial \theta}\mathrm{d}\theta + \frac{\partial \mathbf{x}}{\partial \phi}\mathrm{d}\phi \tag{2.4}$$

The square of the length of this infinitesimal displacement is then given by the dot product of this change of coordinates vector with itself:

$$\|\mathrm{d}\mathbf{s}\|^2 = \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \theta}\mathrm{d}\theta^2 + 2\frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \phi}\mathrm{d}\theta\mathrm{d}\phi + \frac{\partial \mathbf{x}}{\partial \phi} \cdot \frac{\partial \mathbf{x}}{\partial \phi}\mathrm{d}\phi^2 \tag{2.5}$$

$$= \begin{pmatrix} \mathrm{d}\theta & \mathrm{d}\phi \end{pmatrix} \begin{pmatrix} \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \theta} & \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \phi} \\ \frac{\partial \mathbf{x}}{\partial \theta} \cdot \frac{\partial \mathbf{x}}{\partial \phi} & \frac{\partial \mathbf{x}}{\partial \phi} \cdot \frac{\partial \mathbf{x}}{\partial \phi} \end{pmatrix} \begin{pmatrix} \mathrm{d}\theta \\ \mathrm{d}\phi \end{pmatrix} \tag{2.6}$$

$$= \begin{pmatrix} \mathrm{d}\theta & \mathrm{d}\phi \end{pmatrix} \mathsf{G} \begin{pmatrix} \mathrm{d}\theta \\ \mathrm{d}\phi \end{pmatrix} \tag{2.7}$$

where the matrix G is referred to as the Riemannian metric or more succinctly as simply the **metric**. A manifold with a Riemannian metric defined on it is referred to a Riemannian manifold. In the case of the plane, the metric is simply equal to the identity matrix, corresponding to the dot product definition of length in Euclidean space.

A main use of the metric is in calculating the shortest path between two points on a surface. For two points on a plane the path corresponds to a straight line. For a curved surface one can locally determine the shortest path by calculating the distance in the tangent plane and a curve which locally follows the shortest path throughout its entire length is known as a **geodesic**.

On the sphere the geodesics are given by great circles, from which it is clear that a geodesic is not directly identical to the shortest path in Euclidean space, because for any pair of points that are not opposite each other there will be a long way round and a short way round along the geodesic. But provided that it is checked that when there is more than one geodesic path through two points, the shortest one is being used, the geodesic does correspond to the shortest path.

An embedding of the projective plane into Euclidean space that corresponds to the rotations of the eye is provided by the **Veronese surface** [2]. Let a point on a sphere be specified by a unit vector $\mathbf{v} = (v_1, v_2, v_3)$ then the coordinates of the corresponding point on an embedding of the Veronese surface in six-dimensional Euclidean space are given by the map χ:

$$\chi(v_1, v_2, v_3) = (v_1^2, v_2^2, v_3^2, \sqrt{2}v_2v_3, \sqrt{2}v_3v_1, \sqrt{2}v_1v_2) \tag{2.8}$$

Since $\chi(v_1, v_2, v_3) \cdot \chi(v_1, v_2, v_3) = (v_1^2 + v_2^2 + v_3^2)^2$ it follows from the definition of $\mathbf{v}$ as a unit length vector that every point on the Veronese surface is specified by a unit length vector. Hence the collection of points which form the Veronese surface lie on a sphere in six-dimensional Euclidean space.

Given the set Eqs. 2.8 defining an embedding of the projective plane one can go on to calculate the metric and the form of the geodesic curves. These turn out to correspond to rotations of the eye about a fixed axis [2]. Listing's law implies that with the line of fixation in the straight ahead position, the possible axes of rotation for movements made in accordance with Listing's law must all lie in the plane perpendicular to the straight ahead position. A further implication of Listing's law is that the corresponding plane, for an initial direction of the line of fixation away from the straight ahead position, is normal to the direction which bisects the angle between the straight ahead position and the initial direction of the line of fixation. A consequence of this half-angle rule is that a rotation about a fixed axis that starts with the line of fixation away from the primary position will trace out a circle referred to as a **direction circle** on a sphere surrounding the eye [4]. Some examples of direction circles are illustrated in Fig. 2.9.

For horizontal movements of the arm, embedding is straightforward because a disk can be embedded in 2-dimensional Euclidean space. When a human moves

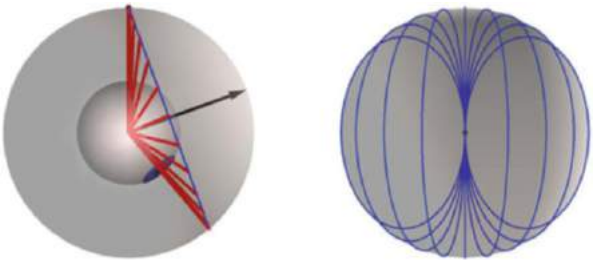


Fig. 2.9 Illustration of the half-angle rule that follows from Listing's law. **Left Figure** Let the initial position of the line of fixation lie in a horizontal plane and make an angle θ with the straight ahead direction. A vertical movement from this initial position made in accordance with Listing's law will consist of a rotation about an axis that makes an angle $\pi + \theta/2$ with the straight ahead direction. In the figure on the left the initial direction of the line of fixation is at an angle of 40° to the straight ahead direction in the horizontal plane and the corresponding axis of rotation, which is indicated by a black arrow, has moved back by 20°. Successive positions of the line of fixation, plotted in red, intersect with a sphere concentric with the eye to form a direction circle, that is plotted in blue. **Right Figure** The arcs of several direction circles, viewed from directly in front of the eye, when the axes of rotation are restricted to the horizontal plane. Because a half turn made in accordance with Listing's law always sweeps the line of fixation into the direction diametrically opposite to the straight ahead direction, the directions intersect at the point where the opposite direction intersects with the sphere surrounding the eye

an arm in the horizontal plane the trajectory of the movement is an approximately straight line [10] which is the geodesic associated with the Euclidean metric.

When the falling cat is initially released it has zero angular momentum and the conservation law requires that the angular momentum remains zero throughout the fall, but by changing its shape the cat is able to generate angular momentum about the $\mathbf{e}_3$ axis and this is balanced by angular movement of the body in the opposite direction. Making a backbend alone does not produce angular momentum about the $\mathbf{e}_3$ axis and making a roll alone only produces angular momentum about the $\mathbf{e}_3$ axis, which is in contradiction to the conservation law. However, when its body is bent the components of angular velocity about the $\mathbf{e}_2$ direction of the front and back halves of the cat cancel giving a net component of angular velocity about the roll axis.

Let the plane containing the two halves of the cat make an angle χ with the vertical. Then from the expression for the angular momentum of the system one can derive an equation relating the angular velocity of the shape changes of the cat to the angular velocity of the roll of the cat in space:

$$\frac{\mathrm{d}\chi}{\mathrm{d}t} = \frac{\mathrm{I_L}\sin(\frac{\psi}{2})}{\mathrm{I_P}\cos(\frac{\psi}{2})^2 + \mathrm{I_L}\sin(\frac{\psi}{2})^2}\frac{\mathrm{d}\theta}{\mathrm{d}t} \tag{2.9}$$

where $\mathrm{I_L}$ is the moment of inertia about the longitudinal axis and $\mathrm{I_P}$ is the moment of inertia about a perpendicular axis of one half of the model cat. For an actual cat an estimate of the ratio $\mathrm{I_L}/\mathrm{I_P}$ is 0.25 [6, 9].

The constraints described by Listing's law and the hinge joints at the elbow determine a relationship between the variables specifying the position of the eye and arm respectively whereas the constraints on the movement of the cat set up a relationship between the velocities of the variables specifying the shape of the cat. The constraints on the velocities result in some paths through the configuration space not being permitted. So a distinction is required between Riemannian manifolds in which all paths are possible and **sub-Riemannian** manifolds in which some paths are excluded.

In the case of the falling cat one can make a separation between the tangent vectors to the configuration space associated with rotation of the cat and those associated with change of its shape. The tangent vectors to the fibre of the configuration space correspond to roll of the whole cat and so are not permitted whereas the tangent vectors to the projective plane of the base manifold of the configuration space are allowed. So although the configuration space of the falling cat can be described by a manifold with a Riemannian metric given by the kinetic energy of the displacement of the cat, it is more accurately described as a sub-Riemannian manifold in which the tangent vectors that ensure conservation of angular momentum are separated from those that do not. In the context of sub-Riemannian geometry the tangent vectors allowed by the constraints are referred to as **horizontal directions** as opposed to the vertical directions, which are not allowed. One can still go to define a geodesic in the sub-Riemannian geometry as a path in the shape manifold with a minimum length, as defined by the kinetic energy metric, because by restricting the path to the horizontal directions one avoids the physically impossible paths.

Three alternative trajectories are illustrated in Fig. 2.10. Montgomery [9] introduced a trajectory which has the advantage that it works whatever values the moments of inertia take, although it requires unfeasible shape changes by the cat.

Rademaker and Ter Braak [13] introduced a trajectory in which the angle ψ between the two halves of the body is kept constant. However this trajectory only works for a particular value of the angle ψ and with an estimated ratio for the moments of inertia of 0.25 this angle is 120°. As can be seen from Fig. 2.10 when the cat is released with a realistic angle of 65° it fails to complete the half turn. From a biological perspective, the trajectory is incorrect because an actual cat bends forward to a much greater extent than it bends back.

The trajectory followed by a living cat is such that if it is initially held upside down with its spine bent forwards and then released, first it bends to one side, then backwards then to the other side and finally forwards again so that it has the same shape on landing as it did on release [6]. This trajectory across the projective plane is similar to that of the eye, as illustrated in Fig. 2.11. With respect to the front half of the cat, the back of the cat rotates around a fixed axis, as does the eye in the orbit. What is different is that the cat movement includes a backbend which requires that the axis of rotation is moved back further than with Listing's law.

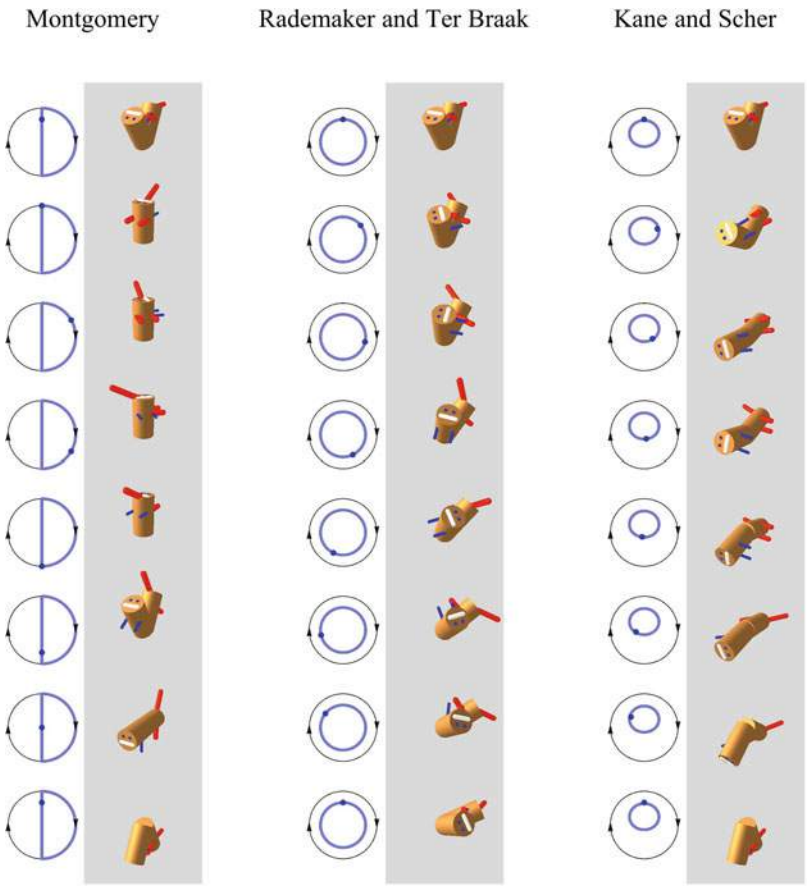


Fig. 2.10 Paths in shape space and the corresponding movement in physical space of three alternative trajectories for the cat righting movement. The Kane-Scher trajectory is closest to that actually used by the cat. Note that it is also the shortest of the three paths in the shape space

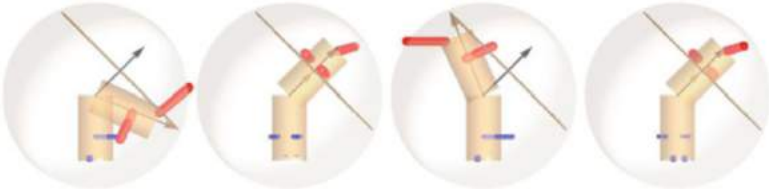


Fig. 2.11 Side view of the Kane-Scher trajectory drawn for comparison with the movement of the eye according to Listing's law illustrated in Fig. 2.9. The longitudinal axis of the front half of the cat has been held fixed corresponding to the constant position of the skull in the eye movement picture. Like the line of fixation of the eye, the longitudinal axis of the back half of the cat rotates around a fixed axis, indicated by a black arrow

2.3 Moving Around

Whilst it may be that discrete movements such as a saccade or a reaching movement can be described by a trajectory through a configuration space, there is no obvious way that this description will extend to more open-ended movements such as exploration of a new environment. Animals move about in their environment to find resources, so they have to be able both to locate the supplies and to be able to get to them. Braitenberg [1] introduced what he referred to as vehicles to illustrate the minimum capabilities required of a brain for it to be able to carry out such tasks. An example of a bilaterally symmetrical vehicle is shown in Fig. 2.12.

These conceptual vehicles provide a simple approach to thinking about how control of movement of the limbs might be extended to control of movement through an animal's environment. In the insects this is done by a process of **local navigation** which involves obstacle avoidance, homing with respect to the landmarks and path integration [5]. Avoidance behaviour is one of the simplest to implement. For instance, in the case of the Braitenberg vehicle, if the connection between the visual sensors and the motors are crossed then the vehicle will be attracted to light, but if they are uncrossed then it will turn away from a light source [1].

A more sophisticated navigation mechanism involves the animal keeping track of its heading direction and how far it has travelled in that direction so that it can always calculate the path back to its original position simply by inverting the sum of all the displacement vectors. In the continuous case, the animal keeps track of its velocity and the overall displacement is calculated by integrating the instantaneous velocity vector, a process referred to as **path integration** .

A key component of path integration is the maintenance of a heading direction vector that gives the instantaneous direction and speed of movement of the animal. Evidence for a mechanism for that maintains a heading direction signal has been found in the brain of the fly, which is illustrated in Fig. 2.13 [16]. The mechanism

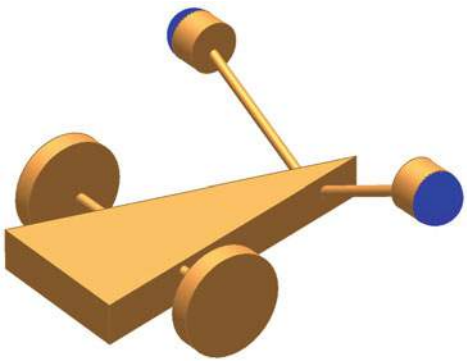


Fig. 2.12 A schematic diagram of the type of vehicle that Braitenberg [1] introduced. The vehicle consists of a pair of light sensors and a pair of wheels each of which is driven by its own motor

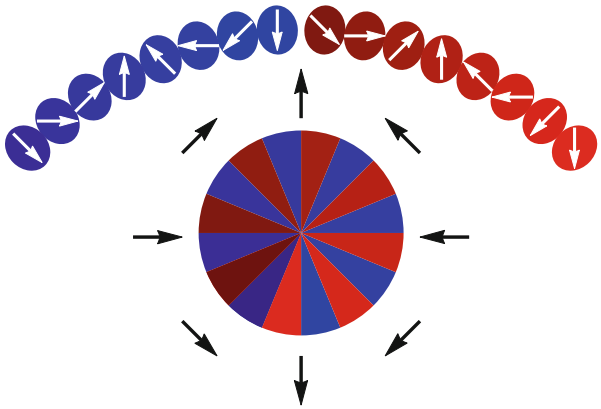


Fig. 2.13 Schematic diagram of the components of the heading direction mechanism of the fly which is comprised of the elongated protocerebral bridge and the ellipsoid body. Both components are made up of regions tuned to a particular heading direction of the fly, and the responses in the regions can be summed to make a population vector representation of heading direction. The protocerebellar bridge is distinguished by having two representations, one on each side

involves a region referred to as the central complex which includes the ellipsoid body and elongated protocerebral bridge both of which are organised into segments. The locus of neural activity in the ellipsoid body correlates with the heading direction of the fly. Associated with this locus of activity are a pair of corresponding regions of localised activity in the protocerebral bridge, one on each side. These bumps of activity are also tuned to the turning direction of the fly so that the activity on the left is greater when the fly is turning left and greater on the right when the fly is turning right.

The guidance circuit involves two classes of neurons that can be labelled E-P and P-E. Both have connections to the protocerebral bridge and ellipsoid body. But the E-P neurons receive input signals from the ellipsoid body and pass them to the protocerebral bridge, whereas the signals in the P-E neurons travel in the opposite direction. The E-P neurons from the active region of the ellipsoid body excite the pair of corresponding regions in the protocerebral bridge. However the projections to the ellipsoid body from the P-E neurons excited by these regions of the protocerebral bridge are offset with respect to the active region of the ellipsoid body, so that if the fly is turning right the active location in the ellipsoid body changes to a more rightward heading direction and similarly for turning left.

A model of these connections is given by a system of equations in which the activity of each region i of the ellipsoid body together with that of the pair of regions in the protocerebral bridge with the same orientation are described by a set of three equations. The first equation describes the activity in the ellipsoid body and the remaining two equations describe the activity in the two sides of the protocerebral bridge.

$$\begin{aligned}
\lambda \frac{\mathrm{d}x_\mathrm{i}}{\mathrm{d}t} &= -x_\mathrm{i} + 30(x_{\mathrm{n+i}} + x_{\mathrm{n+L_i}} + x_{\mathrm{2n+i}} + x_{\mathrm{2n+R_i}})^+ \\
\mu \frac{\mathrm{d}x_{\mathrm{n+i}}}{\mathrm{d}t} &= -x_{\mathrm{n+i}} + (10x_\mathrm{i} - \tfrac{125}{\mathrm{n}} \textstyle\sum_{\mathrm{j}=1}^{\mathrm{n}} x_\mathrm{j} + 1 + \omega^+)^+ \\
\mu \frac{\mathrm{d}x_{\mathrm{2n+i}}}{\mathrm{d}t} &= -x_{\mathrm{2n+i}} + (10x_\mathrm{i} - \tfrac{125}{\mathrm{n}} \textstyle\sum_{\mathrm{j}=1}^{\mathrm{n}} x_\mathrm{j} + 1 + (-\omega)^+)^+
\end{aligned} \tag{2.10}$$

where ω is the angular velocity of the fly, λ and μ are time constants equal to 80 and 65 ms respectively, and L and R are one-dimensional matrices that describe the offset of the connections from the protocerebral bridge to the ellipsoid body. With eight such regions, as illustrated in Fig. 2.13 and used in the simulations shown in Fig. 2.14 $\mathsf{L} = (2,3,4,5,6,7,8,1)$ and $\mathsf{R} = (8,1,2,3,4,5,6,7)$. These equations are based on a more complete model that was shown to replicate the experimental findings of the changes in the activity in the central complex [16].

In rats the heading direction mechanism is combined with circuitry in the hippocampus and neighbouring subiculum and entorhinal cortex that forms a map of the surroundings of the animal. Heading direction neurons whose firing is tuned to a particular heading direction of movement through the environment [15] are found in the entorhinal cortex. But the hippocampal circuits have several other classes of neurons with distinct properties.

Place cells in the hippocampus of rats fire when the animal is in a specific location [11]. The neighbouring **grid cells** in the entorhinal cortex of a rat fire when animal arrives at the vertices of a lattice of equilateral triangles [3], although the positions of the vertices varies from neuron to neuron. A key difference between place cells and grid cells is that whereas different sets of place cells fire in different environments, the vertex spacing and orientation relations between grid cells remains the same. A further class of neurons consist of the boundary cells which fire when the animal is close to a boundary and these are found in both the subiculum and enthorinal cortex [14]. Another form boundary cells, referred to as boundary vector cells, are tuned to respond to a boundary at a specific distance and orientation with respect to the animal [7]. These are found in the subiculum.

Our understanding of the workings of the hippocampal circuitry is still at an early stage, but it offers tantalising insights into how data is organised deeper into the brain. The discovery of the different classes of neurons in the rat should be put in the context that the hippocampus in rats is much larger, relative to the whole brain, than in humans. The hippocampus is also predominantly involved with spatial navigation in rats whereas in humans only the posterior part is linked to navigation and the structure as a whole appears to be more involved with setting up and recalling memories. This finding has led to the hypothesis that the hippocampus in humans builds a representation of a space of memories just as it does for everyday space, but the nature of this space remains to be clarified [8].

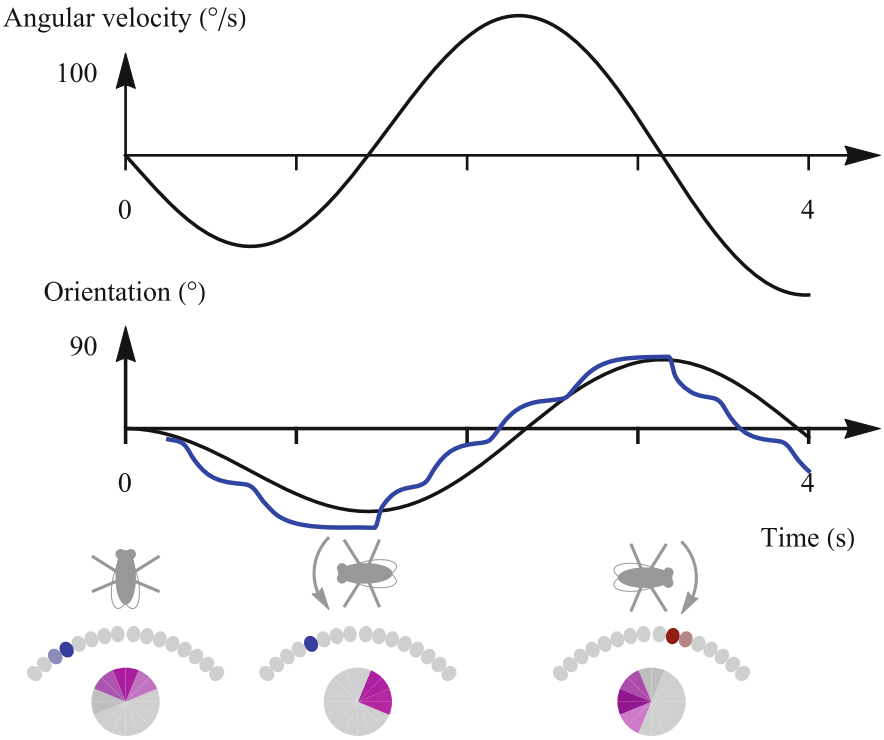


Fig. 2.14 Simulation of how the representation of heading direction by the active region of the ellipsoid body is changed to match the outcome of turning movement of the fly produced by Eqs. 2.10. In this example the fly begins in the straight ahead and then turns left and then right, as indicated by the fly icons drawn in grey. The angular velocity and orientation of the fly are plotted in black, whilst the population vector representation of orientation computed from the output of the model is plotted in blue. The simulated orientation is jerky because of the small number of neurons used in the simulation

References

1. Braitenberg, V.: Vehicles: Experiments in Synthetic Psychology. MIT Press, Cambridge (1984)
2. Clement, R.: Eye movement space. Math. Intell. **38**, 8–13 (2016). https://doi.org/10.1007/s00283-015-9576-3
3. Fyhn, M., Molden, S., Witter, M., Moser, E., Moser, M.B.: Spatial representation in the entorhinal cortex. Science **305**, 183–196 (2004). https://doi.org/10.1126/Science.1099901
4. Helmholtz, H.: Treatise on Physiological Optics. Translated by Southall, J.P.C. Dover, New York (1910)
5. Hübner W., Mallot, H.A.: Metric embedding of view graphs. Auton. Robot. **23**, 183–196 (2007). https://doi.org/10.1007/s10514-007-9040-0
6. Kane, T.R., Scher, M.P.: A dynamical explanation of the falling cat phenomena. Int. J. Solids Struct. **5**, 663–670 (1969). https://doi.org/10.1016/0020-7683(69)90086-9

7. Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., Burgess, N.: Boundary vector cells in the subiculum of the hippocampal formation. J. Neurosci. **29**, 9771–9777 (2009). https://doi.org/10.1523/JNEUROSCI.1319-09.2009
8. Mok, R.M., Love, B.C.: A non-spatial account of place and grid cells based on clustering models of concept learning. Nat. Commun. **10**(1), 1–9 (2019). https://doi.org/10.1038/s41467-019-13760-8
9. Montgomery, R.: Gauge theory of the falling cat. In: Enos, M.J. (ed.) Fields Institute Communications. Dynamics and control of mechanical systems: the falling cat and related problems, pp. 195–281. Fields Institute Communications, New York (1993). https://doi.org/10.1090/fic/001
10. Morasso, P.: Spatial control of arm movements. Exp. Brain. Res. **42**, 223–227 (1981). https://doi.org/10.1007/BF00236911
11. O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. **34**, 171–175 (1971). https://doi.org/10.1016/0006-8993(71)90358-1
12. Polpitiya, A.D., Dayawansa, W.P., Martin, C.F., Ghosh, B.K.: Geometry and control of eye movements. IEEE Trans. Autom. Control **52**, 170–180 (2007). https://doi.org/10.1109/TAC.2006.887902
13. Rademaker, G.G.J., Ter Braak, W.G.: Das Umdrehen der fallenden Katze in der Luft. Acta. Oto-Laryngol. **23**, 313–343 (1936). https://doi.org/10.3109/00016483609123230
14. Solstad, T., Boca, C., Kropff, E., Moser, M.B., Moser, E.: Representation of geometric borders in the entorhinal cortex. Science **322**, 1865–1868 (2008). https://doi.org/10.1126/science.1166466
15. Taube J.S., Muller, R.U., Ranck, J.B.: Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J. Neurosci. **10**, 420–435 (1990). https://doi.org/10.1523/JNEUROSCI.10-02-00420.1990
16. Turner-Evans, D., Wegener, S., Rouault, H., Franconville, R., Wolff, T., Seelig, J.D., Druckmann, S., Jayaraman, V.: Angular velocity integration in a fly heading direction. eLife **6**, 1–39 (2017). https://doi.org/10.7554/eLife.23496

Chapter 3
Measurements

Given that it is possible to describe a biological object in terms of a manifold and a vector field, the brain still has the problem of building the description using only measurements. Colour measurements are used to illustrate that data can be simplified by applying principal components analysis, resulting in a partition of the data according to how the measurements vary with each other. The neural implementation of this procedure corresponds to Hebbian learning and the complementary procedure of anti-Hebbian learning is also introduced using the cerebellum as an example.

3.1 Linear Transformations

Whilst the movements of biological objects can be described in terms of trajectories through configuration spaces, and furthermore such descriptions can be used in specifying the movements, it remains to be shown that the brain can make use of such descriptions when interacting with other animals. The task for the brain is to take physically measurable variables and identify the animals present and the configurations of their bodies.

An extensively investigated example of a measurement made by the visual system is that of colour. In a typical colour matching experiment the subject views a small stimulus which is usually restricted to around 2° in diameter so that only the fovea is stimulated. The field is split into two halves and a pair of lights A and B, presented in one half, have to be matched in colour to a pair of lights C and D, presented in the other half, as illustrated in Fig. 3.1. The main finding of colour matching experiments, referred to as the trichromatic law of colour matching, is that the colour of any stimulus can be matched by an additive mixture of three fixed primary colours [29].

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_3

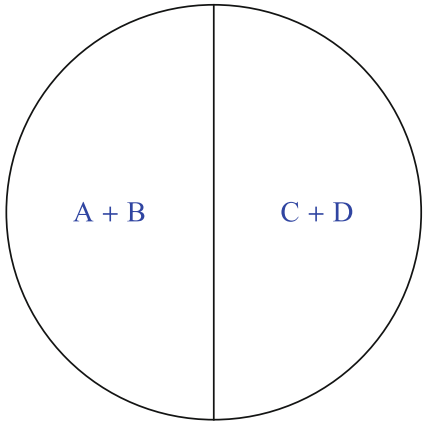


Fig. 3.1 Presentation of the stimuli in a colour matching experiment. The light C is specified by the experimenter. The task for the subject is to alter the intensity of the lights A, B and D until both sides of the presentation appear to have the same colour. In the simplest case the lights are monochromatic and by choosing successive lights from the electromagnetic spectrum for C, the colour matching functions plotted in Fig. 3.2 can be derived

The experimental findings can be described geometrically by treating each colour stimulus as a vector, with the direction of the vector representing the qualitative nature of the colour and the size of the vector representing the amount of the colour. The vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are given by the functions of wavelengths $\mathbf{v}_1(\lambda)$, $\mathbf{v}_2(\lambda)$ and $\mathbf{v}_3(\lambda)$ respectively, which describe the spectral power distribution of the fixed primary light sources. In vector notation, the trichromatic law states that any colour $\mathbf{c}$ can be matched by a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$.

$$\mathbf{c} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 \tag{3.1}$$

If one of the coordinates c_1, c_2 or c_3 is negative, then one can move the term involved to the other side of the equation, to obtain a description of the physical situation where one pair of colours is matched by another pair.

In 1931 a standard colorimetric observer was defined by the 'Commission Internationale de l'Éclairage' (abbreviated to CIE). The experimental results from many colour matching experiments were combined by expressing them all in terms of colour matches with three monochromatic light sources **r**, **g** and **b**, located at wavelengths of 700, 546, and 435.8 nanometres respectively. The relative amounts of the sources were set so that a colour mixture made by equal amounts of **r**, **g** and **b** appears white. The coordinates $r(\lambda)$, $g(\lambda)$ and $b(\lambda)$ required to match a light with a unit radiant power at wavelength λ are functions of λ referred to as colour matching functions. The colour matching functions which form the specification of the C.I.E. observer are plotted in Fig. 3.2.

The C.I.E. also constructed an **xyz** space which simplifies the representation of colour. Negative values in the color matching functions were eliminated by expressing the colour matching functions in terms of hypothetical primary lights referred to as **x**, **y** and **z** which were also chosen so that a linear combination of equal amounts of the primary lights appears white. Further simplification was achieved by selecting the lights so that the **y** light matched the apparent luminance of a colour

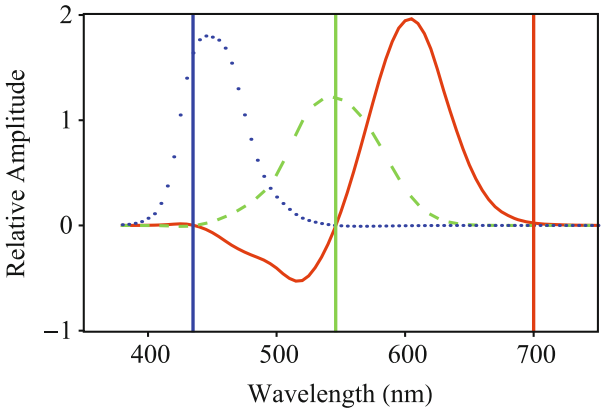


Fig. 3.2 Average colour matching functions obtained with **r**, **g** and **b** monochromatic lights. $r(\lambda)$—red, continuous line, $g(\lambda)$—green, broken line, and $b(\lambda)$—blue, dotted line. The wavelengths of the **r**, **g** and **b** lights are indicated by red, green and blue vertical lines. An individual's colour matching functions will depend on both the viewing conditions and the viewer

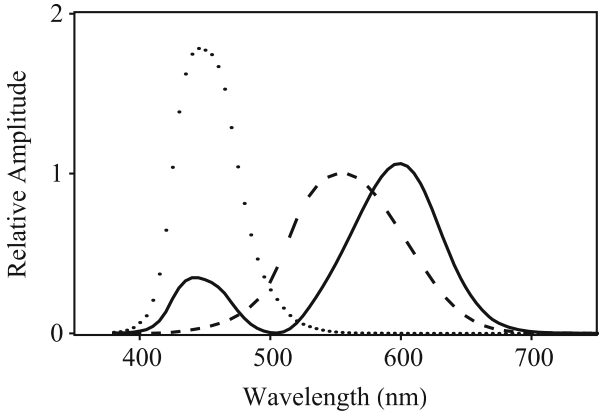


Fig. 3.3 Color matching functions of **xyz** colour space. $x(\lambda)$—continuous line, $y(\lambda)$—broken line, and $z(\lambda)$—dotted line

and the **x** and **z** lights made no contribution to the luminance match. The colour matching functions in this space are plotted in Fig. 3.3.

The transformation from **rgb** to **xyz** colour space was specified by a set of linear equations:

$$\begin{aligned} \mathbf{x} &= 2.3649\mathbf{r} - 0.5156\mathbf{g} + 0.0052\mathbf{b}, \\ \mathbf{y} &= -0.8971\mathbf{r} + 1.4273\mathbf{g} - 0.0144\mathbf{b}, \\ \mathbf{z} &= -0.4678\mathbf{r} - 0.0883\mathbf{g} + 1.0092\mathbf{b}, \end{aligned} \tag{3.2}$$

The transformation between colour spaces is an example of a linear transformation because the **xyz** basis vectors are specified by linear functions of the **rgb** basis vectors. For a **linear transformation** T it holds that for every vector $\mathbf{p}$ and every vector $\mathbf{q}$:

$$\mathsf{T}(\mathbf{p}+\mathbf{q}) = \mathsf{T}(\mathbf{p}) + \mathsf{T}(\mathbf{p}) \tag{3.3}$$

and if a is a constant used to multiply the length of a vector then:

$$\mathsf{T}(\mathrm{a}\mathbf{p}) = \mathrm{a}\mathsf{T}(\mathbf{p}) \tag{3.4}$$

The coefficients of a linear combination can be written as a row of numbers, and the coefficients of a system of linear equations can be arranged into an m row by n column array of numbers referred to as a **matrix**. A matrix is denoted by a single upper case letter such as A and an individual number in the array is referred to as an element and is denoted by a lower case letter such as a. The element in the ith row and jth column of a matrix A is specified by adding subscripts to the label for the element.

The product of two matrices A and B is denoted putting the capital letters in sequence AB and is defined so that a system of linear equations can be written more succinctly as a matrix equation. The rule is that the element in the ith row and jth column of the product is equal to the sum of the products of successive elements in the ith row the first matrix with successive elements in the jth column of the second matrix. This rule can be written more succinctly using the sigma notation as:

$$ab_{ij} = \sum_{k=1}^{\mathrm{n}} a_{ik} b_{kj} \tag{3.5}$$

where n is the number of elements in the rows of the first matrix and also the number of elements in the columns of the second matrix. Applying the matrix notation, the description of the linear transformation from **rgb** to **xyz** colour space can be written as:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} 2.3649 & -0.5156 & 0.0052 \\ -0.8971 & 1.4273 & -0.0144 \\ -0.4678 & -0.0883 & 1.0092 \end{pmatrix} \begin{pmatrix} \mathbf{r} \\ \mathbf{g} \\ \mathbf{b} \end{pmatrix} \tag{3.6}$$

If the linear transformation describes a change of the basis of a vector space then the matrix of the transformation will have the same number of rows and columns. Matrices with this property are referred to as **square** matrices .

The colour matching functions can be directly related to the properties of the three classes of light sensitive cones in the retina, by a linear transformation from the **xyz** space to the cone space. The three classes of cones have peaks in their spectral sensitivities at long, medium and short wavelengths and will be denoted

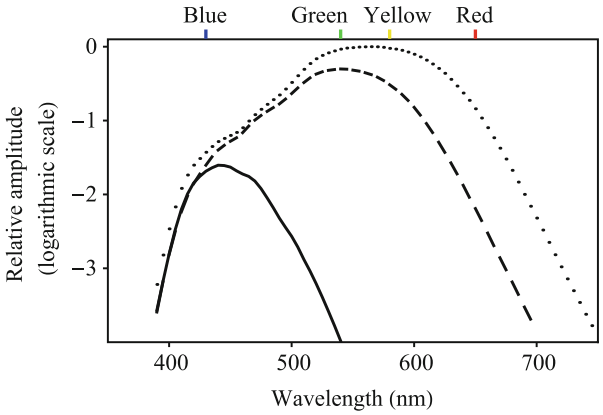


Fig. 3.4 Cone spectral sensitivities of the **l** (dotted line), **m** (dashed line) and **s** (continuous line) cones. B, G, Y and R correspond to the wavelengths which appear blue, green, yellow and red respectively

by **l**, **m** and **s** respectively. Estimates of the spectral sensitivities of the cones were initially derived from colour matching experiments with subjects who lacked one of the cone types [25, 28] and were corroborated by subsequent direct physiological measurement of the cone sensitivities. Subsequently these estimates have been revised to take into account individual differences in the genotypes [26]. The ratios of the maximum sensitivities of the **l**, **m** and **s** cones have been estimated as 1: 0.5 : 0.025 respectively, and these ratios reflect the relative number of the different types of cone in the retina [3]. The spectral sensitivities of the cones are plotted in Fig. 3.4

As the colour data can be described by coordinates of points in any number of sets of three basis vectors, the first step in analysing the data involves finding a set of basis vectors that simplifies the task of making use of the data. This can be done by assuming that a linear transformation T turns a set of basis vectors into the required set and then changing the basis vectors so that T becomes the identity transformation that does not change anything. The matrix of the **identity** transformation **I** has the diagonal form:

$$\mathsf{I} = \begin{pmatrix} 1\,0\,.\,0 \\ 0\,1\,.\,0 \\ .\quad .\quad .\quad . \\ 0\,0\,.\,1 \end{pmatrix} \tag{3.7}$$

One can take a linear transformation defined by a square matrix T, which turns a set of coordinates U into a set of coordinates V, and combine it with a change of basis matrix S, that turns the sets of coordinates X and Y into U and V respectively.

Applying these transformations in sequence gives:

$$\mathsf{SY} = \mathsf{TSX} \tag{3.8}$$

so

$$\mathsf{Y} = \mathsf{S}^{-1}\mathsf{TSX} = \mathsf{PX} \tag{3.9}$$

where S^{-1} stands for the **inverse** of the matrix S which has the property that multiplication of a square matrix by its inverse gives the identity matrix I.

The matrices P and $\mathsf{S}^{-1}\mathsf{TS}$ both describe the same linear transformation and are referred to as **similar** matrices. The key to finding the simplest diagonal form of the linear transformation T involves finding a type of vector referred to as an **eigenvector**, which changes size but not direction under the transformation:

$$\mathsf{T}(\mathsf{u}) = \lambda\mathsf{u} \tag{3.10}$$

where λ is a constant specifying the change in size that is referred to as the **eigenvalue**. If one can find n different eigenvalues, where n is the number of basis vectors, then one can write them into the rows of the matrix S, and apply the definition of an eigenvector given in Eq. 3.10 to obtain the equivalence relation:

$$\mathsf{TS} = \Lambda\mathsf{S} \tag{3.11}$$

where Λ is a matrix with all elements equal to zero except for the diagonal elements which are equal to the eigenvalues of T.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & . & 0 \\ 0 & \lambda_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \lambda_n \end{pmatrix} \tag{3.12}$$

from which it follows that $\mathsf{S}^{-1}\mathsf{TS}$ must be a diagonal matrix.

This diagonalisation procedure reveals the simplest description of a linear transformation and will be applied in the next section.

3.2 Principal Components Analysis

Every pth stimulus generates responses in a collection of n neurons. This n-tuple of responses $(v_1^p, v_2^p, \ldots, v_n^p)$ can be used to specify the coordinates of a vector $\mathbf{v}^p$ in a vector space referred to in this context as a **data space**. Given m measurement

n-tuples, the **mean** of the ith coordinate v_i is defined as:

$$\text{mean}(v_i) = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} v_i^p \tag{3.13}$$

and the **variance** of the ith coordinate v_i is defined to be:

$$\text{variance}(v_i) = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} (v_i^p - \text{mean}(v_i))^2 \tag{3.14}$$

Similarly the **covariance** between coordinate v_i and coordinate v_j is defined as:

$$\text{covariance}(v_{ij}) = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} (v_i^p - \text{mean}(v_i)) \times (v_j^p - \text{mean}(v_j)) \tag{3.15}$$

Analysis of the data can often be simplified by subtracting the mean of each coordinate from every data point. For mean-subtracted data mean $(v_i) = 0$ for every coordinate, so the variance becomes:

$$\text{variance}(v_i) = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} v_i^p \times v_i^p = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} (v_i^p)^2 \tag{3.16}$$

and the covariance becomes:

$$\text{covariance}(v_{ij}) = \frac{1}{\text{m}} \sum_{p=1}^{\text{m}} v_i^p \times v_j^p \tag{3.17}$$

Let the mean-subtracted data be represented by a matrix V that is formed by writing the coordinates of successive data vectors into successive rows.

$$\mathsf{V} = \begin{pmatrix} v_1^1 & v_2^1 & . & v_n^1 \\ v_1^2 & v_2^2 & . & v_n^2 \\ . & . & . & . \\ v_1^{\text{m}} & v_2^{\text{m}} & . & v_n^{\text{m}} \end{pmatrix} \tag{3.18}$$

The **transpose** of a matrix is obtained by writing successive rows of the matrix into successive columns of the new matrix, and is denoted by superscript $^\mathsf{T}$. By using the transpose operation the covariances of the mean-subtracted data can be conveniently represented by an n row by n column covariance matrix given by $\frac{1}{\text{m}}\mathsf{V}^\top\mathsf{V}$ since for

each element of the covariance matrix it holds that:

$$\frac{1}{\mathrm{m}}(\mathsf{V}^\top \mathsf{V})_{ij} = \frac{1}{\mathrm{m}} \sum_{p=1}^{\mathrm{m}} v_i^p \times v_j^p = \mathrm{covariance}(v_{ij}) \tag{3.19}$$

The covariance matrix is **symmetric** in that for every element c_{ij} it holds that $c_{ij} = c_{ji}$.

Now any symmetric matrix can be used to define a quadric surface in Euclidean space. A **quadric surface** is defined by a quadratic equation. The surface can be described by a matrix equation of the form

$$\mathsf{X}^\top \mathsf{A}\mathsf{X} + \mathsf{B}^\top \mathsf{X} + \mathrm{c} = 0 \tag{3.20}$$

where A is a n $\times$ n symmetric matrix, B is a $n \times 1$ matrix, c is a constant and the transposition operations are necessary to ensure that the matrix products can be computed.

The simplest equations for quadric surfaces occur when A is a diagonal matrix and the matrix B is zero so that the surface is centered at the origin. Written out in full the quadratic equation is:

$$a_{11}x_1^2 + a_{22}x_2^2 + \ldots + a_{\mathrm{nn}}x_{\mathrm{n}}^2 + \mathrm{c} = 0 \tag{3.21}$$

and some examples of surfaces described by this equation when c is a positive number are illustrated in Fig. 3.5.

If the matrix A contains nonzero elements off the diagonal then the quadric surface will be rotated with respect to the basis vectors as illustrated in Fig. 3.6. The representation of the surface by a diagonal matrix can be recovered by applying an inverse rotation to the coordinates.

A set of vectors are **orthogonal** if the inner product between every possible pair of them is zero. In two dimensions this corresponds to the vectors being perpendicular to each other. If the set of orthogonal vectors are also all of unit length, as is the case for the basis vectors of Euclidean space, then they are referred to as orthonormal. Linear transformations that preserve the orthogonality of vectors, as is the case for transformation between Euclidean spaces, are referred to as orthogonal transformations and the corresponding matrices as orthogonal matrices.

For the simplest case of rotation of points in two-dimensional Euclidean space through an angle θ, as shown in Fig. 3.7, the corresponding orthogonal matrix R_θ is:

$$\mathsf{R}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{3.22}$$

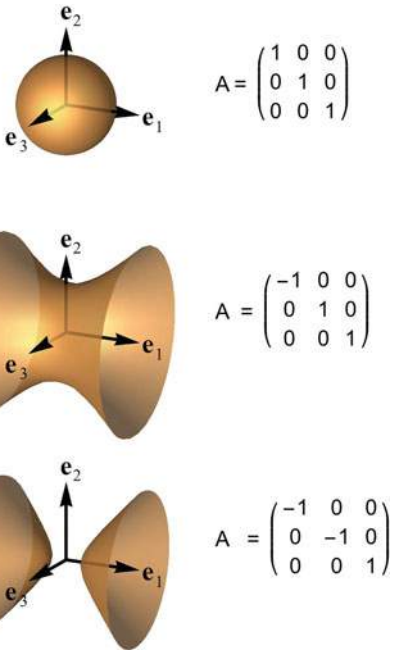


Fig. 3.5 Examples of central quadrics. **Upper figure**. Sphere. **Middle figure** Hyperboloid of one sheet. **Lower figure** Hyperboloid of two sheets

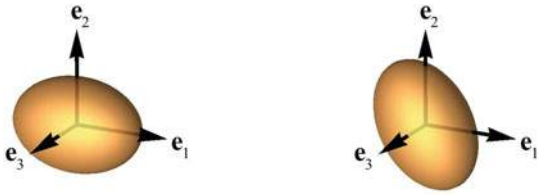


Fig. 3.6 Rotation of a quadric surface associated with nonzero elements off the diagonal of the matrix A

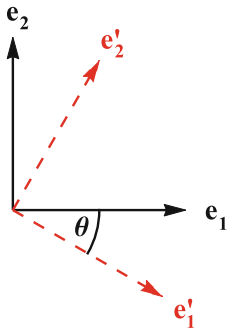


Fig. 3.7 Clockwise rotation of a pair of Euclidean basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ through an angle θ into the new pair of Euclidean basis vectors $\mathbf{e}'_1$ and $\mathbf{e}'_2$

This matrix describes a clockwise rotation of the basis vectors. For an anticlockwise rotation $R_{-\theta}$ the matrix is:

$$\mathsf{R}_{-\theta} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \mathsf{R}_{\theta}^{\top} \tag{3.23}$$

As the anticlockwise rotation is the inverse of the clockwise rotation, this result illustrates a distinctive property of orthogonal matrices which is that their transpose is equal to their inverse.

$$\mathsf{R}\mathsf{R}^{-1} = \mathsf{R}\mathsf{R}^{\top} = \mathsf{I} \tag{3.24}$$

The diagonalisation procedure introduced with similar matrices can be applied to the matrix of coefficients **A** of a quadric form to eliminate the off-diagonal coefficients. The required rotation matrix can be found by forming a matrix of the eigenvectors of the matrix **A** that defines the quadric surface. For example, in the case of the matrix used to illustrate the rotated quadric surface in Fig. 3.6.

$$\mathsf{A} = \begin{pmatrix} \frac{7}{4} & \frac{\sqrt{3}}{4} & 0 \\ \frac{\sqrt{3}}{4} & \frac{5}{4} & 0 \\ 0 & 0 & 3 \end{pmatrix} \tag{3.25}$$

The matrix of unit length eigenvectors **S** is:

$$\mathsf{S} = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.26}$$

so $\mathsf{S}^{\top}\mathsf{A}\mathsf{S}$ can be calculated to be:

$$\begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{7}{4} & \frac{\sqrt{3}}{4} & 0 \\ \frac{\sqrt{3}}{4} & \frac{5}{4} & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \tag{3.27}$$

Diagonalisation of the covariance matrix of a set of data answers the question of how to choose an appropriate set of basis vectors for data. The matrix of the n eigenvectors of the symmetric covariance matrix is an orthogonal matrix. When the orthogonal transformation is applied to the data then the covariance matrix becomes diagonal. The transformation of the data so that the coordinates of the first basis vector have the largest amount of variance, the coordinates on the second basis vector have the next largest amount of variance and so on, is referred to as **principal components analysis**.

An example of the application of principal components analysis by the nervous system is provide by the early stages of colour processing. A range of surface reflectance functions that approximately corresponds to those found in everyday life is provided by the Munsell chip set. This is a set of coloured chips that have been selected so that the changes between neighbouring chips of hue, saturation and lightness appear equal. Measurements of the reflectance functions have been made at five nanometre intervals [17] and tabulated values of the cone sensitivities at 5 nm intervals have also been made available [26] so the response of a cone can be approximated by the sum.

$$\text{response of cone} = \sum_{i=0}^{74} v(380 + 5i)c(380 + 5i) \tag{3.28}$$

where the range 380 to 750($= 380 + 5 \times 75$) nm is the range over which the cones respond. To be equivalent to the integrated response each element of the sum should be multiplied by the 5 nm interval between samples, but inclusion of this scale factor does not affect the subsequent analysis of the data.

The effect of diagonalisation of the covariance matrix of the cone responses is illustrated in Fig. 3.8. The eigenvalues of the covariance matrix are in the ratio 33033: 94: 1, showing that by far the largest component of the variance is associated with the first eigenvector, which acts as a light-dark filter. The next largest component of covariance is provided by the second eigenvector which acts an opponent red-green filter and the third eigenvector corresponds to blue-yellow opponent filter [4]. These are plotted in Fig. 3.9.

A technique for measuring the weighting of the light-dark filter involves adjusting the relative brightness of two alternately flickering lights so that the apparent flicker disappears. Two different alternately flickering hues merge into a single apparent hue at a low flicker frequency, but if they differ in brightness they will

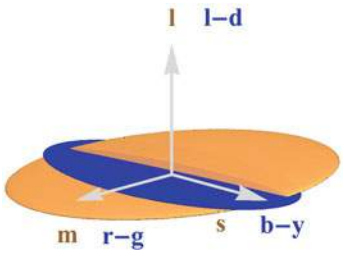


Fig. 3.8 Illustration of the ellipsoid defined by the covariance matrix of the cone responses to the Munsell chip set, before and after diagonalisation. Note that the elements of the matrix A defining the quadric surface have been taken directly from the covariance matrix so a large covariance leads to a flattening of the ellipsoid in that direction. In statistics presentations the reciprocal of the covariance is usually taken in the definition of the ellipsoid so that a large variance gives an elongation of the ellipsoid in that direction

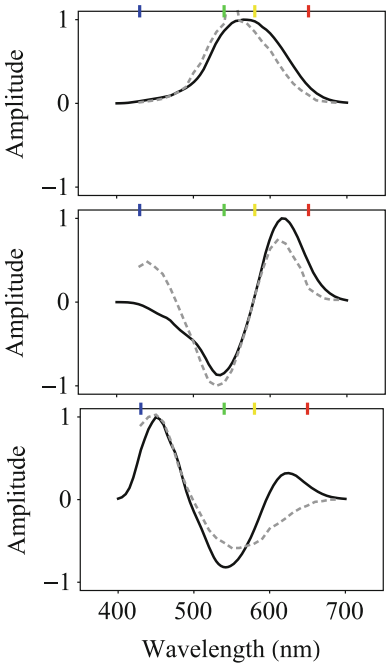


Fig. 3.9 Basis vectors for approximating cone responses to the Munsell chip set. The eigenvectors of the principal components are plotted as continuous lines and the experimental data are plotted as dashed lines

continue to flicker at a higher frequency because lights differing in brightness fuse at a higher frequency. The weighting of the red-green opponent-colour filters has been investigated experimentally by a hue cancellation technique. This technique involves taking a light that produces a red sensation and adding a light that produces a green sensation until a stimulus is produced that appears neither red nor green to the observer. By carrying out this procedure with a range of monochromatic lights the weighting of the opponent colour filter can be determined experimentally and a similar procedure can be applied to the blue-yellow filter [11]. The experimentally determined weightings are plotted as dashed lines in Fig. 3.9.

3.3 Hebbian Learning

The drawback of population coding is that many neurons fire in response to a single stimulus and this leaves the brain with the problem of how to tell that all these responses are due to the same stimulus. This problem can be overcome by a neural mechanism referred to as **Hebbian learning** after its originator [10]. The idea is captured by the succinct description that "neurons wire together if they fire together" [13]. This behaviour can be incorporated in the model neuron by continuously changing the weights according to the formula:

$$\Delta w_i = \eta x_i y \tag{3.29}$$

where η is a learning rate parameter. But with this formulation the weights can grow without limit so an additional specification is required to ensure this does not happen. If the weight vector is divided by its length after each change the learning rule becomes

$$\Delta w_i = \eta y(x_i - y w_i) \tag{3.30}$$

and this modification ensures that the weight vector retains its unit length. In effect, growth of the term $x_i y$ that determines the weight change is countered by the feedback term $y w_i$ [15]. The rule for implementing Hebbian learning given in Eq. 3.30 is closely related to principal components analysis in that after the learning has finished the weights of the neuron will be equal to the eigenvector with the largest eigenvalue. Furthermore, the rule can be extended to more than one neuron. If the outputs of m neurons are indexed by the variable j then the extended rule is:

$$\Delta w_{ij} = \eta y_j (x_i - \sum_{k=1}^{i} y_k w_{ki}) \tag{3.31}$$

and with this learning rule, the weights associated with successive neurons correspond to the successive eigenvectors of the covariance matrix of the input [21].

Whilst it is easiest to understand the components in the data if they are listed in order of the amount of variation in the data which they account for, allowing each neuron to signal a linear combination of the eigenvectors has the advantage that loss of some of the neurons will result in a degradation of the signalling of the data, but not a loss of one or more of the components. So in the case of colour vision, loss of some of the colour opponent neurons would be expected to result in degraded red/green and blue/yellow discrimination but not loss of one type of discrimination only. Evidence that neurons carry a mixture of the components comes from principal components analysis of colour sensitive neurons in the thalamus. The variation of the responses could be accounted for by three factors, corresponding to the light/dark, red/green and blue/yellow components, but it was found that individual neurons responded to a linear combination of the components [30].

Although the assumption that the cone numbers ratios l:m:s equal 1:0.5:0.025 is valid for an average over many subjects, individuals show great variability in their cone ratios, with l:m ratios ranging between 10:1 and 1:2. If the opponent colour channels were fixed linear combinations of the cone signals, then one would expect the wavelength corresponding to the percept of unique yellow, which appears neither reddish or greenish to vary with the cone ratios [5]. However experimental investigations have revealed no correlation between l:m ratios and the wavelength of the unique yellow, which varies little from 580 nm [14]. The conclusion drawn from these experiments was that a mechanism of neural plasticity is involved in setting the relative weights of the cones in the opponent colour channels. The Hebbian learning mechanism of principal components analysis explains the invariance of the unique yellow, because the predicted zero crossings of the red-green basis vector with l:m

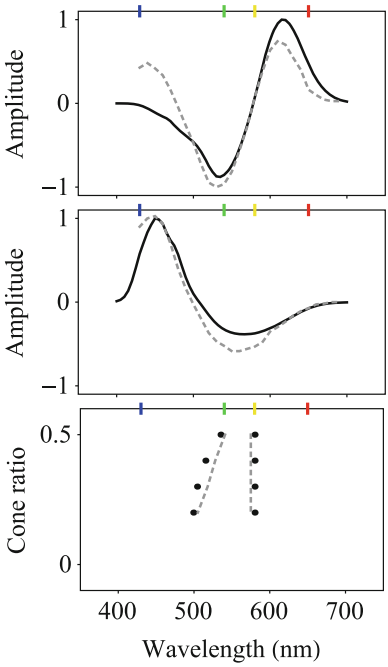


Fig. 3.10 Illustration of the effect of a 4.5 ° rotation of the opponent color basis vectors. **Upper Figure** Red-green opponent colour matching function after rotation of the basis vectors. **Middle Figure** Blue-yellow opponent colour matching function after rotation of the basis vectors. The opponent colour filter now matches more closely that measured experimentally. **Lower Figure** With rotation of the opponent-colour channels the location of the unique green changes with the ratio m/(l+m) of l and m cones, as is found experimentally [23]. In all the graphs the experimental data are shown by dashed lines

ratios of 10:1, 2:1 and 1:2 are at wavelengths of 576, 578, and 578 nanometres respectively.

Unlike unique yellow, unique green does change wavelength with changes in the l:m cone ratio. A possible explanation for this is that the principal components are rotated so that the blue-yellow sensitivity function more closely matches that found experimentally as illustrated in Fig. 3.10.

Often the data will consist of successive samples from a spatial or temporal continuum. In the case of the colour data, the samples were taken from successive points along the electromagnetic spectrum. The distance separating the position of the first and last sample defines the length of the sampling window, and the results of principal components analysis depends on both the covariance structure of the data and the length of the window used to sample the data.

The effect of the window length is easiest to appreciate when the covariance between sampling points depends only on the distance between the sampling points and not on the selection of samples used in the calculation, as is the case for instance for photographic images of natural scenes. If the window is small the successive eigenvalues correspond to filters that compute successive derivatives of the data, and ideally will match the best fixed window length filters that can be made for calculating successive derivatives. These can be described by discrete Legendre functions [8] and Fig. 3.11 shows that principal components analysis of data taken from a one-dimensional strip of photographs of natural images [9] results in eigenvectors that closely match the Legendre functions.

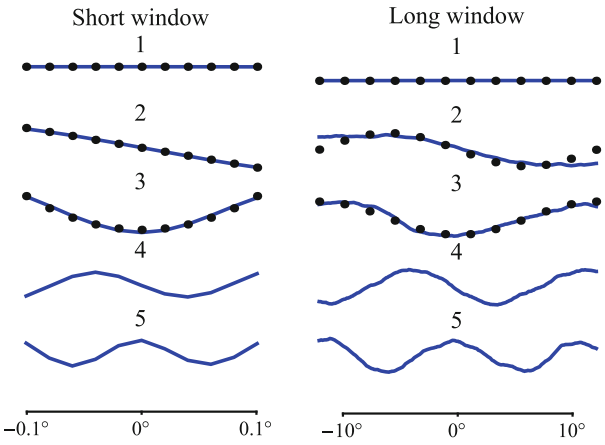


Fig. 3.11 Effect of length of the sampling window on principal components analysis of visual images. **Left Figure** With a short window length of 0.2 ° of visual angle, the eigenvectors are filters that compute successive derivatives of the samples. With a uniform covariance these converge to the discrete Legendre functions shown by dotted lines for the first three eigenvectors. **Right Figure** With a long window length of 20 ° of visual angle, the eigenvectors converge to approximations of successive components of the Fourier series. The first three sinusoidal components are again shown by dotted lines

At the other extreme, if the window is infinitely long the eigenvectors will correspond to the sinusoidal functions of the Fourier series [27]. The change in the form of the eigenvectors with increasing window length is also illustrated in Fig. 3.11 by plotting the eigenvectors with a long window alongside those obtained with a short window. Although the window is not infinitely long, there is a noticeable change in the form of the eigenvectors towards the sinusoidal components of the Fourier series. It is possible that the shift between the place cell responses in the hippocampus and the grid cell responses in the entorhinal cortex correspond to a just such change from a local representation to a Fourier representation [20] associated with a change in the size of the area processed by the neurons.

Associated with the change in eigenfunctions is a change in the distribution of eigenvalues, as illustrated in Fig. 3.12. With small windows, the eigenvalues of successive eigenvectors are initially large but decline rapidly in size, whereas with large windows the eigenvalues start much smaller but decline much more slowly with successive eigenvectors. So in practice it is usual to use a small window so that the first few eigenvectors account for all the covariance of the data.

In many cases, the largest eigenvectors account for most of the covariance of the data, and this fact can be exploited to transform data from lying in a high-dimensional space to a low-dimensional space, a procedure referred to as **dimensionality reduction**. For example, take the length changes of the six eye muscles as the gaze is transferred from one target to another. There are six extraocular muscles involved in rotation of the eye, which are grouped in three pairs:

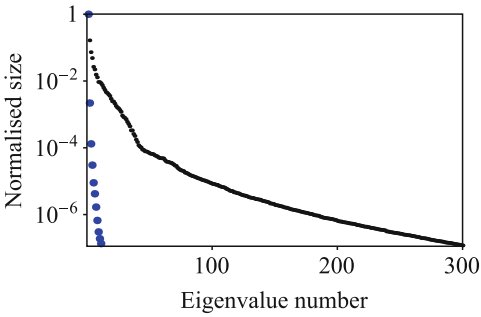


Fig. 3.12 Comparison of the distribution of eigenvalues obtained by principal component analysis of natural images with a short window (plotted in blue) or a long window (plotted in black)

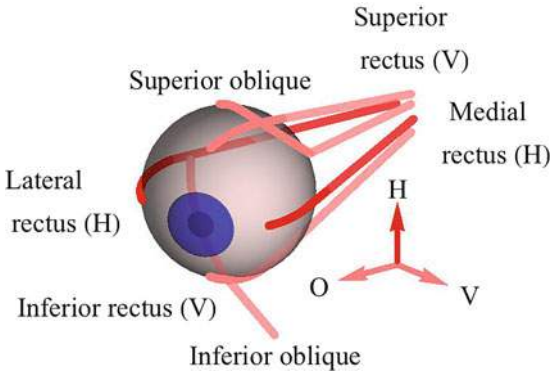


Fig. 3.13 Illustration of the paired actions of the extraocular muscles of the right eye. The horizontal recti, indicated by the letter H, rotate the eye about a vertical axis. The vertical recti, indicated by the letter V, rotate the eye about an axis in the horizontal plane. The remaining oblique muscles also rotate the eye about an axis in the horizontal plane

the lateral and medial rectus, the superior and inferior rectus and the superior and inferior oblique muscles. The recti muscles have their origin at the back of the orbit, as does the superior oblique, whilst the inferior oblique arises at the front of the orbit on the nasal side. The lateral and medial recti, referred to as the horizontal recti, lie on either side of the eye, and the superior and inferior recti, referred to as the vertical recti, lie above and below the eye respectively. The superior oblique passes through a loop of cartilage at the front of the orbit and back over the upper portion of the eye, attaching to the eye underneath the superior rectus. The inferior rectus passes under the eye, outside the inferior rectus and attaches to the eye underneath the lateral rectus. The layout of the muscles is illustrated schematically in Fig. 3.13.

The length changes required to hold the eye in a particular position can be computed from a model of the mechanics of the extraocular muscles that has successfully predicted the outcomes of squint surgery [18, 19, 24]. Calculation of the length changes associated with a selection of target directions within 25 ° of the straight ahead direction generates a set of six-dimensional data and the eigenvalues generated by principal components analysis are plotted in Fig. 3.14. The plot shows that almost all the variance of the data can be generated by linear combinations of the first two eigenvectors so in effect, the data lies in a two-dimensional space spanned by the first two eigenvectors.

As a final point, the data is often organised into an n × m matrix. In the case of the cone responses m equals three and n equals the number of Munsell chips.

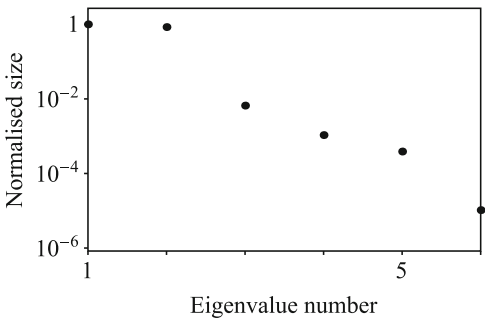


Fig. 3.14 Relative sizes of the six eigenvalues obtained by principal components analysis of length changes of the six eye muscles over a range of gaze directions. The figure illustrates that almost all the variance is described by linear combinations of the first two eigenvectors

It can be computationally efficient to work directly with these rectangular matrices without having to form the square covariance matrix. For the n × m matrix A it is possible to find associated matrices X, Y and Z such that $\mathsf{AX} = \mathsf{ZY}$ and $\mathsf{A}^\top\mathsf{Z} = \mathsf{XY}^\top$ where Y is a matrix of zeroes except for the first n or m (whichever is smaller) diagonal elements. The columns of X are referred to as the right singular vectors of A and the columns of Z as the left singular vectors of A. The right singular vectors are identical to the eigenvectors of $\mathsf{A}^\top\mathsf{A}$ and the left singular vectors are identical to the eigenvectors of $\mathsf{AA}^\top$ since:

$$\mathsf{A}^\top\mathsf{AX} = \mathsf{A}^\top\mathsf{ZY} = \mathsf{XY}^\top\mathsf{Y} \tag{3.32}$$

and

$$\mathsf{AA}^\top\mathsf{Z} = \mathsf{AXY}^\top = \mathsf{ZYY}^\top \tag{3.33}$$

The values of the non-zero elements of Y are referred to as the **singular values** of A and it also follows that the singular values correspond to the square roots of the eigenvalues of $\mathsf{A}^\top\mathsf{A}$. Since the eigenvectors of a symmetric matrix are orthonormal, the original matrix A is equal to the product $\mathsf{ZYX}^\top$ of the three matrices X, Y and Z. This representation of A is referred to as its **singular value decomposition**. If A is an m row by n column matrix, X is an n row by n column matrix, Y is an n row by n column matrix of zeroes, except for the first n or m diagonal elements and Z is an m row by m column matrix [12, 22].

3.4 Anti-Hebbian Learning

The complement to Hebbian learning is **anti-Hebbian** learning in which co-firing of inputs results in a reduction of the output of a neuron. The linear filter model neuron illustrated in Fig. 1.6 can be made to show anti-Hebbian learning by applying

the learning rule:

$$\Delta w_i = -\eta x_i y \tag{3.34}$$

where η is a constant that specifies the learning rate. By comparison with the Hebbian learning rule given by Eq. 3.29 it can be seen that all that is involved in obtaining anti-Hebbian learning is a sign change in the update rule.

Whereas the result of Hebbian learning is that the neuron responds most to the component of the input which is varying the most, anti-Hebbian learning results in the neuron responding least to this component. So if the output of the neuron y is replaced by another signal s, the neuron can effectively be trained not to respond to the main component of variation of s. A version of anti-Hebbian learning that ensures the neuron does not respond to the main component of variation of the training signal is based on the covariance update rule:

$$\Delta w_i = -\eta \,\text{covariance}(s, x_i) \tag{3.35}$$

This type of learning turns out to be very useful for eliminating sensory signals that are not due to a change in the environment but are a by-product of an animal's own actions.

For example, electric fish have adapted muscle cells to produce electric pulses and have receptors that respond to the electric currents created by the pulses. The currents depend on the surroundings of the fishes and can be used to guide their interactions with their environment. But the changes in the currents produced by objects close to the fish are very small compared with those produced by the pulses themselves and have to be separated out before they can be used [2]. This task is performed by the cerebellum-like structure of the fish illustrated in Fig. 3.15 which consists of granule cells and associated parallel fibres that form synapses with ganglion cells. The granule neurons receive sensory information about the activity of muscles in the body of the fish and the generation of the electric pulses and the ganglion cells receive sensory information from the electric current receptors. Anti-Hebbian learning could be used by the cerebellum-like structure to create a signal that predicts the sensory input due to self-motion of the animal, enabling its removal.

Similar processing is required for motor behaviour. Neural signals always show a degree of variation and drift and so the signals have to be constantly monitored and updated and the cerebellum is highly involved in this task. Damage to the cerebellum results in inaccurate movements and a failure to adapt to changes in muscle strengths [16]. This role is particularly significant for more demanding movement control tasks, such as dealing with interaction torques associated with shape changes, since these vary with the movement speed. Patients with cerebellar ataxia make more variable movements because they do not change their neural control signals to match the changes in the interaction torques [1].

The main input to the dentate and fastigial nuclei of the cerebellum comes from the Purkinje cells, which are inhibitory. The behaviour of the Purkinje cells is controlled by a characteristic circuit, a simplified diagram of which is shown

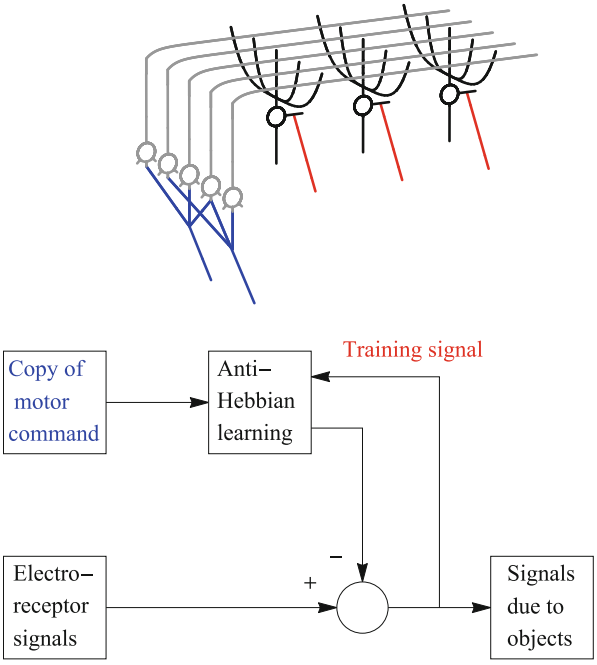


Fig. 3.15 Possible role of anti-Hebbian learning in sensory signal processing in the electric fish. **Upper Figure** Schematic diagram of the circuit found in the electrosensory lateral line lobe of the electric fish. The granule cells (drawn in grey) form a separate cluster. They receive input signals associated with movement of the animal's body and the generation of electrical pulses by the animal. The ganglion cells (drawn in black) receive inputs from both the parallel fibres and from electroreceptors in the animal's skin (drawn in red). **Lower Figure** Block diagram of the proposed signal flow in the cerebellar-like structure. The weights of the inhibitory ganglion cells are adjusted until there is no covariance between the motor command and the training signal, whereupon their output is equal to the component of the electrical signal produced by the activity of the animal's own muscles [6]

in Fig. 3.16. In the circuit signals are processed by granule cells and then passed to the Purkinje cells. The axons of the granule cells are unmyelinated and spread out to form parallel fibres that run transverse to the dendritic trees of the Purkinje cells. Long term depression of synapses of the parallel fibres with the Purkinje cell dendrites can be induced by climbing fibre inputs which originate in the inferior olive. Each Purkinje cell has only one climbing fibre connected to it, but the fibre branches and wraps around the dendrites of the cell. Parallel fibre excitation results in a single spikes from the Purkinje cells, which are referred to as simple spikes to distinguish them from the multiple spikes resulting from climbing fibre excitation that are referred to as complex spikes.

A possible scheme for how the cerebellum could act to compensate for the variations in the strengths of eye movement commands is illustrated in Fig. 3.16 [7]. In this case the task is to maintain calibration of the vestibulo-ocular reflex

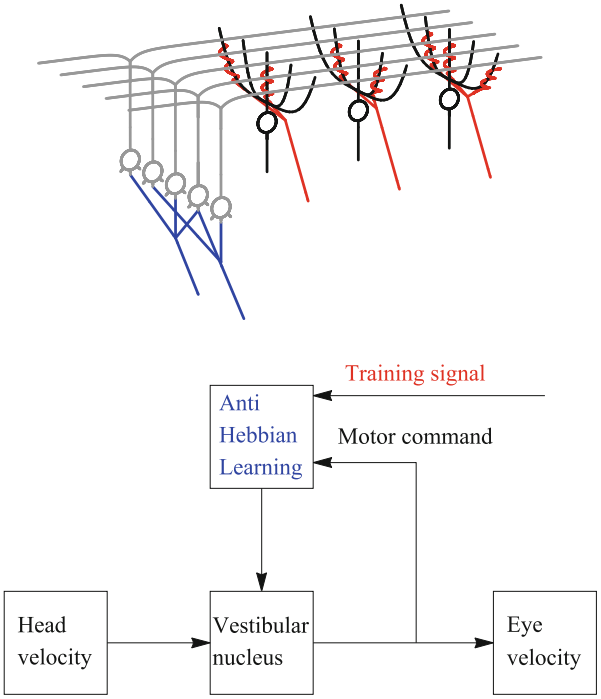


Fig. 3.16 Possible role for anti-Hebbian learning in processing of motor signals. **Upper Figure** Schematic diagram of the circuit repeated in the cerebellar cortex. Purkinje cells (drawn in black) receive inputs from climbing fibres (drawn in red) and from a dense layer of granule cells (drawn in grey). The granule cells have only a few short dendrites that receive excitatory inputs from mossy fibres (drawn in blue). Although this diagram captures part of the circuitry it does not include all the features. Not included are the large-bodied Golgi cells, the stellate and the basket cells. All of these receive inputs from the parallel fibres and have inhibitory outputs. The Golgi cells connect with the granule cells, and the stellate and basket cells form synapses with the Purkinje cells. **Lower Figure** Block diagram of the proposed signal flow in the vestibulo-ocular reflex. The mossy fibres carry motor commands, the climbing fibres deliver the amount of retinal slip and anti-Hebbian learning ensures that output from the Purkinje fibres alters until the motor command generates an eye movement exactly matching the head velocity [7]

which acts to keep the gaze direction aligned with the target and so ensure a stable retinal image despite head movements. If the head moves in one direction, the reflex produces a compensatory eye movement which is in the opposite direction to the head movement but of the same size.

The reflex depends on a neural pathway from the semicircular canals, which signal head velocity, on to the vestibular nucleus and ending on the oculomotor neurons, that drive the eye muscles. If the neural pathway is not perfectly reliable it will require calibration, and the region of the cerebellum involved with the reflex is the flocculus. Within this part of the cerebellum some of the Purkinje neurons send outputs directly to the vestibular nucleus. Assuming that the mossy fibres deliver a

copy of the motor command and that the training signal provided by the climbing fibres consists of the slippage of the retinal image, then after training, the output of the Purkinje neurons will carry the signal that must be added to the head velocity to ensure no slippage.

References

1. Bastian, A.J., Martin, T.A., Keating, J.G., Thach, W.T.: Cerebellar ataxia: abnormal control of interaction torques across multiple joints. J. Neurophysiol. **76**, 492–509 (1996). https://doi.org/10.1152/jn.1996.76.1.492
2. Bell, C.C., Han, V., Sawmill, N.B.: Cerebellar-like structures and their implications for cerebellar function. Annu. Rev. Neurosci. **31**, 1–24 (2008). https://doi.org/10.1146/annurev.neuro.30.051606.094225
3. Bowmaker, J.D., Dartnall, H.J., Lythgoe, J.N., Mollon, J.D.: The visual pigments and cones in the rhesus monkey, Macaca mulatta. J. Physiol. **274**, 329–348 (1978). https://doi.org/10.1113/jphysiol.1978.sp012151
4. Buchsbaum, G., Gottschalk, A.: Trichromacy, opponent colour coding and optimum colour information in the retina. Proc. R. Soc. Lond. B **220**, 89–113 (1983). https://doi.org/10.1098/rspb.1983.0090
5. Cicerone, C.M.: Constraints placed on color vision models by the relative number of different cone classes in human fovea centralis. Farbe **34**, 59–66 (1987)
6. Dean, P., Porrill, J.: The cerebellum as an adaptive filter: a general model? Funct. Neurol. **25**, 1–4 (2010)
7. Dean, P., Porrill, J. Stone, J.V.: Decorrelation control by the cerebellum achieves oculomotor plant compensation in simulated vestibulo-ocular reflex. Proc. R. Soc. B **269**, 1895–1904 (2002). https://doi.org/10.1098/rspb.2002.2103
8. Gibson, J., Farmer, J., Casdagli, M., Eubank, S.: An analytic approach to practical state space reconstruction. Physica D. **57**, 1–30 (1992). https://doi.org/10.1016/0167-2789(92)90085-2
9. Hateren, J., der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. B **265**, 359–366 (1998). https://doi.org/10.1098/rspb.1998.0303
10. Hebb, D.: The Organisation of Behavior: A Neuropsychological Theory. Wiley, New York (1949)
11. Hurvich, L., Jameson, D.: An opponent process theory of color vision. Psychol. Rev. **64**, 384–404 (1957). https://doi.org/10.1037/h0041403
12. Kirby, M.: Geometric Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns. Wiley, New York (2001)
13. Löwel, S., Singer, W.: Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. Science **255**, 209–212 (1992). https://doi.org/10.1126/science.1372754
14. Neitz, J., Carroll, J., Yamaauchi, Y., Neitz, M., Williams, D.R.: Color perception is mediated by a plastic neural mechanism that is adjustable in adults. Neuron **35**, 783–792 (2002). https://doi.org/10.1016/S0896-6273(02)00818-8
15. Oja, E.: A simplified neuron model as a principal component analyser. J. Math. Biol.**15**, 267–273 (1982). https://doi.org/10.1007/BF00275687
16. Optican, L., Miles, F.: Visually induced changes in primate saccadic oculomotor control signals. J. Neurophysiol. **54**, 940–958 (1985). https://doi.org/10.1152/jn.1985.54.4.940
17. Parkkinen, J.P.S., Hallikainen, J., Jaaskelainen, T.: Characteristic spectra of Munsell colors. J. Opt. Soc. Am. A **6**, 318–322 (1989). https://doi.org/10.1364/JOSAA.6.000318

18. Porrill, J., Warren, P.A., Dean, P.: A simple control model generates Listing's positions in a detailed model of the extraocular muscle system. Vis. Res. **40**, 3743–3758 (2000). https://doi.org/10.1016/S0042-6989(00)00211-X
19. Robinson, D.A.: A quantitative analysis of extraocular muscle cooperation and squint. Invest. Ophthalmol. Vis. Sci. **14**, 801–825 (1975)
20. Rodríguez-Dominguez, U., Caplan, J.B.: A hexagonal Fourier model of grid cells. Hippocampus **29**, 37–41 (2019). https://doi.org/10.1002/hipo.23028
21. Sanger, T.: Optimal unsupervised learning in a single-layer feed-forward neural network. Neural Netw. **2**, 459–473 (1989). https://doi.org/10.1016/0893-6080(89)90044-0
22. Schiff, S.J.: Neural Control Engineering. The MIT Press, Cambridge (2012)
23. Schmidt, B.P., Touch, P., Neitz, M., Neitz, J.: Circuitry to explain how the relative number of l and m cones shapes color ewqaxperience. J. Vis. **16**, 1–17 (2016). https://doi.org/10.1167/16.8.18
24. Simonsz, H.J., Spekreijse, H.: Robinson's computerized strabismus model comes of age. Strabismus **4**, 25–40 (1996). https://doi.org/10.3109/09273979609087734
25. Smith, P., Pokorny, J.: Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. Vis. Res. **15**, 161–171 (1975). https://doi.org/10.1016/0042-6989(75)90203-5
26. Stockman, A., Sharpe, L.T.: The spectral sensitivities of the middle- and long-wavelengthy-sensitive cones derived from measurements in observers of known genotype. Vis. Res. **40**, 1711–1737 (2000). https://doi.org/10.1016/S0042-6989(00)00021-3
27. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Phys. D Nonlinear Phenom. **35**, 395–424 (1989). https://doi.org/10.1016/0167-2789(89)90077-8
28. Vos, J.J., Walraven, P.L.: On the derivation of the foveal receptor primaries. Vis. Res. **11**, 799–818 (1971). https://doi.org/10.1016/0042-6989(71)90003-4
29. Wright, W.: The Measurement of Color. Adam Hilger, Cambridge (1969)
30. Young, R.A.: Principal-component analysis of macaque lateral geniculate nucleus data. Vis. Res. J. Opt. Soc. Am. A **3**, 1735–1742 (1986). https://doi.org/10.1364/JOSAA.3.001735

Chapter 4
From Local to Global

Individual sensory neurons can only signal measurements from a restricted region of space. To combine the individual descriptions of an object the brain has to be able to turn a local description into a global description. The task of making use of a retinal image is used as an example of how measurements from the outline of a manifold can be used to recover a global description by applying Morse theory.

4.1 Critical Points

Given a smooth manifold embedded in n-dimensional Euclidean space one can go on to use the value of any one of the coordinates of a point to indicate the height of the point on the manifold. Any function which can be used to assign a height to every point on a manifold is referred to as a **height function**. By investigating how the arrangement of points varies with the height one can gain insight into the shape of the manifold. For example, given a set of global coordinates (x_1, x_2, x_3) for a point on a surface embedded in everyday three-dimensional space, one can assume that the x_3 coordinate gives the height of the point. In particular, the points on the manifold where the tangent plane is orthogonal to the height function mark changes in the way the shape depends on the height as illustrated in Fig. 4.1 and are referred to as **critical points**.

Also illustrated by the example in Fig. 4.1, is that the tangent planes of the surface can be orthogonal to the direction of the height function at several different heights. In order to be able to differentiate the height functions with respect to the points on the surface it must be turned into a single-valued function. This can be done by introducing local coordinates. A local description of an m-dimensional manifold is provided by a function ϕ from each point in a neighbourhood of the manifold to an m-dimensional Euclidean space. In this way every point in the neighbourhood is assigned a unique set of coordinates $(\mathrm{x}_1, \mathrm{x}_2, \ldots \mathrm{x}_\mathrm{m})$, referred to as **local coordinates**.

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_4

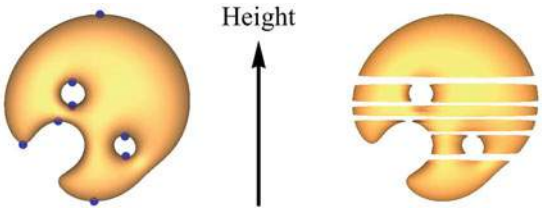


Fig. 4.1 An example of how a height function can by used to isolate the component parts of a three-dimensional surface. **Left Figure** A surface with the critical points associated with the height function plotted in blue. The critical points mark changes in the shape of the surface. **Right Figure** Illustration of how the surface is comprised of component shapes separated by the critical points. From the top these are: a bag, two tubes, one tube, a bag and a tube, two tubes and a bag

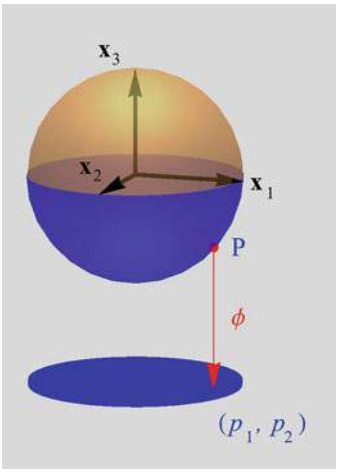


Fig. 4.2 Illustration of a system of local coordinates for the sphere. In effect, given a sphere specified by the equation $x_1^2 + x_2^2 + x_3^2 = r^2$, the hemisphere drawn in blue is being treated as the graph of the function $f(x_1, x_2) = \sqrt{r^2 - x_1^2 - x_2^2}$ where the negative sign outside the square root symbol is used to indicate that the negative square root is being taken. The local coordinates (p_1, p_2) of a point P on the graph are given by the projection of the point onto the plane spanned by the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$

An example of how local coordinates are set up for a portion of a sphere is shown in Fig. 4.2.

Although the theory of height functions extends to manifolds embedded in n-dimensional Euclidean space, only its application to surfaces embedded in 3-dimensional Euclidean space will be considered here, because this aspect of the theory is pertinent to the perception of solid objects. Let $f(x_1, x_2)$ be a smooth function of a system of local coordinates (x_1, x_2) which describes a patch of a surface. The local coordinates can be set up with the origin at the critical point so that $f(0, 0) = 0$. Then by the fundamental theorem of calculus the value of the

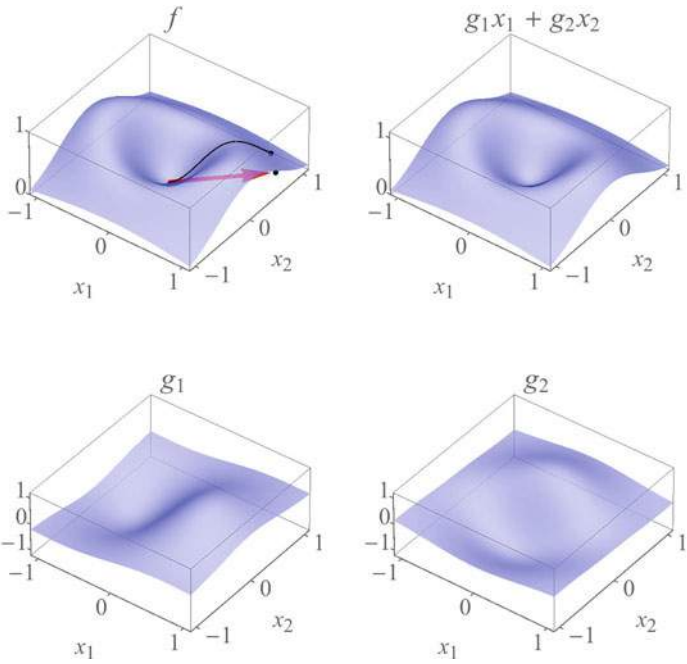


Fig. 4.3 Local description of the behaviour of a function $f(x_1, x_2)$ in the neighbourhood of a critical point at the origin (0,0). A plot of the function is shown in the top left hand figure. The straight line to an example pair of local coordinates (x_1, x_2), is indicated by a red arrow. The graph of the corresponding function $f(tx_1, tx_2)$ of a variable t, where t varies from 0 to 1, is shown in black. This function is used to derive the component functions g_1 and g_2 as described in the text

function at any point with local coordinates (x_1, x_2) is equal to the integral of the derivative of the function along the straight line from the origin to the point (x_1, x_2). If the distance along the straight line is specified by a variable t, where t is zero at the origin and one at the point (x_1, x_2) then this description of the function can be written as:

$$f(x_1, x_2) = \int_0^1 \frac{\mathrm{d}f(tx_1, tx_2)}{\mathrm{d}t}\mathrm{d}t = \int_0^1 \frac{\partial f}{\partial x_1}\Big|_{(tx_1, tx_2)} x_1 \mathrm{d}t + \int_0^1 \frac{\partial f}{\partial x_2}\Big|_{(tx_1, tx_2)} x_2 \mathrm{d}t \tag{4.1}$$

By putting $g_i(x_1, x_2) = \int_0^1 \frac{\partial f}{\partial x_i}\Big|_{(tx_1, tx_2)} \mathrm{d}t$ this equivalence can be rewritten as:

$$f(x_1, x_2) = g_1(x_1, x_2)x_1 + g_2(x_1, x_2)x_2 \tag{4.2}$$

as illustrated in Fig. 4.3.

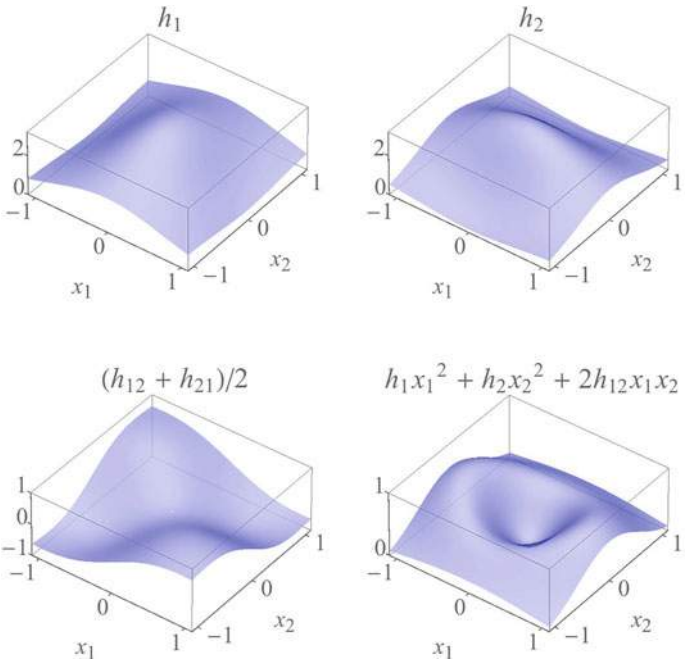


Fig. 4.4 Local description of the behaviour of a function in the neighbourhood of a critical point by a quadratic function

Since the origin of the system of local coordinates is a critical point all the g_j functions are zero at the origin and one can repeat this argument with each of them:

$$g_j(x_1, x_2) = \int_0^1 \frac{\mathrm{d}g_j(tx_1, tx_2)}{\mathrm{d}t}\mathrm{d}t = \int_0^1 \frac{\partial g_j}{\partial x_1}\Big|_{(tx_1,tx_2)} x_1 \mathrm{d}t + \int_0^1 \frac{\partial g_j}{\partial x_2}\Big|_{(tx_1,tx_2)} x_2 \mathrm{d}t \tag{4.3}$$

Proceeding as before, putting $h_{ij}(x_1, x_2) = \int_0^1 \frac{\partial g_j}{\partial x_i}\Big|_{(tx_1,tx_2)} \mathrm{d}t$ the expression for f in terms of two functions g_1 and g_2 becomes an expression in terms of four functions $h_{11}, h_{12}, h_{2,1}$ and $h_{2,2}$:

$$f(x_1, x_2) = h_{11}(x_1, x_2)x_1^2 + h_{12}(x_1, x_2)x_1x_2 + h_{21}(x_1, x_2)x_2x_1 + h_{22}(x_1, x_2)x_2^2 \tag{4.4}$$

as illustrated in Fig. 4.4

By putting:

$$a_{1,1} = h_{1,1},\ a_{2,2} = h_{2,2} \text{ and } a_{1,2} = a_{2,1} = \frac{h_{1,2} + h_{2,1}}{2} \tag{4.5}$$

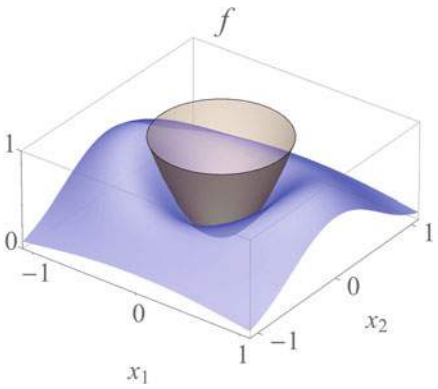


Fig. 4.5 Illustration of the approximation of a function in a small neighbourood of a critical point by a quadric surface

the local definition of the function f can be written as a quadratic form:

$$f(x_1, x_2) = \mathsf{X}^\top \mathsf{A}\mathsf{X} \tag{4.6}$$

where X is the 2 row by 1 column matrix of local coordinates. The symmetric matrix A can be diagonalised so in the neighbourhood of the critical point the shape of the surface patch can be described by a linear combination of the signed square of each of the local coordinates [12], as illustrated in Fig. 4.5. This finding is referred to as the **Morse lemma**.

4.2 Receptive and Association Fields

To interact effectively with other animals the brain has to be able to identify the positions and velocities of the limbs of the animals. But most of the sense organs do not give direct access to quantitative values for the positions of physical surfaces, only to measurements of some property of the surfaces. In the case of the retinal image, measurements of the intensity distribution of light reflected from the surfaces in view can be described by a function of position on the retina and the graph of this function defines a new surface embedded in three-dimensional Euclidean space, as illustrated in Fig. 4.6.

Analysis of the retinal image in vertebrates begins with rods and cones. But whilst an individual rod or cone receptor signals the intensity at a particular position on the surface, this signal does not reach the brain as it has to be turned into nerve impulses prior to transmission along the length of optic nerve. This conversion is done by the retinal ganglion cells which combine the responses of many receptors. A distinctive feature of the ganglion cells in the retina is that each neuron receives inputs from a limited area of the retina, referred to as its **receptive field** [7].

The strengths of the connections associated with each position in the receptive field can be found experimentally by flashing small spots of light at each position

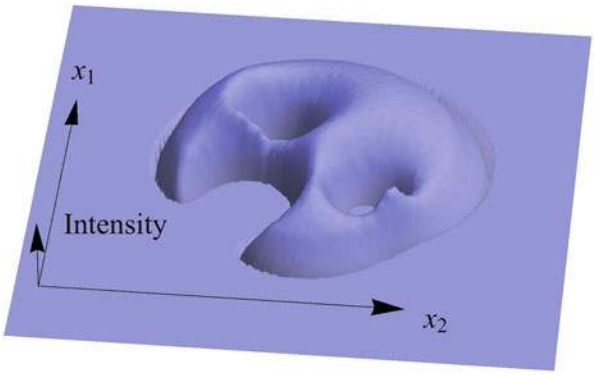


Fig. 4.6 Graph of the image intensity of the left half of Fig. 4.1 as a function of retinal position

and measuring the response of the neuron. In the retina the receptive fields of the cells have a concentric organisation consisting of a centre and a surrounding ring. The centre responds to a stimulus being turned either on or off, while the surround responds to a light action that is opposite to that of the centre. The output of the neuron therefore depends on the difference between the stimulation of the centre and surround, for if a large light spot covers both the centre and surround then the response is less than with stimulation of either the centre or surround alone [9], so the cell can be thought of as signalling the local contrast of the image.

For cells which respond linearly, both the centre and surround sensitivity profiles of retinal receptive fields can be described by Gaussian functions of position [13] and the overall receptive field profile can be described by the difference between the surround Gaussian and the centre Gaussian. Such Difference-of-Gaussian (abbreviated to DoG) functions are described by the equation :

$$DoG(x) = \frac{\mathrm{a}}{\sigma\sqrt{2\pi}} \exp - \left(\frac{x}{\sqrt{2}\sigma}\right)^2 - \frac{\mathrm{b}}{\mathrm{s}\sigma\sqrt{2\pi}} \exp - \left(\frac{x}{\sqrt{2}\mathrm{s}\sigma}\right)^2 \tag{4.7}$$

where σ and sσ specify the square root of the variances of the Gaussian functions of the centre and surround respectively, and s, a and b are constants. If a = b = 1, the Gaussian functions used are normal distributions, so each has an area under the curve equal to one, and if the entire receptive field is stimulated uniformly, the responses from the centre and surround cancel out and a zero response is produced. An example of a Difference-of-Gaussian weighting function is plotted in Fig. 4.7.

The property of the receptive fields of cortical cells that distinguishes them from the receptive fields of cells in the retina and lateral geniculate nucleus is that they are elongated in one direction and so respond best to slits of light, dark bars and edges [8]. A distinction can be made between simple cells, whose receptive fields can be explored with stationary spots of flashing light, and complex cells which respond poorly to stationary spots of flashing light, but do respond to moving bars

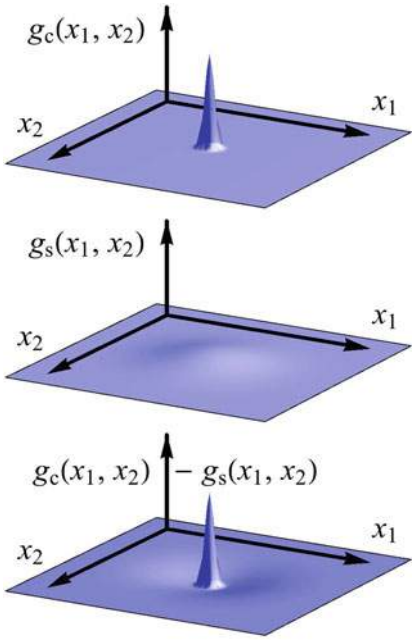


Fig. 4.7 Description of a rotationally symmetric receptive field by the Difference-of-Gaussian function defined by Eq. 4.7. If the variables x_1 and x_2 are taken to be the visual angle in degrees, then region illustrated corresponds to a 1° by 1° retinal patch. **Upper Figure** The excitatory centre is $g_c(x_1, x_2)$ is described by a normal distribution with $\sigma = 0.023$. **Middle Figure** The broader inhibitory surround $g_s(x_1, x_2)$ is described by a normal distribution with $s\sigma = 0.194$. **Lower Figure** Graph of the Difference-of-Gaussian function resulting from subtracting the inhibitory surround from the excitatory centre

and edges. The receptive fields of simple cells can be divided into mutually exclusive regions, each of which responds to light on or light off. Complex cells do not in general have subdivisions of their receptive fields although they too are tuned to stimuli with specific widths and orientations.

The Difference-of-Gaussian model of receptive fields can be extended to simple cells by using a one-dimensional Difference-of-Gaussian function to describe the cross-section of the receptive field and a Gaussian function to describe the lengthways profile of the receptive field. Typically this is three times the width of the receptive field as illustrated in Fig. 4.8.

Another difference between cortical receptive fields and retinal receptive fields is in the greater range of profiles. In the retina, most linear receptive fields are approximately radially symmetric, but in the cortex the weighting $w(x)$ of the cross-section can vary from being symmetric, for which $w(x) = w(-x)$ to being antisymmetic for which $w(x) = -w(-x)$. This may reflect the ability of cortical neurons to calculate not just local functions but also derivatives of these functions. For example, the derivative of the symmetric Difference-of-Gaussian weighting function is antisymmetric as illustrated in Fig. 4.8.

Evidence for the existence of spatial frequency filters early on in the visual pathway comes not just from neurophysiology but also from psychophysics [1], and the latter technique has provided quantitative estimates of the filters of the human visual system. An example set of such filters, which were described by receptive fields formed by differences of either two or three Gaussian functions, is illustrated in Fig. 4.9 [16]. One characteristic of a filter tuned in the frequency domain is its

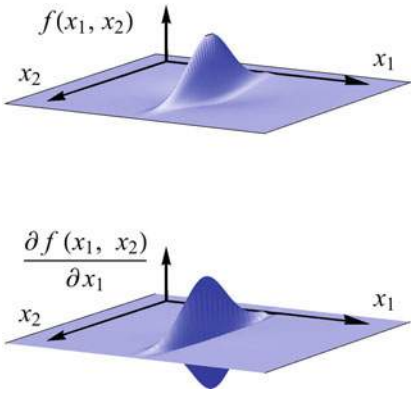


Fig. 4.8 Description of the receptive field weighting of a simple cell by the product of a cross-section function with a Gaussian function lengthways. **Upper Figure** A receptive field with a cross-section described by Difference-of-Gaussian function with same parameters as those used in Fig. 4.7. **Lower Figure** A receptive field with a cross-section corresponding to the rate of change of the Difference-of-Gaussian function with position

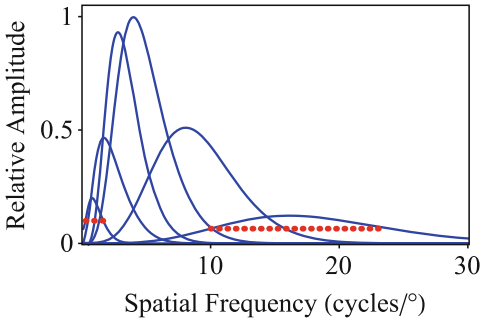


Fig. 4.9 Spatial frequency tuning of filters in the human visual system. The bandwidth of the filters change with frequency, as can be seen by comparing the bandwidths of the filter tuned to the lowest spatial frequency with that of the filter tuned to the highest spatial frequency. The bandwidths are indicated by the dotted red lines

bandwidth, which can be specified in terms of the difference between the upper and lower frequencies at which the filter produces half of its maximum response. The psychophysical findings have revealed that the bandwidths of the filters are not the same in every filter but change with the frequency to which the filter is tuned. So it is clear that a range of different filters is involved in vision and the particular profiles shown in Figs. 4.7 and 4.8 were chosen simply to reflect the average of the range.

Whilst a cortical cell does not respond to a stimulus presented outside its receptive field, such stimuli can markedly alter the response of the neuron to a stimulus lying within its receptive field. In particular, extending a contour with the preferred orientation of a neuron beyond its receptive field increases its response, whilst the presence of a contour with a different orientation decreases its response

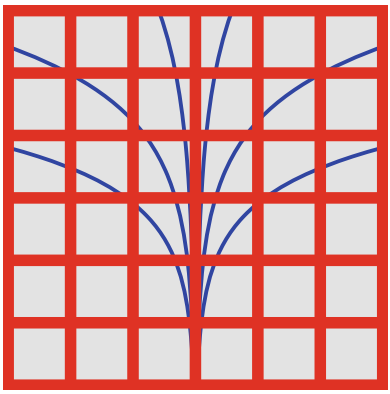


Fig. 4.10 Combination of separate parts of a smooth curve into a single extended curve. Although many parts of the blue lines are hidden they can clearly be seen to form a set of six curved lines

[11]. The enhancement of the responses of the neurons of the visual cortex helps to bring together the responses to different parts of a contour into a single whole, as illustrated in Fig. 4.10. By analogy with the idea of a receptive field, the cortical neuron can also be thought of as having an **association field** that extends well beyond its receptive field and results in enhanced responses to smooth contours [6].

Insight into how a contour is formed can be gained by assuming that it is generated by the path of a Braitenberg vehicle such as that portrayed in Fig. 2.12. The simplest case arises when the vehicle moves over a flat surface on which the wheels do not slip, so that the vehicle can move forwards or backwards, or turn, or make a combination of these movements, but what it cannot do is move sideways. In visual terms, the no-slip condition is just what is required to ensure that only smooth contours result in edge facilitation.

As illustrated in Fig. 4.11, the configuration space for the vehicle is a three-dimensional Euclidean space in which the first two axes can be conveniently positioned in the plane so the remaining axis is located perpendicular to the plane. Following through with this description the retina corresponds to a base manifold R2 and each hypercolumn of the visual cortex, which consists a set of orientation columns that span the entire range of orientations, corresponds to the principal fibre bundle R2 $\times$ S1.

The no-slip constraint requires that the possible velocities of the vehicle lie in a two-dimensional portion of the tangent space, spanned by a vector $\mathbf{x}_1$ which has a length and direction specified by the linear velocity of the vehicle, and a vertical vector $\mathbf{x}_2$ that describes the rate of change of orientation of the linear velocity vector. So, as with the manifold of the falling cat, the tangent space of the sub-Riemannian configuration space can be divided into horizontal and vertical directions. In this case the horizontal directions are contained in the tangent plane of allowed movements spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

To apply the vehicle analogy to contour processing one can assume that the vehicle travels at unit velocity so that its linear velocity is given by the tangent vector to the retinal outline $(\sin(\theta), \cos(\theta), 0)$ where θ is the orientation of the edge

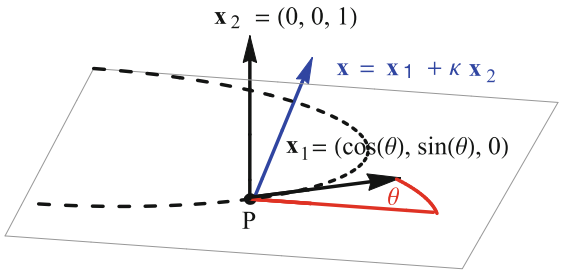


Fig. 4.11 Tangent space for movements of the Braitenberg vehicle portrayed in Fig. 2.12 along a retinal contour. The path of the vehicle is shown by a dashed line. At a given point P on the ground, the tangent to the path makes an angle θ with the horizontal meridian and this angle is used to characterise the retinal tangent vector $\mathbf{x}_1$. The change in the orientation of the vehicle is specified by a displacement in the vertical direction $\mathbf{x}_2$, and the combination of displacements is given by the vector sum $\mathbf{x} = \mathbf{x}_1 + \kappa\mathbf{x}_2$ where κ is the curvature of the path of the vehicle at the point P

on the retina. Locally, a smooth curve follows the path of a circle and the reciprocal of the radius of the circle is referred to as the curvature of the curve, and is denoted by κ. For a straight line the curvature is zero. By putting the size of the vertical shift equal to the curvature of the outline, the three-dimensional curve keeps track of the curvature of the retinal contour. So within the geometric framework, the processing carried out in the cortex effectively converts an outline on the retina into a curve in three-dimensional space.

The linear combination defining the tangent vector in three-dimensional Euclidean space is then specified by Eq. 4.8:

$$\mathbf{x} = (\sin(\theta), \cos(\theta), 0) + \kappa(0, 0, 1)) = \mathbf{x_1} + \kappa\mathbf{x_2} \tag{4.8}$$

As the vehicle moves along its path the vector $\mathbf{x}$ will trace a path in three-dimensional Euclidean space and the length of this path can be can be calculated by applying the Euclidean metric. Hence at any given position on a contour the length of the tangent vector $\mathbf{x} = \mathbf{x_1} + \kappa\mathbf{x_2}$ is given by:

$$\|\mathbf{x}\| = \|\mathbf{x_1} + \kappa\mathbf{x_2}\| = \sqrt{1 + \kappa^2} \tag{4.9}$$

The length γ of the three-dimensional path between two points on the retinal contour is then given by the integral:

$$\gamma = \int_0^1 \sqrt{1 + \kappa^2} dt \tag{4.10}$$

where t is a parameter specifying the position on the retinal contour that is equal to 0 at the start point and 1 at the end point [4].

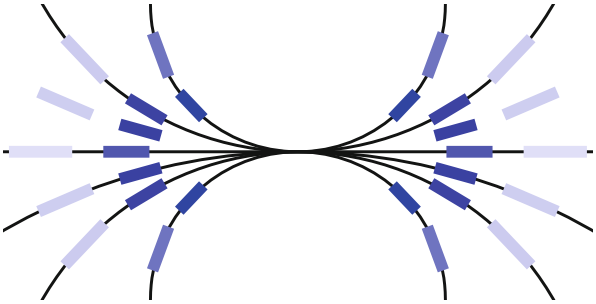


Fig. 4.12 Illustration of the role of an association field. The black lines indicate possible paths of smooth outlines passing through a horizontally-orientated receptive field. The blue line segments show the orientations of examples of receptive fields stimulated by such paths, which are grouped in a characteristic bow tie shape. The association field is formed by a weighted combination of the stimulated receptive fields and activation of the association field enhances the response of the horizontally-orientated receptive field. The weights are specified by a Gaussian function of the path lengths given by the sub-Riemannian metric defined by Eq. 4.10 and are indicated by the saturation of the blue colour of the example receptive fields

Given a metric for a contour one can model an association field by using a Gaussian function of the distance from the receptive field centre to specify the weight distribution. This is illustrated in Fig. 4.12, where the weight assigned to the output of neighbouring receptive fields is indicated by the saturation of the blue icons. The quantitative description of the orientated filters and the association fields can be combined to model the behaviour of the early stages of the visual cortex. Neurons with the same preferred orientation respond to changes along a height function perpendicular to the orientation to which they are tuned as illustrated in Fig. 4.13. It turns out that the neurons that respond to the first derivative of the local contrast function are ideally suited to signalling the location of the critical points.

4.3 Bump Functions

Research in neuroscience is distinctive in that one always has, at the back of one's mind, the thought that an experimental finding might have implications for understanding our subjective experiences. But our subjective worlds are a result of so much neural elaboration that it is difficult to gain insight into the early stages of processing of the sensory stimuli. Just because we find a neural signal for critical points is no guarantee that the nervous system cares about this one way or another. But certainly, a capability of the visual system must be to be able to take local measurements from an object and turn them into a unified description of it. One has to answer the question of what sort of description is being used.

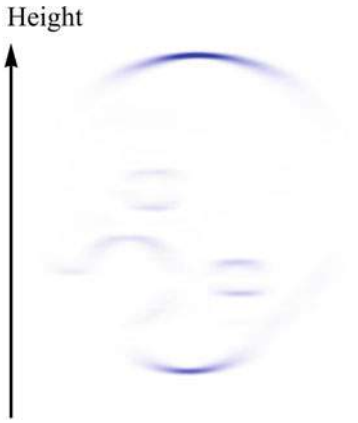


Fig. 4.13 An illustration of how cortical neurons that respond to the first derivative of contrast could find the critical points of surface. The direction of the height function is indicated by an arrow. The responses are maximal at the critical points as can be seen by comparison with Fig. 4.1

A prerequisite for successful interaction with an object is its shape because it cannot be effectively manipulated without this knowledge. The most fundamental properties of shape are topological and concern questions such as how many holes does the object have and discrimination experiments have revealed that animals are adept at distinguishing the presence of holes [2, 3]. The presence of holes can be detected by application of the Morse lemma. For a surface the signs of the squares can be both positive, both negative or they can be opposite. The shapes associated with such conditions, correspond to a local minimum, a local maximum or a saddle. A sphere will have a local maximum and a local minimum at the intersections with the height function and a hole will have a saddles at opposite intersections. So the local shapes are grouped in pairs and the number of local shapes will vary with the number of holes. To exploit this relationship one can begin by calculating the Euler characteristic $\chi(S)$ given by:

$$\chi(S) = \text{number of minima} + \text{number of maxima- number of saddles} \tag{4.11}$$

For the example shown in Fig. 4.1 there is one maximum, two minima and five saddles, giving an Euler characteristic of $1 + 2 - 5 = -2$. The Euler characteristic is also equal to two minus twice the number of holes g so g = (2 - $\chi(S)$)/2 which is this case is equal to 2, as is the case.

But to be able to handle an object one often needs the slant of the holes and protrusions so as to grasp it safely. In the neighbourhood of a point on the outline of a surface, it follows from the Morse lemma that the shape of the surface can be approximated by a quadratic function. Given a pair of critical points which are adjacent to each other with respect to the height function one can interpolate between the critical points to obtain the shape of the portion of the surface lying between them. It is even simpler to interpolate the slant of the tangent planes, which is a linear function of position along the height function, as illustrated in Fig. 4.14. But this presupposes that the brain can identify pairs of corresponding critical points and the separation between them. Furthermore, the interpolation procedure acts

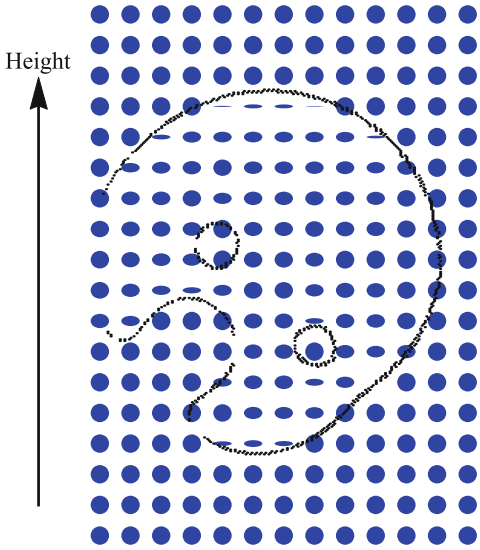


Fig. 4.14 Illustration of surface shape estimation by tangent plane interpolation. An estimation of surface slant and how it changes with height can be obtained from points on the outline of the figure. The estimates are indicated by the appearance of a disk which becomes more elongated with increasing slant. A single height function only gives an estimate of slant in the direction of the height function and to obtain an estimate of the slant of the surface in all directions, the results from height functions with different orientations would have to be combined

along the direction of the height function as illustrated in Fig. 4.14 and to obtain the overall shape of the object the results of applying the procedure along all the possible directions of the height function would have to be combined.

The analysis of critical points depends on having the manifold embedded in Euclidean space and a source of ideas about how the brain might combine information from critical points comes from the theory of the embedding process. It turns out that a smooth manifold can always be embedded in a high-dimensional Euclidean space [15]. The simplest proof of the embedding theorem applies only to manifolds that can be covered by a finite number N of neighbourhoods. The neighbourhoods can be large or small and can be located anywhere on the manifold, as illustrated in Fig. 4.15 for a two-dimensional manifold, so that by defining enough neighbourhoods one can cover the entire manifold. A neighbourhood can also be defined at a boundary by aligning the direction of one of the local coordinates with the boundary.

The task of amalgamating the collection of local coordinates into a single global system of coordinates requires a procedure to smoothly combine the local descriptions. A useful tool for this task is the **bump function** $b(x_1, x_2)$ which is a smooth function that is non-zero for all points in the neighbourhood of a given point

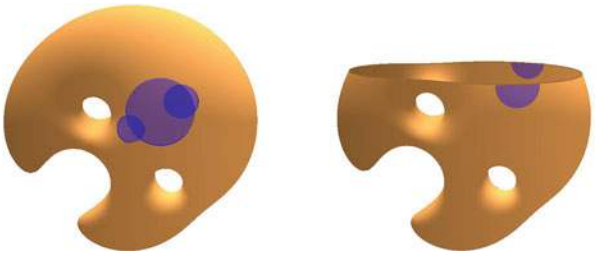


Fig. 4.15 Examples of different sizes and positions of possible neighbourhoods of a surface. Each of the blue regions indicates a neighbourhood of a point on the manifold. At a boundary of the manifold, the shape of the neighbourhood can be matched to that of the boundary

on the surface, and zero everywhere else. An example of a bump function is:

$$b(x_1, x_2) = \begin{cases} exp\left(\frac{1}{x_1^2+x_2^2-1}\right), & \text{when}(x_1^2 + x_2^2) < 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.12}$$

and this function is plotted in Fig. 4.16.

The embedding of a surface M can be constructed by combining the local coordinates into a set of smooth functions of position on the surface, each of which specifies a coordinate. For each neighbourhood region $\mathrm{R^i}$ on M set up a system of local coordinates ϕ^{i} that maps each point in $\mathrm{R^i}$ into a point with coordinates (x_1^i, x_2^i) in 2-dimensional Euclidean space. Next, form the bump functions $b^i(x_1^i, x_2^i)$, $x_1^i b^i(x_1^i, x_2^i)$ and $x_2^i b^i(x_1^i, x_2^i)$. Then every point on the surface has a set of coordinates in a 2s + s dimensional Euclidean space:

$$\begin{aligned} \mathrm{P} = (&b^1(x_1^1, x_2^1),\ x_1^1 b^1(x_1^1, x_2^1),\ x_2^1 b^1(x_1^1, x_2^1), \\ &b^2(x_1^2, x_2^2), \quad x_1^2 b^2(x_1^2, x_2^2),\ x_2^2 b^2(x_1^2, x_2^2), \\ &\cdots \\ &b^s(x_1^s, x_2^s), \quad x_1^s b^s(x_1^s, x_2^s),\ x_2^s b^s(x_1^s, x_2^s)) \end{aligned} \tag{4.13}$$

where s is the number of neighbourhood regions. Now if one takes two distinct points P and Q on the surface then if P is in region $\mathrm{R^i}$ and Q is not then $b^i(p_1^i, p_2^i) \neq 0$ and $b^i(q_1^i, q_2^i) = 0$ so the points have distinguishing coordinates. If instead, P and Q are both in $\mathrm{R^i}$ then if $b^i(p_1^i, p_2^i) \neq b^i(q_1^i, q_2^i)$ the coordinates of the points are different. If the bump functions of the two points are the same then their coordinates will still be different as either $p_1^i, b^i(p_1^i, p_2^i) \neq q_1^i, b^i(q_1^i, q_2^i)$ or $p_2^i, b^i(p_1^i, p_2^i) \neq q_2^i, b^i(q_1^i, q_2^i)$. The implications of this embedding theorem is that any smooth manifold can be embedded in Euclidean space (albeit a high-

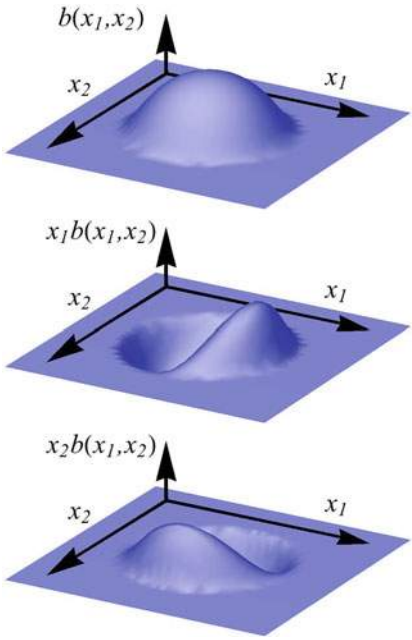


Fig. 4.16 Graph of an example bump function $b(x_1, x_2)$ that can be used to combine local descriptions of surfaces. Also shown are graphs of the functions $x_1 b(x_1, x_2)$ and $x_2 b(x_1, x_2)$ which can be used to ensure unique coordinates for each point on the surface

dimensional one) and so the properties of any smooth manifold can be analysed in terms of relations between sets of coordinates.

Coordinates are most often used in the specification of a function or vector field that varies smoothly over a manifold and bump functions are also useful for forming a global specification of functions and vector fields given in local coordinates. For example, if a point P on a surface lies within the neighbourhoods of k bump functions then one can define k representations of the value of a function g defined on the surface by $b^i(x_1^i, x_2^i)$ where $1 \leq i \leq k$. One can rescale these representations so that they sum to unity and then recover the value of g(P) simply by summing them:

$$\hat{b}^i(x_1^i, x_2^i)g(P) = \frac{b^1(x_1^1, x_2^1)g(P) + b^2(x_1^2, x_2^2)g(P) + \ldots + b^k(x_1^k, x_2^k)g(P)}{b^1(x_1^1, x_2^1) + b^2(x_1^2, x_2^2) + \ldots + b^k(x_1^k, x_2^k)} \tag{4.14}$$

Now this sum in Eq. 4.14 is the same as that given in Eq. 1.9 for reconstruction of the direction and length of a vector in a plane by a population of model neurons. The analogy is not entirely correct because the derivative of a bump function must tend to zero as the value of the bump function tends to zero to ensure smooth blending of the functions and this is not true for the positive half of the cosine function, but it does indicate how a neural system could combine the outputs of a population of neurons to reconstruct the values of functions and vector fields over entire surfaces.

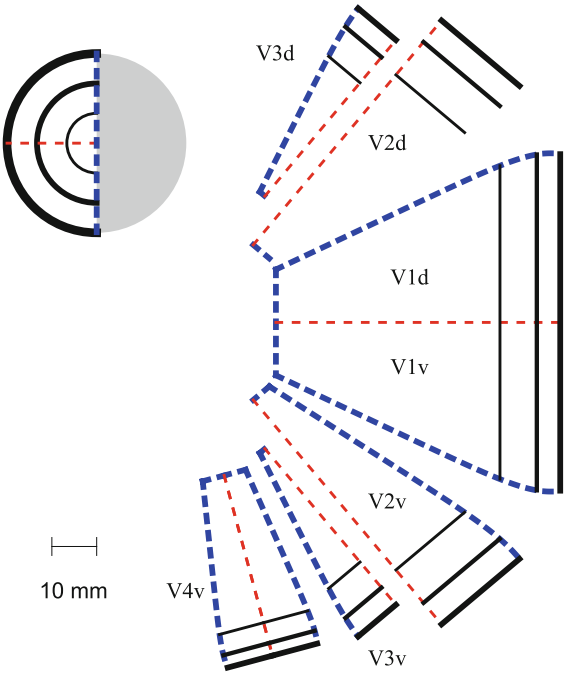


Fig. 4.17 Schematic diagram of the topographic arrangement of the larger areas of the right visual cortex which receives input from the left visual field as shown by the inset diagram with one half greyed out. The horizontal meridian is plotted in red and the vertical meridian is plotted in blue. Contours of 10, 20, and 30° of visual angle with respect to the foveal direction and plotted as black lines of increasing thickness. The cortical areas abut along either a horizontal or a vertical meridian, which minimises connection lengths between the representation of a visual direction in different cortical areas. The locations of the areas are distinguished by the letter d for dorsal and v for ventral

The idea that the visual cortex could use bump functions to interpolate measurements is no more than speculation, but it is in keeping with the presence of both simple and complex neurons in the visual cortex. Whilst simple cells, which have receptive fields that can be divided into mutually exclusive regions, have the properties required for the image filters, complex cells are suited to carrying out the task of the bump functions. Furthermore, there are many cortical areas beyond the primary visual cortex, as illustrated in Fig. 4.17, containing neurons with larger and larger receptive fields. Given that the receptive field sizes of neurons in the visual cortex approximately double in size at successive stages of the pathway it may be that the surface is reconstructed in stages by using successively more extensive bump functions, each of which responds to the sum of the smaller bump functions. By incorporating bump functions with a range of neighbourhood sizes the spatially separated orientations of the tangent planes at the critical points could be merged across the whole of the surface.

4.4 Gradient Space

An obvious criticism of emphasising the constraints available from smooth surfaces is that not all surfaces are smooth. But such surfaces also have constraints which can be exploited. In particular, if the surface is not smooth, but polyhedral, then the intersection of two flat faces will produce a line on the retina that provides a restriction on the possible orientations of the faces [10, 14]. The derivation of this constraint follows the viewing geometry illustrated in Fig. 4.18.

The depth of a point is specified by the x_3 coordinate and its retinal image is given by the intersection of the ray from the origin to the point with the $x_3 = -1$ plane and so the flat surface of a face of the polyhedron can be described by a function of the x_1 and x_2 variables:

$$-x_3 = f(x_1, x_2) \tag{4.15}$$

Given this function, the orientation of the surface is defined by its gradient:

$$(g_1, g_2) = \left(\frac{\partial f(x_1, x_2)}{\partial x_1}, \frac{\partial f(x_1, x_2)}{\partial x_2}\right) \tag{4.16}$$

and the collection of all possible coordinates (g_1, g_2) are referred to as the **gradient space**.

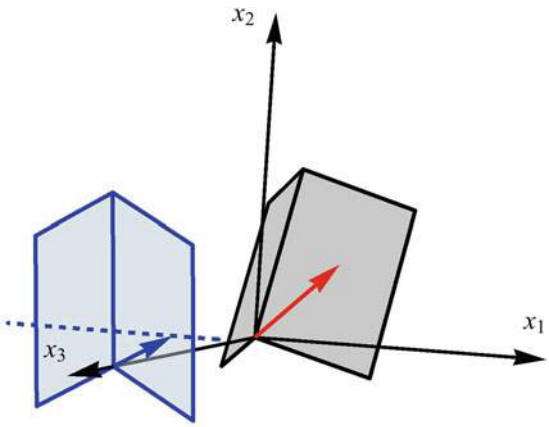


Fig. 4.18 Illustration of the gradient space constraint on the orientations of intersecting flat surfaces. Two surfaces that intersect in three-dimensional space are drawn in grey, and the corresponding retinal projections are drawn in blue. A vector orthogonal to one of the surfaces is plotted in red and the corresponding vector in gradient space is plotted in blue. The gradient space constraint is that the gradient space representations of the orientations of both surfaces must lie on the broken, blue line perpendicular to the retinal projection of the edge formed by the intersection of the faces

With this description of surface orientation $(1, 0, -g_1)$ and $(0, 1, -g_2)$ are tangent vectors in the x_1 and x_2 directions and $(g_1, g_2, 1)$ is a vector orthogonal to the surface. It follows that if (v_1, v_2, v_3) is a line lying in the surface:

$$(v_1, v_2, v_3) \cdot (g_1, g_2, 1) = 0 \text{ so } (v_1, v_2) \cdot (g_1, g_2) = -v_3 = \mathbf{v} \cdot \mathbf{g} \tag{4.17}$$

where $\mathbf{v}$ is a vector with coordinates (v_1, v_2) that specifies the direction of the line on the retina and $\mathbf{g}$ is a vector with coordinates (g_1, g_2) that specifies the gradient space representation of the orientation of the surface. When two faces intersect the rate of change of depth with distance along the retinal projection of the edge given by Eq. 4.17 has to be the same for both surfaces so:

$$\mathbf{v} \cdot \mathbf{g^1} = \mathbf{v} \cdot \mathbf{g^2} = -v_3 \text{ so } \mathbf{v} \cdot (\mathbf{g^1} - \mathbf{g^2}) = 0 \tag{4.18}$$

where $\mathbf{g^1}$ and $\mathbf{g^2}$ are the gradient space representations of the surfaces. It follows that the gradient space representations of the surfaces must lie on a line perpendicular to the retinal projection of the edge. So if the visual system does interpolate tangent planes between critical points, it can extend the tangent plane description of the surface to polyhedral surfaces when the information from critical points is no longer available.

The folded card illustrated in Fig. 4.18 looks flatter as the angle between the two halves is increased, so a card with one side folded further back will require that the gradient space representations of each half shift from the their positions with the symmetrical figures if the gradient space constraint is to be obeyed. The apparent orientation of each half of the card can be measured subjectively by setting a picture line so that it appears orthogonal to the surface of the card in depth, and such measurements show that the gradient space representations do indeed shift so that the edge formed by the intersection of the two halves has a slant approximately midway between the two symmetrical cases [5].

References

1. Blakemore, C., Campbell, F.W.: On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. J. Physiol. **203**, 237–260 (1969). https://doi.org/10.1113/jphysiol.1969.sp008862
2. Chen, L.: Topological structure in visual perception. Science **218**, 699–700 (1982). https://doi.org/10.1126/science.7134969
3. Chen, L., Zhang, S., Srinivasan, M.V.: Global perception in small brains: topological pattern recognition in honey bees. PNAS **100**, 6884–6889 (2003). https://doi.org/10.1073/pnas.0732090100
4. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. J. Math. Imaging Vis. **24**, 307–326 (2006). https://doi.org/10.1007/s10851-005-3630-2
5. Clement, R.A.: Apparent shape of the Mach book. Perception **25**, 313–319 (1996). https://doi.org/10.1068/p250313

6. Field, D.J., Hayes, A., Hess, R.F.: Contour integration by the human visual system: evidence for a local "association field". Vis. Res. **33**, 173–193 (1993). https://doi.org/10.1016/0042-6989(93)90156-Q
7. Hartline, H.: The response of single optic fibers of the vertebrate eye to illumination of the retina. Am. J. Physiol. **121**, 400–415 (1938). https://doi.org/10.1152/ajplegacy.1938.121.2.400
8. Hubel, D., Wiesel, T.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. **160**, 106–154 (1962). https://doi.org/10.1113/jphysiol.1962.sp006837
9. Kuffler, S.W.: Discharge patterns and functional organisation of the mammalian retina. J. Neurophysiol. **46**, 37–68 (1953). https://doi.org/10.1152/jn.1953.16.1.37
10. Mackworth, A.K.: Interpreting pictures of polyhedral scenes. Artif. Intell. **4**, 121–137 (1973). https://doi.org/10.1016/0004-3702(73)90003-9
11. Maffei, L., Fiorentini, A.: The unresponsive regions of visual cortical receptive fields. Vis. Res. **16**, 1131–1139, IN4-IN5 (1976). https://doi.org/10.1016/0042-6989(76)90253-4
12. Poston, T., Stewart, I.N.: Catastrophe Theory and its Applications. Pitman, London (1978)
13. Rodieck, R.: Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vis. Res. **5**, 583–601 (1965). https://doi.org/10.1016/0042-6989(65)90033-7
14. Shafer, S.A., Kanade, T., Kender, J.: Gradient space under orthography and perspective. Comput. Vis. Graph. Image Process. **24**, 182–199 (1983). https://doi.org/10.1016/0734-189X(83)90042-7
15. Wallace, A.: Differential Topology: First Steps. W.A.Benjamin Inc., New York (1968)
16. Wilson, H.R.: Psychophysical models of spatial vision and hyperacuity. In: Regan, D. (ed.) Spatial Vision, vol. 10, Visual Dysfunction, pp.64–86. Macmillan Press, London (1991)

Chapter 5
Actions

The simplest form of interaction with a biological object that an animal can make is a discrete behavioural event or an action. A geometric description of an action as a slow-fast system is introduced and applied to eye and arm movements. It is then extended to include the decision process of which action to make, which is described in terms of alternative trajectories.

5.1 Slow-Fast Systems

Any system which exhibits an alternation of slowly varying and rapidly varying behaviours can be modelled by a **slow-fast** system in which the rate of change of one of the differential equations describing the system is multiplied by a small amount. The simplest class of such a system consists of a pair of equations. Let the small parameter be denoted by ϵ then the pair of equations are:

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2) \\ \epsilon \frac{dx_2}{dt} &= f_2(x_1, x_2) \end{aligned} \tag{5.1}$$

where f_1 and f_2 are both functions of the variables x_1, and x_1 and ϵ is greater than zero but much less than 1.

The behaviour of the variable x_2 evolves on a much faster time scale that of the other variable, so x_2 is a referred to as a fast variable and x_1 is referred to as a slow variable. The parameter ϵ determines the timescale on which the fast variable behaves. If one can assume that the magnitude of the derivative of the fast variable is of order unity then the equation of the fast variable becomes:

$$0 \approx f_2(x_1, x_2) \tag{5.2}$$

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes
in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_5

So with a very small value for ϵ this equation is a close approximation to the equation:

$$0 = f_2(x_1, x_2) \tag{5.3}$$

which describes a curve on the plane, referred to as the **slow manifold**.

Two simple examples of slow-fast systems are defined by Eqs. 5.4 and 5.5 in which the slow manifold is specified by a cubic function [15]:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 \\ \epsilon \frac{dx_2}{dt} &= -(x_2^3 + x_2 + x_1) \end{aligned} \tag{5.4}$$

The state space behaviour of this system is illustrated in Fig. 5.1. Points in the plane will move rapidly onto the slow manifold and then home slowly to the stable equilibrium of the system, so the entire slow manifold in this case attracts points towards itself.

The second example system can be made simply by changing the sign of one of the terms in the equation for the slow manifold:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 \\ \epsilon \frac{dx_2}{dt} &= -(x_2^3 - x_2 + x_1) \end{aligned} \tag{5.5}$$

and the behaviour of this system is also illustrated in Fig. 5.1. In this system the slow manifold has two attracting portions and one repelling portion. Points in the plane move rapidly onto an attracting portion of the manifold and then slowly towards the fold between the attracting portion and the repelling portion. The point then jumps from the end of one attracting portion onto the other attracting portion and again continues slowly towards the repelling portion. So every point eventually becomes trapped in a loop leading to repetitive behaviour referred to as a **limit cycle**.

It turns out the slow-fast equations capture the abrupt threshold that was lacking in the model for an action developed at the end of the first chapter. If a constant is added to the slow equation of the second example system of Eqs. 5.5 the equilibrium position is moved away from the origin and the behaviour of the system is that required for an action.

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 - 1 \\ \epsilon \frac{dx_2}{dt} &= -(x_2^3 - x_2 + x_1) \end{aligned} \tag{5.6}$$

The behaviour of this system of equations is illustrated in the upper half of Fig. 5.2. The figure illustrates that Eq. 5.6. describes an action that involves a jump return to equilibrium. A smooth return can be made by using a slow manifold consisting of a

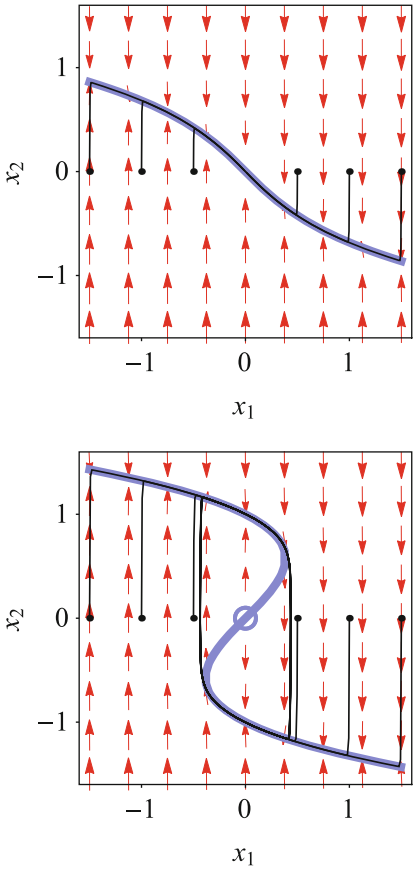


Fig. 5.1 Behaviour of two examples of a slow-fast system. The vector field defined by the system equations is represented by the red arrows. Six example trajectories are shown. Each trajectory begins at the point marked with a black dot. The slow manifold is marked by a thick blue line. **Upper Figure** The system defined by Eqs. 5.4 has a stable fixed point marked by a blue dot. The trajectories all end up at this point. **Lower Figure** In the system defined by Eqs. 5.5 the fixed point is marked by a blue circle to indicate that it is unstable. The trajectories all move towards the slow manifold and then follow a clockwise path along it until they reach a fold whereupon they follow the vector field onto another part of the slow manifold leading to cyclical behaviour

surface with a fold. An example of such behaviour is provided by the system three equations:

$$\begin{aligned} \frac{dx_1}{dt} &= -x_2 - 1 \\ \frac{dx_2}{dt} &= -x_2 - x_3 \\ \epsilon \frac{dx_3}{dt} &= -(x_3^3 + x_2 x_3 + x_1) \end{aligned} \tag{5.7}$$

and the behaviour of this system is illustrated in the lower portion of Fig. 5.2.

With this system of equations an action is deemed to have occurred when the state of the system is moved over the fold and drops rapidly back onto the slow manifold. The rapid change in state models the switching that occurs at the onset of an action. The threshold for initiating an action is specified by the locus of critical points on the slow manifold. The projection of this fold curve into the $(x_1(s), x_2(s))$

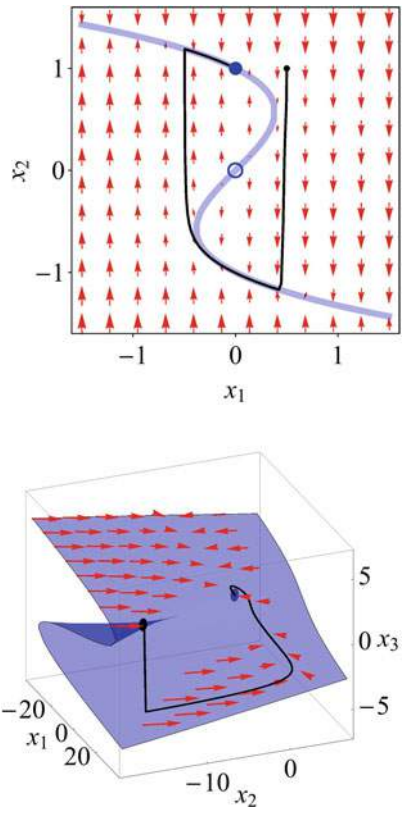


Fig. 5.2 Behaviour of two possible models for an action. **Upper Figure** Behaviour of Eq. 5.6, which has an abrupt return to equilibrium. **Lower Figure** Behaviour of Eq. 5.7, which has a smooth return to equilibrium

plane is described by the parametric equations

$$\begin{aligned} x_1(s) &= 2s^3 \\ x_2(s) &= -s^2 \end{aligned} \tag{5.8}$$

where s is a variable that describes how far along the fold curve the point $(x_1(s), x_2(s))$ is from the origin.

5.2 Velocity Commands

An example of an action is provided by the eye movements used to transfer gaze from one part of the scene to another. Such movements are referred to as **saccades**. The simplest case of purely horizontal movements can be used to illustrate the distinctive characteristics of saccadic eye movements. Saccades show a relatively invariant relationship between the size of the movement and the peak velocity and duration of the movement which is illustrated in Fig. 5.3. The peak velocity of

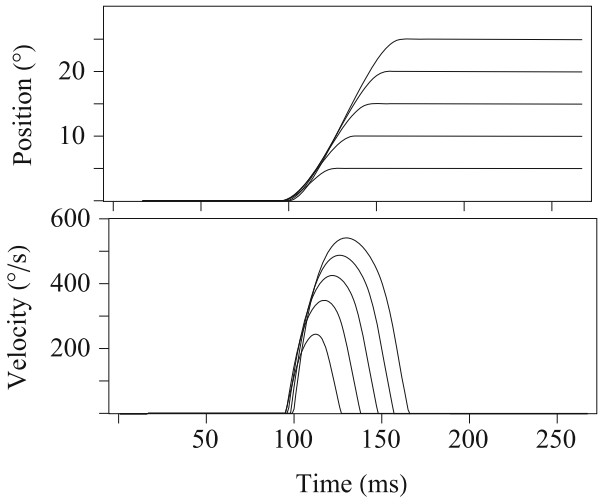


Fig. 5.3 Characteristics of saccadic eye movements. **Upper Figure** Changes of eye position with time for a range of saccade sizes. **Lower Figure** Corresponding changes of eye velocity as a function of time

saccades typically varies from thirty to seven hundred degrees per second and their duration varies from 30 to 100 ms for eye movements of a half to forty degrees in amplitude. The peak velocity progressively saturates with increasing saccade amplitude after 20 ° whereas the duration of the movements varies approximately linearly with amplitude. This consistent relationship between amplitude, duration and peak velocity of saccades is termed the **main sequence** [2].

The most direct route for signals involved in the control of saccades runs from the retina to the superior colliculus, then on to the brainstem, and terminates at the oculomotor nuclei. At the physiological level, neuronal recordings in awake animals have provided detailed information about the circuitry involved in the generation of saccades. The motoneurons that innervate the extraocular muscles are located in the nuclei of cranial nerves III, IV and VI and have a burst-tonic pattern of discharge. The rate at which spikes are produced during the burst correlates with eye velocity and the steady level of tonic activity correlates with the eye position at the end of the saccade [12, 13].

The burst of activity is generated by burst neurons which fire during saccadic movements in a preferred direction (their on direction) and are silent for movements in the opposite direction (their off direction). They can be classified into long-lead burst neurons, which steadily increase their firing before a saccade, and short-lead burst neurons, which only begin firing shortly before the start of the saccade. In addition to the burst units in the brainstem, there is also a class of cells referred to as omnipause neurons. These fire continuously except just before and during saccades in any direction, during which time they cease firing.

The action involved in generation of a saccade can be described by a slow/fast system of equations with three variables x_1, x_2 and x_3. In effect the x_1 variable holds the size of the required eye displacement and its value decreases according to the integral of a velocity command specified by the x_2 variable. For convenience, in the context of slow-fast modelling, the velocity command neurons will be referred to as **burst neurons** and the neurons specifying the slow manifold will be referred to as **pause neurons**.

The Eqs. 5.7 of the slow/fast action system have to be modified to include a time constant λ that can be set to ensure that the rate at which the velocity command changes matches that found in saccades.

$$\begin{aligned} \lambda \frac{dx_1}{dt} &= -x_2 - 1 \\ \lambda \frac{dx_2}{dt} &= -x_2 - x_3 \\ \lambda\epsilon \frac{dx_3}{dt} &= -(x_3^3 + x_2 x_3 + x_1) \end{aligned} \tag{5.9}$$

A value of 0.01 for ϵ ensures that the fast x_3 variable changes much faster than the other two variables and a value of 0.018 for λ results in simulated saccades which follow the main sequence for humans.

An advantage of specifying the velocity of the eye, as opposed to its position, is that the steady level of firing of the oculomotor neurons related to eye position can be generated by integrating the burst neuron signal. The **oculomotor integrator** output n is given by the integral of the velocity command y^+. As the value of the n variable corresponds to the position of the eye, the system of equations obtained by combining the three slow/fast Eqs. 5.9 with an equation for the oculomotor integrator:

$$\begin{aligned} \lambda \frac{dx_1}{dt} &= -x_2 - 1 \\ \lambda \frac{dx_2}{dt} &= -x_2 - x_3 \\ \lambda\epsilon \frac{dx_3}{dt} &= -(x_3^3 + x_2 x_3 + x_1) \\ \frac{dn}{dt} &= \kappa x_2^+ \end{aligned} \tag{5.10}$$

gives a model of the brainstem saccadic control circuit. The parameter κ scales the burst signal generated by the behavioural model to the eye velocities found experimentally, and this model was used to generate the representative curves shown in Fig. 5.3 [5].

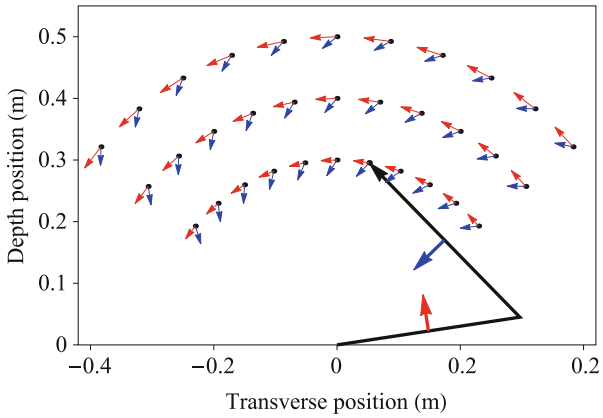


Fig. 5.4 Illustration of the movement of the hand resulting from a five degree change of the shoulder angle (red arrows) and of the elbow angle (blue arrows) from various starting positions. The relative orientations of the upper arm and lower arm are shown in black for an example position. The ratio of the sizes of the hand displacements resulting from the shoulder angle and elbow angle changes depending on the position of the hand

Arm movements are more complicated than eye movements because the displacement of the hand for a given set of joint angle changes varies with the arm configuration. This effect is illustrated in Fig. 5.4. This constraint is embodied in the responses of neurons in the motor cortex. In a reaching task, in which monkeys were trained to move their hand from a starting position straight ahead to a target on a diagonal, neurons in the primary motor cortex were found to have a preferred direction of movement for which they responded maximally and cosine tuning with respect to this preferred direction and these are the properties required for a population vector code of direction [7]. Subsequent recordings made during a drawing task revealed that the direction of the population vector correlated with the direction of the instantaneous displacement of the monkey's hand and its length correlated with the speed of hand movement [10]. However, when the arm movements were restricted to the horizontal plane, but required flexion of the elbow, then the population code was found to be no longer aligned with direction of the hand movement but with the relative changes in the joint angles [11] as illustrated in Fig. 5.5.

To ensure straight paths of hand movement, the velocity signal given to the joint angle integrators has to be scaled by the projection of the direction of hand movement resulting from a change in one of the joint angles onto the target direction

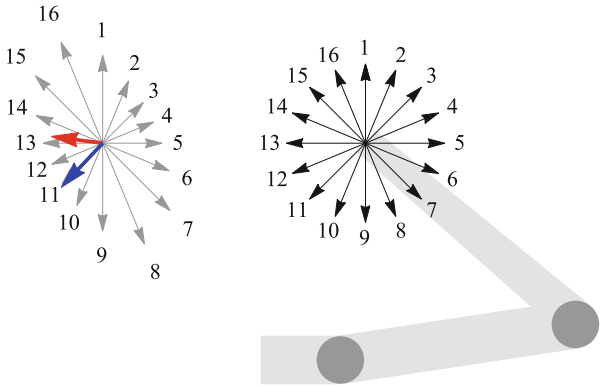


Fig. 5.5 Illustration of the alignment of the responses of a population of neurons in the motor cortex. Black arrows are used to indicate sixteen equal-sized hand movements made from the example starting position of Fig. 5.4. For a hand movement in a particular direction a fixed ratio of shoulder and elbow rotations is required, and this is specified by the orientation and size of the corresponding grey arrow shown on the left. Each of these vectors is a linear combination of the displacements due to a shoulder movement (red arrow) and an elbow movement (blue arrow) alone. The responses of neurons in the motor cortex were found to be aligned with the joint angle change vectors rather than with the hand displacement vectors [11]

(T_x,T_y). The model for eye movements can be adapted to incorporate this constraint giving a model for arm movements:

$$\begin{aligned}
\lambda \frac{dx_1}{dt} &= -x_2 - 1 \\
\lambda \frac{dx_2}{dt} &= -x_2 - x_3 \\
\lambda\epsilon \frac{dx_3}{dt} &= -(x_3^3 + x_2 x_3 + x_1) \\
\frac{d\phi}{dt} &= (j_{11}, j_{12}) \cdot (T_x, T_y)\kappa x_2^+ \\
\frac{d\theta}{dt} &= (j_{21}, j_{22}) \cdot (T_x, T_y)\kappa x_2^+
\end{aligned} \tag{5.11}$$

where ϕ, is the shoulder angle, θ is the elbow angle, j_{11}, j_{12}, j_{21}, and j_{22} are elements of the matrix of the transformation of changes in the (x,y) coordinates of the hand into changes in the (ϕ, θ) coordinates of the shoulder and elbow. This model was used to produced the simulations shown in Fig. 5.6. The behaviour of the eye movement variables are plotted alongside to illustrate the common role of burst and pause neurons in the two types of movement.

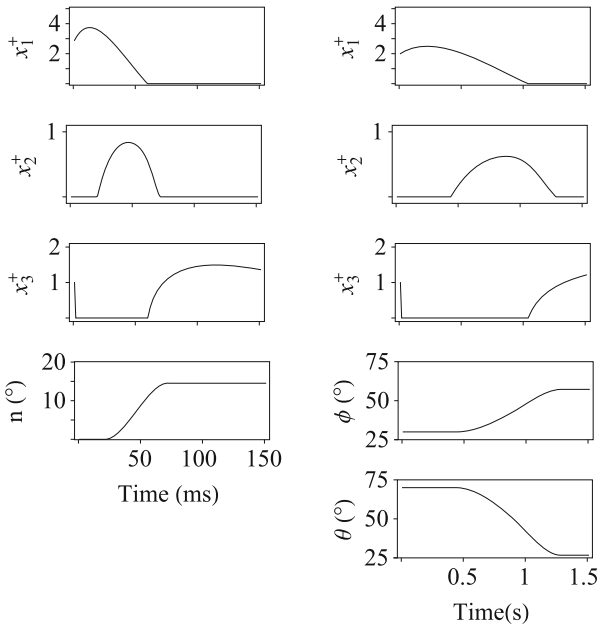


Fig. 5.6 Positive values of the slow-fast model variables for an eye movement, specified by Eqs. 5.10, and a horizontal arm movement, specified by Eqs. 5.11. The positive values of the variables x_1, x_2 and x_3 correspond to the firing rates of three classes of neurons. The x_1 neurons have a pattern of firing that matches the difference between the current eye or arm position and the required one, a quantity referred to as motor error. The x_2 neurons have a burst pattern of firing identical to the instantaneous velocity of the movement and the x_3 have a pause pattern of firing

Although the arm movement model appears different from the eye movement model in the involvement of the variables used in the neural integration, this difference becomes less if the eye movements involve both horizontal and vertical directions. In this case accurate integration of angular velocity requires knowledge of the current position [14], as in the arm movement model. However, in both cases further processing is required to convert a neural signal corresponding to the velocity of movement to one that further ensures the mechanical constraints of muscle and connective tissue are satisfied. In this respect arm movements are more complicated than eye movements because the load and shape of the arm can change leading to altered force requirements. The additional neural signals required to overcome these constraints are supplied in part by the stretch reflex, which the eye does not have, and in part by the cerebellum.

5.3 Deciding What to interact With

Whilst it is relatively straightforward to formulate a mechanism for carrying out an action, that still leaves open the question of how to select what to interact with. A simple example of this task consists of the decision as to which of two indistinct targets to look at; the one on the left or the one on the right. The longer one waits the more sure one becomes of where the target is so there is a trade-off between making an accurate eye movement and making a quick decision. This type of behaviour can be modelled by a pair of neurons, each of which integrates the evidence for one of the objects and is referred to in this context as an **accumulator**. When the level of one of the accumulators reaches a threshold then an eye movement to the corresponding target is made.

The build-up to an action during this decision phase can be incorporated into the slow-fast model of an action by adding an accumulator:

$$\begin{aligned}
\lambda \frac{\mathrm{d}a}{\mathrm{d}t} &= \Theta(a)z \\
\lambda \frac{\mathrm{d}x_1}{\mathrm{d}t} &= -x_2 - 1 \\
\lambda \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -x_2 - x_3 - \mu a \\
\lambda\epsilon \frac{\mathrm{d}x_3}{\mathrm{d}t} &= -(x_3^3 + x_2 x_3 + x_1) \\
\frac{\mathrm{d}n}{\mathrm{d}t} &= -\frac{n}{T_n} + \kappa y^+
\end{aligned} \tag{5.12}$$

where Θ(a) is the Heaviside theta function which is 1 if a is greater than zero and zero otherwise and μ is a positive constant. The Heaviside function ensures a steady build-up of the accumulator when the output of x_3 variable is positive and that the output of the accumulator never becomes negative. The amplitude of the saccade depends on the value of μ. The behaviour of this model is illustrated in Fig. 5.7. Unlike the original slow-fast model there is no step initiation of the action, but rather a steady build-up to the action.

In most cases an interaction with the world involves more than just a decision to move right or left. One way to be able to make decisions about whether to look at a target anywhere in the visual field is to have an accumulator for every direction, and this type of structure is found in the superior colliculus. The superior colliculus is a layered structure with visual cells located superficially and multi-sensory and motor cells located in intermediate and deeper layers. The visual cells in the superficial layers are arranged topographically so that the retinal area that must be stimulated for a cell to respond depends on the position of the cell in the colliculus. Electrical stimulation of the deep layers of the superior colliculus results in a saccade to a

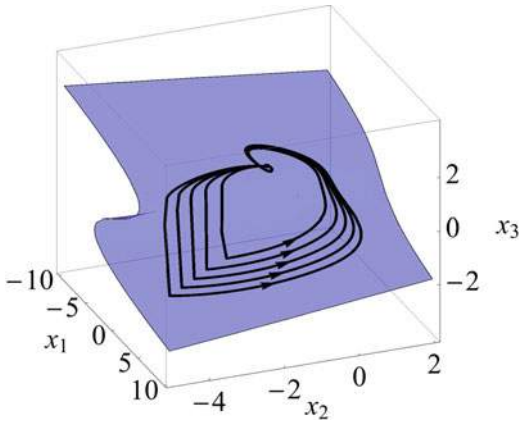


Fig. 5.7 Projection of trajectories of the 4-Dimensional system specified by Eqs. 5.12 into the 3-dimensional space of the slow-fast system of equations. The state of the system moves steadily away from the equilibrium state as the accumulator output increases

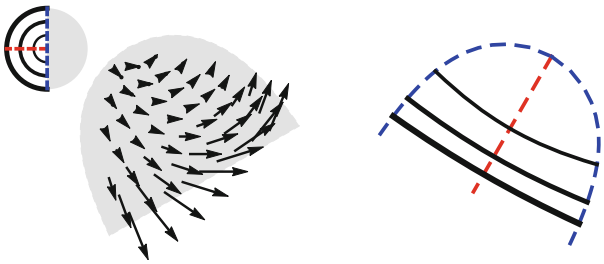


Fig. 5.8 Topographic map of the saccades elicited from one half of the superior colliculus. The inset picture illustrates the system of coordinates used to specify the direction of the line of fixation with respect to the head. The arrows on the left half of the colliculus, which is plotted in grey, indicate the direction and size of a saccade following electrical stimulation of the corresponding part of the colliculus. The right half of the colliculus has been used to indicate the positions of the horizontal meridian, plotted as a red dashed line and the vertical meridian, plotted as a blue dashed line. The black lines indicate points at ten, twenty and thirty degree angles with the line of fixation

position in the contralateral visual field which is approximately the same as that of the receptive fields in the overlying sensory layer, as illustrated in Fig. 5.8.

The classes of neurons found in the colliculus match those found in the brainstem. The build-up cells in the colliculus are distinguished from burst cells by showing a steady increase in firing rate tens of milliseconds before the onset of a saccade and the fixation cells behave like the omnipause neurons in that they fire steadily except during a saccade. The main difference between the collicular neurons and the brainstem neurons is that in keeping with the topographic organisation of the colliculus the build-up and burst cells respond maximally in

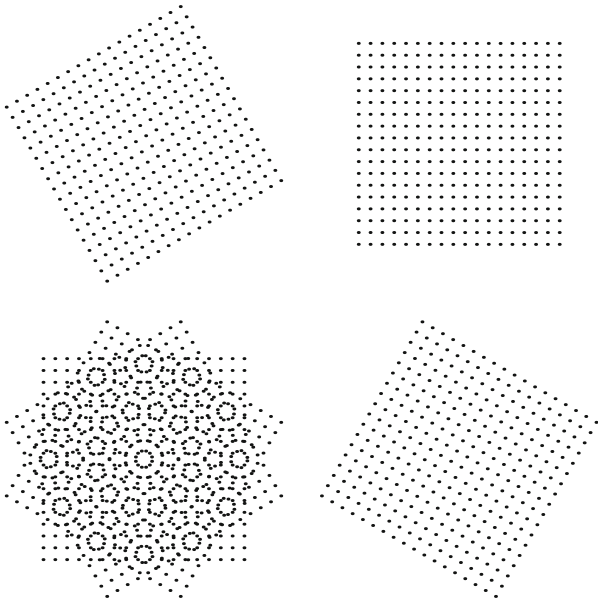


Fig. 5.9 Construction of the Marroquin pattern by overlaying three arrays of dots, differing in orientation by sixty degree steps. With a larger array of dots, fluctuating patterns of circles and polygons are seen [9]

association with eye movements of a particular amplitude and direction. Fixation cells increase their activity during periods of active fixation. These are found at the rostral pole of the colliculus, at the location of the collicular representation of the fovea.

But sometimes it is by no means obvious how a brain decides that a target is present. For example, in even a simple pattern of dots, such as the **Marroquin pattern**, [9] it can be difficult to say exactly what the pattern consists of, with circles and patterns of polygons appearing and disappearing all the time. This is surprising as the pattern is simply constructed by overlaying three rectangular grids with orientations differing by 60° as illustrated in Fig. 5.9.

When viewing the Marroquin figure subjects look predominantly at the centres of the circles, but the paths taken between circles vary between subjects as illustrated in Fig. 5.10. So successive cycles of a decision followed by an action do not appear to be deterministic and are better described by a probabilistic process. The nature of this probabilistic process is constrained by the distributions of the amplitudes of the saccades and the intervals between them which are also plotted in Fig. 5.10 [1].

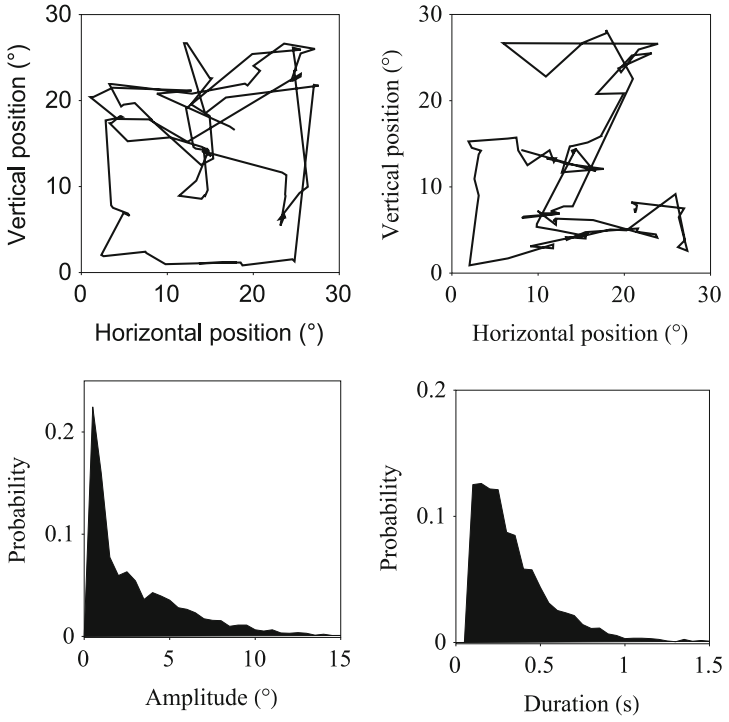


Fig. 5.10 Scanpaths made whilst viewing the Marroquin pattern. **Upper Figure** Illustration of the different scanpaths made by two different viewers. **Lower Figure** Saccade amplitude and fixation duration distributions, made by averaging the results from eight subjects [1]

These distributions are compatible with the idea that the exploration is carried in the same way as optimal foraging, in that the task is to visit as many target sites as possible whilst travelling the shortest distance possible. In particular, the long tail of the amplitude distribution is a signature of such a strategy [3].

But the control of saccades also involves many higher regions of the brain. The cortical frontal eye fields, supplementary eye fields and lateral intraparietal area are all involved in specification of the target position for a saccade, and signals from these areas are passed via the basal ganglia to the superior colliculus. In addition, the firing patterns of many neurons in the cerebellum correlate with eye movements, especially in the context of adaptive behaviour. Cells have also been found in the enthorinal cortex that act like grid cells for the direction of saccades in a scene [8]. In this context, it may be that it is difficult to assemble successive fixations of the Marroquin pattern precisely because the grids are spaced at orientations of sixty degrees apart which may overload the grid neurons.

5.4 Unexpected Consequences of Nonlinear Behaviour

At the end of the first chapter, the slow-fast model of an action was used to introduce the geometric approach and although the design of the model was straightforward it turns out that it has additional equilibrium behaviour to that intended. The unexpected behaviour occurs because the slow manifold obtained by putting ϵ equal to zero in the equation for the fast variable does not match what is actually happening in the model along the fold line. Along the fold of the slow manifold the partial derivative of the fast variable with respect to the slow variable x_1 ceases to be defined.

By projecting the vector field of the slow manifold onto the plane spanned by the x_2 and x_3 variables one can remove the singularities along the fold line and investigate the behaviour close to the fold. It transpires that there is an unstable equilibrium on the fold line that causes some trajectories to loop back as illustrated in Fig. 5.11 [4].

It is also possible for the trajectory to pass close to the fixed point on the fold line and pass onto the unstable portion of the slow manifold. This can lead to dramatic

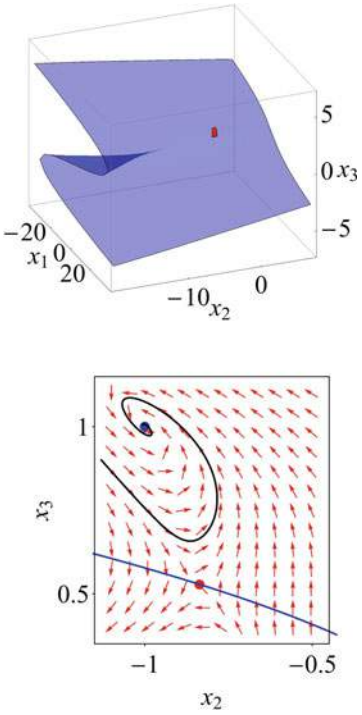


Fig. 5.11 Projection of the vector field on the slow manifold onto the plane spanned by the x_2 and x_3 directions. **Upper Figure** The region projected is indicated by the small red marking on the three-dimensional figure. **Lower Figure** Projection of a small region close to the equilibrium point of the slow manifold. The fold is marked by the blue line. Close to the unstable equilibrium position, marked by the red dot on the fold line, trajectories are attracted to the point along one diagonal and repelled along the opposite diagonal

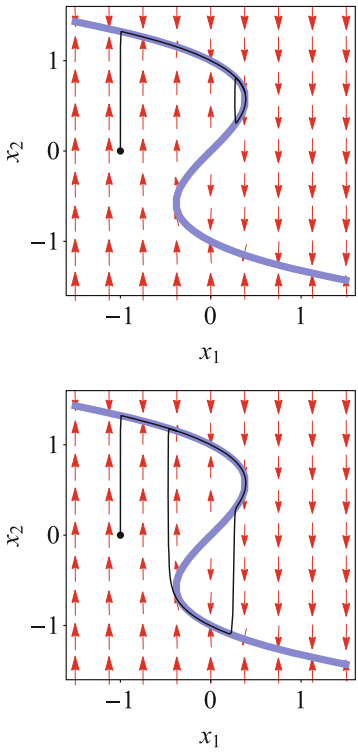


Fig. 5.12 Behaviour of Eqs. 5.13 showing a dramatic change in behaviour with a very small parameter change. **Upper Figure** Onset of small oscillations with $a = 0.5766230683749112$ **Lower Figure** Onset of much longer cycles with $a = 0.5766230683749111$ The abrupt transition from a small amplitude oscillation to a large amplitude oscillation, with no intervening medium-sized oscillation, is referred to as a canard [6]

changes in behaviour, which are easiest to illustrate in two dimensions using the system of equations:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_2 - a \\ \epsilon \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -(x_2^3 - x_2 + x_1) \end{aligned} \tag{5.13}$$

When $a = 0$ this system of equations becomes identical to the second example of a slow-fast system given by Eqs. 5.5 and the behaviour of the system is a limit cycle. However, if $a = 1$ this system of equations becomes identical to the first model for an action given by Eqs. 5.6 and the behaviour of the system consists of a trajectory towards a stable equilibrium. The transition between these two types of behaviour occurs around the value $a = 0.577$ when the trajectoriy starts to move onto the unstable portion of the slow manifold and is very abrupt as can be seen from Fig. 5.12. Such transitions are difficult to pick up experimentally, because they happen in such a narrow parameter range, but they do highlight how a sudden transition from stable to oscillatory behaviour can happen suddenly in nonlinear systems.

References

1. Akman, O.E., Clement, R.A., Broomhead, D.S., Mannan, S., Moorhead, I., Wilson, H.R.: Probing bottom-up processing with multistable images. J. Eye Move. Res. **1**(3), 1–7 (2009). https://doi.org/10.16910/jemr.1.3.4
2. Bahill, A.T., Clark, M.R., Stark, L. R.: The main sequence, a tool for studying human eye movements. Math. Biosci. **24**, 191–204 (1975). https://doi.org/10.1016/0025-5564(75)90075-9
3. Brockmann., D., Geisel., T.: The ecology of gaze shifts. Neurocomputing **32–33**, 643–650 (2000). https://doi.org/10.1016/S0925-2312(00)00227-7
4. Broer, H.W., Kaper, T.J., Krupa, M.: Geometric desingularization of a cusp singularity in slow-fast systems with applications to Zeeman's examples. J. Dyn. Differ. Equ. **25**, 925–928 (2013). https://doi.org/10.1007/s10884-013-9322-5
5. Clement, R.A., Akman, O. E.: Slow-fast control of eye movements: an instance of Zeeman's model for an action. Biol. Cybern. **114**, 519–532 (2020). https://doi.org/10.1007/s00422-020-00845-7
6. Diener, M.: The canard unchained *or* how fast/slow dynamical systems bifurcate. Math. Intell. **6**, 38–49 (1984). https://doi.org/10.1007/BF03024127
7. Georgopoulos, A., Kalaska, J., Caminiti, R., Massi, J.: On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci. **2**, 1527–1537 (1982). https://doi.org/10.1523/JNEUROSCI.02-11-01527.1982
8. Killian, N.J., Buffalo, E.A.: Grid cells map the visual world. Nat. Neurosci. **21**, 161–162 (2018). https://doi.org/10.1038/s41593-017-0062-4
9. Marroquin, J.L.: Human Visual Perception of Structure. Master's Thesis, Department of Electrical Engineering and Computer Science, MIT, New York (1976)
10. Schwartz, A.: Direct cortical representation of drawing. Nat. Rev. Neurosci. **265**, 540–542 (1994). https://doi.org/10.1126/science.8036499
11. Scott, S., Gribble, P., Graham, A., Cabel, D.: Dissociation between hand motion and population vectors from neural activity in motor cortex. Nature **413**, 161–165 (2001). https://doi.org/10.1038/35093102
12. Scudder, C., Kaneko, K., Fuchs, A.: The brainstem burst generator for saccadic eye movement: a modern synthesis. Exp. Brain Res. **142**, 439–462 (2002). https://doi.org/10.1007/s00221-001-0912-9
13. Sparks, D.: The brainstem control of saccadic eye movements. Nat. Rev. Neurosci. **3**, 952–964 (2002). https://doi.org/10.1038/nrn986
14. Tweed, D., Vilis, T.: Implications of rotational kinematics for the oculomotor system in three dimensions. J. Neurophysiol. **58**, 832–849 (1987). https://doi.org/10.1152/jn.1987.58.4.832
15. Zeeman, E.C.: Differential equations for the heartbeat and nerve impulse. In: Waddington, C.H. (eds.) Biological Processes in Living Systems, pp.8–67. Routledge, New York (1972). https://doi.org/10.4324/9781351297165

Chapter 6
Brain and Body

Given a specification of an action, in terms of a trajectory over a configuration space, additional control signals are required to ensure the mechanical constraints are obeyed throughout the movement. Linear modelling of biological tissue is introduced and applied to the orbital plant and the oculomotor neural integrator. The linear approach is extended to simple nonlinear behaviour, using the larynx as an example. Nonlinearities lead to a greater range of behaviours of which heteroclinic cycles form one example.

6.1 Mechanics of Biological Tissue

Animal movements are effected by neural control of patterns of contraction and relaxation of muscles. The force exerted by a muscle on an object attached to it is referred to as tension and it can be described by the effect of an actively contracting element on a combination of spring and viscous damping components.

The major part of the behaviour of a mechanical spring can be described by a linear relationship between the tension f of the spring and its extension x from its resting length:

$$f = \mathrm{k}x \tag{6.1}$$

where k is a constant referred to as the **stiffness** of the spring. A spring is portrayed diagrammatically by coil as this is the typically form of metal spring.

The damping force produced by the passive viscous tissues acts to oppose changes in the length of a muscle and is linearly proportional to the rate of change of the length of the muscle. The standard example of viscous damping is a mechanical bicycle pump which gives little resistance when you push the piston in slowly but a lot more if you do it quickly, and for this reason a viscous element

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_6

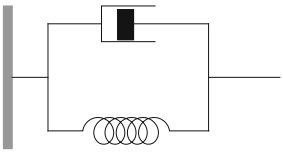


Fig. 6.1 Schematic diagram of a spring and viscous damper acting in parallel

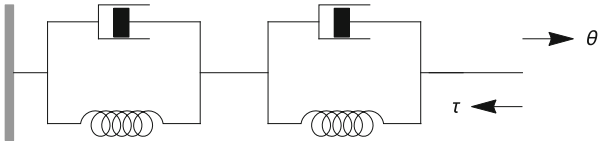


Fig. 6.2 A mechanical system equivalent to that of the oculomotor plant. The angle of rotation of the eye is specified by θ and the torque produced by the plant by τ. The system is comprised of two sub-systems acting in series. Each of the subsystems consist of a spring and a viscous element in parallel

is portrayed pictorially by a simplified cross-section of the piston mechanism. A viscous damping force f can be described by the differential equation:

$$f = \mathrm{r}\frac{\mathrm{d}x}{\mathrm{d}t} \tag{6.2}$$

where r is the **damping constant** of the element.

The mechanical system that has to be changed to obtain the required movement is referred to as the **plant**. The mechanical behaviour of biological tissue can often be described by a combination of spring and viscous damping elements acting in parallel, as illustrated in Fig. 6.1. For example, the spring and viscous components of the muscles and connective tissues that hold the eye in place can be described by a pair of spring and damper combinations acting in series as shown in Fig. 6.2.

As movements of the eye usually only involve rotations, the displacement of the eye can be described by a rotation angle and the forces resulting from the displacement act to rotate the eye and so are examples of torques. Let θ be the rotation of the eye away from the straight ahead direction and τ the torque produced by the stretching of the tissue attached to the eye, then the equation for the orbital plant can be written as:

$$((\mathrm{T}_1\mathrm{T}_2\frac{\mathrm{d}^2}{\mathrm{d}t^2} + (\mathrm{T}_1 + \mathrm{T}_2)\frac{\mathrm{d}}{\mathrm{d}t} + 1)\theta = \mathrm{k}(\mathrm{T}_3\frac{\mathrm{d}}{\mathrm{d}t} + 1)\tau \tag{6.3}$$

where T_1, T_2 and T_3 are time constants determined by the spring constants and viscosities of the two elements and k is an overall stiffness constant that depends on the stiffnesses of the individual springs. For the human eye the stiffness of the plant is approximately 0.56 N/rad. Typical values of the time constant in humans are 0.012, 0.260 and 0.072 [6].

The effect of neural signals on muscles was initially investigated by electrically stimulating an isolated muscle. A single pulse produces a brief, transient rise in tension known as a twitch, but if the pulse is repeated in a train the twitches merge and there is a sustained increase in tension referred to as tetanus. As the rate at which the pulses are repeated increases, eventually all trace of the individual twitches disappears and the muscle tension increases no further. Passive stretching of a contracted muscle is opposed by a force which can be modelled by an elastic element placed in series with the actively contracting portion of the muscle. Also when a stretched muscle is suddenly released there is an immediate reduction in the length of the muscle, and concomitant decrease of the force exerted by the spring component, followed by an exponential decrease in length while the tension remains constant at its new level. This can be explained by assuming that there is a spring and viscous damper in parallel with the actively contracting element [10]. These considerations lead to the mechanical model of a muscle illustrated in Fig. 6.3.

An example of how muscles act to produce movement is provided by saccadic eye movements. Combining the muscle and plant model, as illustrated in Fig. 6.4, gives a model for the mechanics of eye movements. Note that the part of the spring components of the muscle that acts along its whole length have been incorporated into the plant model [14].

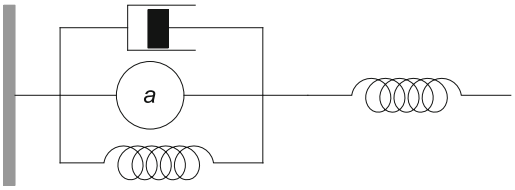


Fig. 6.3 Muscle model. The tension produced by the action of a neural signal on the muscle is denoted by a

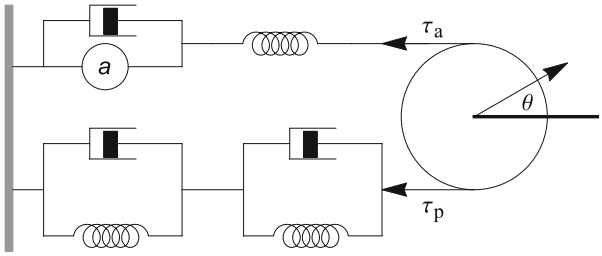


Fig. 6.4 A mechanical model of the extraocular muscles and passive tissues of the orbit acting in the horizontal plane. The angle of rotation of the eye is specified by θ and in the equilibrium position the torque τ_a generated by the muscle is equal to the opposing torque τ_p produced by the passive tissue

The equation for the muscle model can be written as:

$$f_m = -\frac{\mathrm{r}}{\mathrm{k}}\frac{\mathrm{d}f_m}{\mathrm{d}t} - \mathrm{r}\frac{\mathrm{d}\theta}{\mathrm{d}t} - a \tag{6.4}$$

As the forces on the mass of the globe must balance one has that:

$$\mathrm{m}\frac{\mathrm{d}x^2}{\mathrm{d}t^2} = f_m - f_p \tag{6.5}$$

and so one can rewrite Eq. 6.6 with the passive force f_p expressed in terms of the force developed by the muscle:

$$((\mathrm{T}_1\mathrm{T}_2\frac{\mathrm{d}^2}{\mathrm{d}t^2} + (\mathrm{T}_1 + \mathrm{T}_2)\frac{\mathrm{d}}{\mathrm{d}t} + 1)\theta = \mathrm{k}(\mathrm{T}_3\frac{\mathrm{d}}{\mathrm{d}t} + 1)(f_m - \mathrm{m}\frac{\mathrm{d}x^2}{\mathrm{d}t^2}) \tag{6.6}$$

One implication of this equation that the eye will not stay in a particular position unless there is a constant force which overcomes the force due to the spring extension. As was pointed out in the previous chapter, the correct neural signal for the constant force can be obtained by integrating the eye velocity command, rather than specifying the eye position independently of eye velocity.

But the mechanics of the oculomotor plant are such that the response of the system also depends on the rate of change of the applied force and to ensure the velocity of a saccade matches that specified by the burst neurons an additional slide component of neural activity is needed. An example of the profiles of the velocity pulse, neural integrator step and slide components are shown in Fig. 6.5. Both the required slide and integrated velocity command signals can be calculated by linear transformations of the velocity command [12] and the neural implementation of the velocity integration will be outlined in the next section.

6.2 Linear Behaviour

The behaviour of a mechanical system is encapsulated in Newton's three laws. In particular, Newton's second law states that for a particle of mass m at a position x the effect of a force f is to cause an acceleration of the particle

$$\mathrm{m}\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = f \tag{6.7}$$

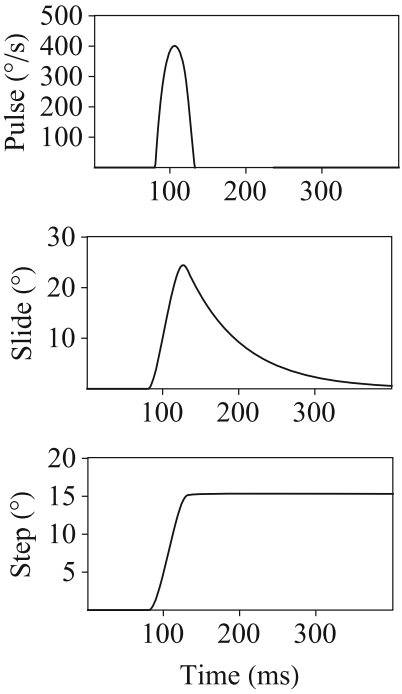


Fig. 6.5 Components of the neural signal required for a fifteen degree saccade. The components can be distinguished by the shapes of their graphs. The pulse component is given by the neural signal that specifies the velocity of the eye movement. The slide component is required to overcome the changes in the viscous forces on the eye, which last well beyond the end of the pulse. The remaining step component is formed by integrating the velocity signal. This position signal is needed to balance the changes in the spring forces resulting from the eye movement. The required signal for the saccade is given by the sum of these components

This second order equation can be rewritten as a system of first order differential equations:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}y} &= y \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= \frac{f}{\mathrm{m}} \end{aligned} \tag{6.8}$$

where the pair of values (x, y) correspond to the instantaneous position and velocity of the system and specify the state of the system.

The simplest form of this system of equations consists of a pair of linear differential equations with constant coefficients. This form of the system of differential equations can be written in matrix notation as:

$$\begin{pmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} \mathrm{a}_{11} & \mathrm{a}_{12} \\ \mathrm{a}_{21} & \mathrm{a}_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.9}$$

where the matrix elements a_{11}, a_{11}, a_{11} and a_{22} are all constants.

Let V be an eigenvector of the matrix A and let λ be the corresponding eigenvalue, so that:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{6.10}$$

Then a solution to the system of differential equations is given by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = e^{\lambda t} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{6.11}$$

since differentiating both sides of the solution with respect to t gives:

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \end{pmatrix} = \lambda e^{\lambda t} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = e^{\lambda t} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.12}$$

All the possible solutions can be made by linear combination of the two eigenvectors λ_1 and λ_2 of the system:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = p\lambda e^{\lambda t} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = q\lambda e^{\lambda t} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{6.13}$$

where p and q are constants.

In the case of a two dimensional linear system with constant coefficients the manifold consisting of all the possible states has the simple form of a plane, so it is straightforward to make a visual representation of the vector field of the system by computing $(\frac{dx_1}{dt}, \frac{dx_2}{dt})$ at an array of values (x_1, x_2) and then plotting the values of the derivatives as vectors, as illustrated in Figs. 6.6 and 6.7.

If both of the eigenvalues are real and positive then the arrows point away from the origin and the system is said to have a **source** at the origin. The simplest such system is described by the system of equations:

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.14}$$

Similarly if both of the eigenvalues are real and negative then the arrows point towards the origin and the system is said to have a **sink** at the origin. The simplest such system is described by the system of equations:

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.15}$$

A third possibility is that both the eigenvalues are real but of opposite sign so that the arrows point towards the origin along one direction and away from the origin

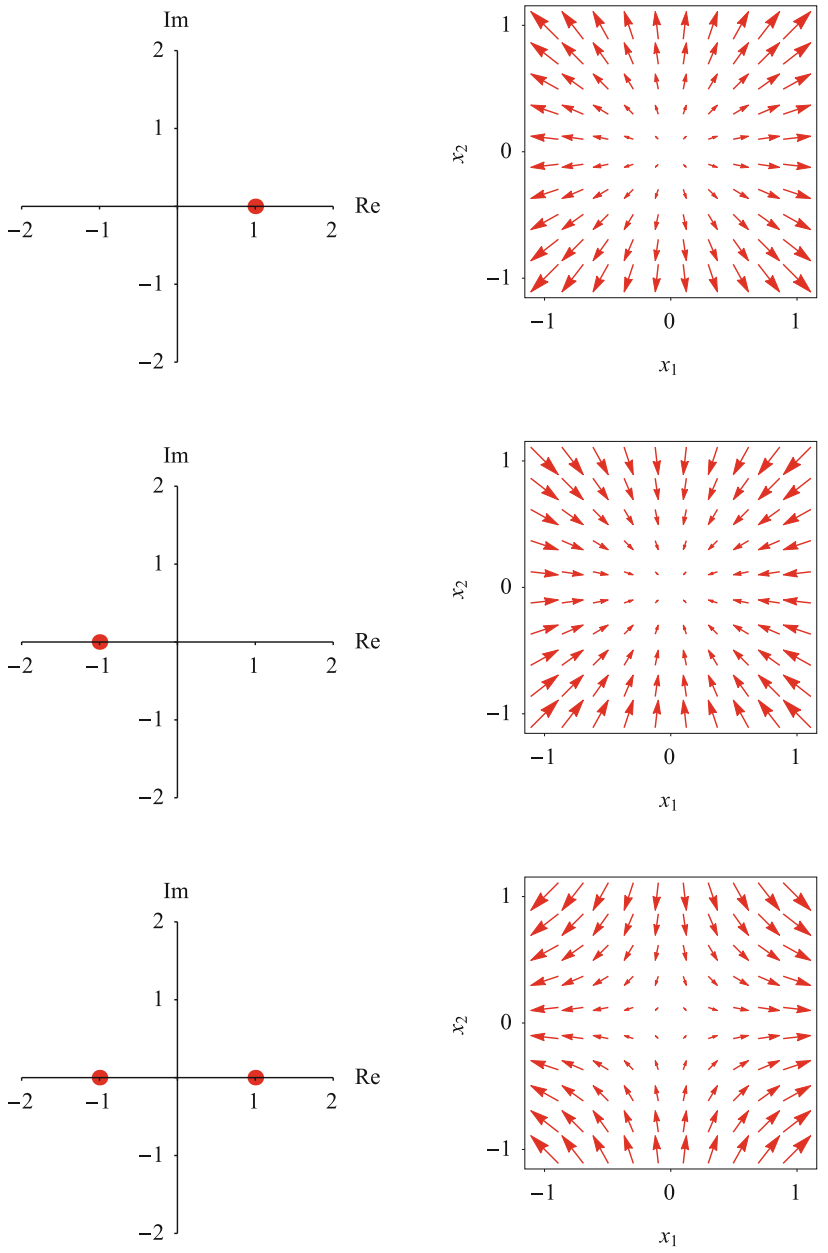


Fig. 6.6 Plots of the eigenvalues and vector fields associated with a source, sink and saddle.The directions and lengths of each of the arrows represent the vectors associated with a sample of points from the vector field

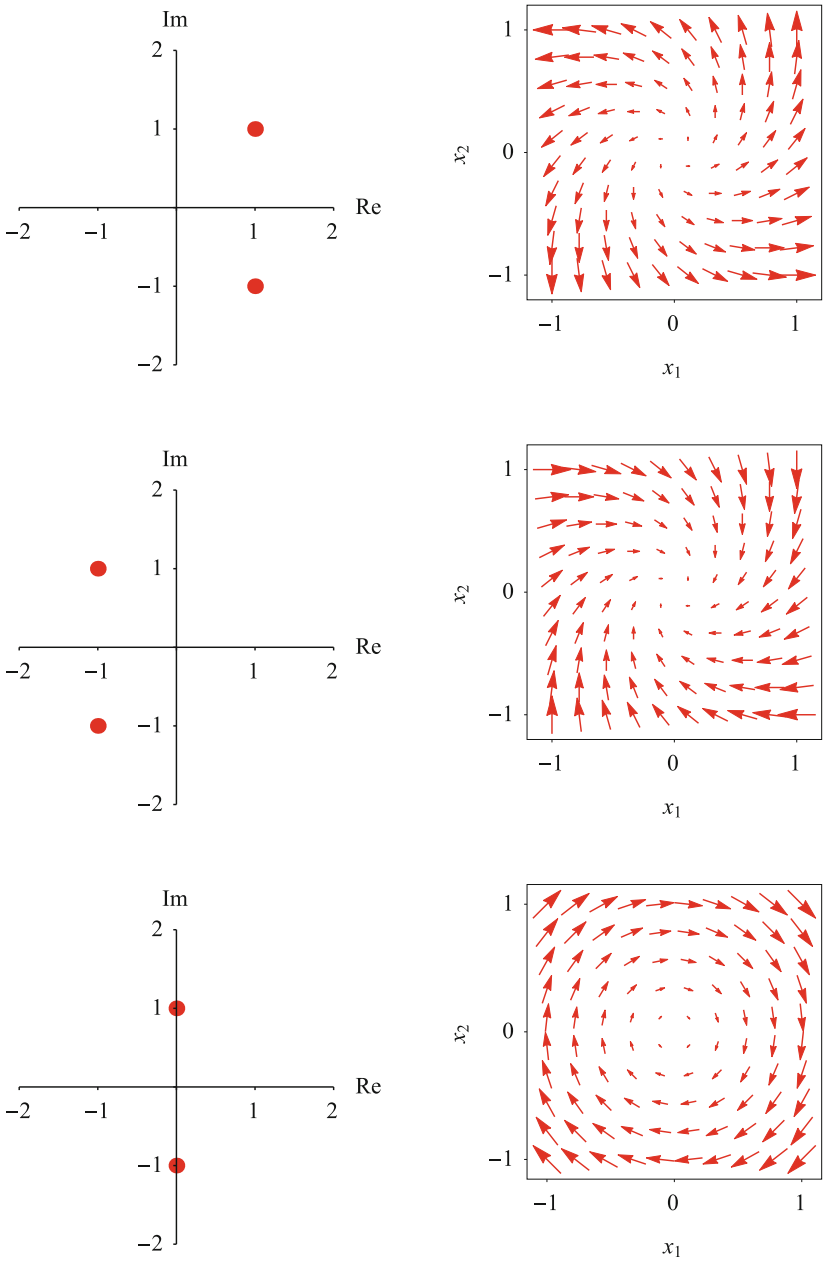


Fig. 6.7 Plots of the eigenvalues and vector fields associated with a spiral source, spiral sink and centre

along another direction. Such a system is said to have a **saddle** at the origin and the simplest example is described by the system of equations:

$$\begin{pmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.16}$$

If an eigenvalue $\lambda = \mathrm{a} + \mathrm{bi}$ is complex then each eigenvalue produces two real solutions. Starting with the solution:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathrm{e}^{(a+b\mathrm{i})t} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{6.17}$$

one can apply Euler's formula to the complex exponential to obtain an alternative form for the solution:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathrm{e}^{\mathrm{at}}(\cos(\mathrm{b}t) + \mathrm{i}\sin(\mathrm{b}t)) \times \left(\mathrm{Re} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \mathrm{Im} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) \tag{6.18}$$

Separating real and imaginary parts gives two solutions:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathrm{e}^{\mathrm{at}} \cos(\mathrm{b}t) \mathrm{Re} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} - \mathrm{e}^{\mathrm{at}} \sin(\mathrm{b}t) \mathrm{Im} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{6.19}$$

The complex conjugate eigenvalue gives the same solutions but with different signs.

If both of the real parts of the eigenvalues are positive then the arrows point away from the origin and the system is said to have a **spiral source** at the origin. A simple example of such a system is described by the system of equations:

$$\begin{pmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.20}$$

Similarly if both of the eigenvalues are real and negative then the arrows point towards the origin and the system is said to have a **spiral sink** at the origin. A simple example of such a system is described by the system of equations:

$$\begin{pmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.21}$$

A third possibility is that both the eigenvalues are imaginary but of opposite sign. In this case the arrows lie tangent to circles centred at the origin. Such a system is said to have a **centre** at the origin and the simplest example is described by the system of equations:

$$\begin{pmatrix} \frac{\mathrm{d}x_1}{\mathrm{d}t} \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6.22}$$

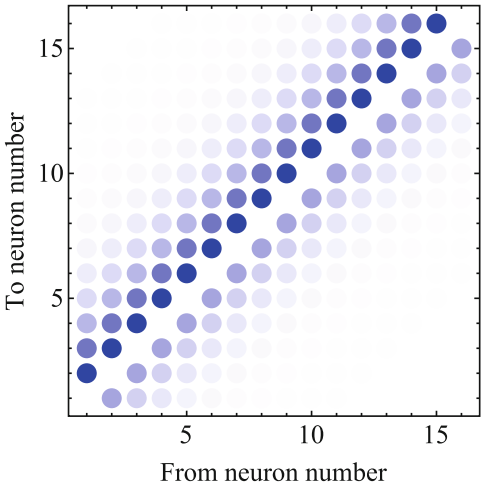


Fig. 6.8 Graphical illustration of the strengths of the connections in a model neural integrator consisting of sixteen neurons arranged in a row. The saturation of the blue dot with coordinates (i, j) represents the strength of the connection between the ith and jth neurons in the row. The ith column represents all connections from the ith neuron in the row. Each neuron has the same pattern of connections so the dot pattern simply shifts vertically with successive neurons in the row

Initial ideas about the neural mechanism of the oculomotor integrator were based on the ideas of positive feedback to a single neuron or mutual inhibition between pairs of antagonistic neurons that were introduced in the first chapter [5]. But simultaneous measurements of the behaviour of many neurons in the oculomotor integrator revealed that their firing persists during steady fixation over a wide range of times with the shortest times being associated with neurons located dorsally and rostrally and the longest times with neurons located ventrally and caudally [11]. A simple model for the neural integrator consists of a row of neurons in which there are stronger excitatory connections going forwards along the row and a weaker excitatory connections going backwards [7, 11]. A possible set of relative strengths of the connections for a row of sixteen neurons is illustrated in Fig. 6.8.

The behaviour of the model is described by a system of linear differential equations:

$$\begin{pmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \cdot \\ \cdot \\ \frac{dx_n}{dt} \end{pmatrix} = \begin{pmatrix} w_{11} & w_{12} & \cdot & w_{1n} \\ w_{21} & w_{22} & \cdot & w_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ w_{n1} & w_{n2} & \cdot & w_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} + \begin{pmatrix} s_1 \\ \cdot \\ s_k \\ 0 \\ 0 \end{pmatrix} \tag{6.23}$$

In the model only the first few inputs s_1 to s_k are active. If one applies a pulse to the network then as one goes along the row the pulse occurs later and is more spread out in time. Summing all the outputs of the neurons gives an integrated signal. This behaviour is illustrated for a 128 neuron model in Fig. 6.9.

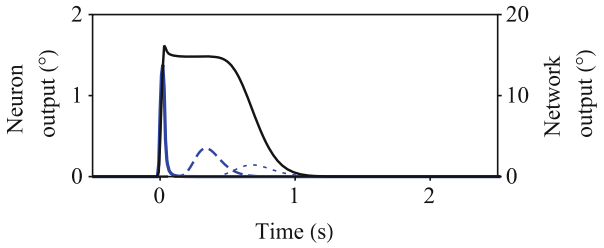


Fig. 6.9 Response of a neural integrator model consisting of 128 neurons with the pattern of connections illustrated in Fig. 6.8. Only the first four neurons in the row received a velocity input and the time constant of the individual neurons was 5 ms. The responses of the first, middle and last neurons in the row are plotted in blue. The network output is given by the sum of the individual responses is plotted in black

6.3 Nonlinear Behaviour

To recapitulate, in geometrical terms, a dynamical system specifies how every point on a manifold will move. As the instantaneous change in the position is equivalent to a vector, a dynamical system effectively assigns a vector to every point on the manifold. A dynamical system can therefore be considered to be a manifold M with a vector field F defined on it. Let the coordinates of a point on the manifold be given by an n-tuple of numbers $\mathbf{x} = (x_1, x_2, \ldots x_n)$, then the behaviour of the dynamical system can be described by a set of equations which specifies how the system changes from one state to the next as time goes by:

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, x_2, \ldots, x_n) \\ \frac{dx_2}{dt} &= f_2(x_1, x_2, \ldots, x_n) \\ &\ldots \\ \frac{dx_n}{dt} &= f_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{6.24}$$

where $(x_1, x_2, \ldots x_n)$ are all functions of time.

Two relatively simple examples of two-dimensional dynamical systems are provided by the slow-fast systems specified by Eqs. 5.4 and 5.5. As illustrated in Fig. 5.1 they have quite different behaviours in that in the first system all the trajectories end up at an equilibrium whereas in the second system all the trajectories end up in a limit cycle, and yet their equations are not so different. One way of understanding why the behaviour has changed is to begin by investigating what is required for the behaviour of the system to persist despite small changes in the equations, a condition referred to as being **structurally stable**.

The stability of the system of equations at a fixed point $\mathbf{x}*$ can be investigated by assuming that the behaviour of the nonlinear system close to the equilibrium state is the same as that of the linear system obtained by projecting the trajectories onto the tangent space to the manifold at the fixed point. The projection is defined by the **Jacobian matrix** J of the partial derivatives of the functions $(f_1, f_2, \ldots f_\mathrm{n})$ of the coordinates evaluated at the fixed point $\mathbf{x}*$:

$$\mathsf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}\Big|_{x*} & \frac{\partial f_1}{\partial x_2}\Big|_{x*} & \cdots & \frac{\partial f_1}{\partial x_\mathrm{n}}\Big|_{x*} \\ \frac{\partial f_2}{\partial x_2}\Big|_{x*} & \frac{\partial f_2}{\partial x_2}\Big|_{x*} & \cdots & \frac{\partial f_2}{\partial x_\mathrm{n}}\Big|_{x*} \\ \cdots & & & \\ \frac{\partial f_\mathrm{n}}{\partial x_1}\Big|_{x*} & \frac{\partial f_\mathrm{n}}{\partial x_2}\Big|_{x*} & \cdots & \frac{\partial f_\mathrm{n}}{\partial x_\mathrm{n}}\Big|_{x*} \end{pmatrix} \tag{6.25}$$

and with the linearisation assumption the system of Eqs. 6.24 for the dynamical system simplify to a matrix equation at the fixed point.

$$\frac{\mathrm{d}(\mathsf{X} - \mathsf{X}*)}{\mathrm{d}t} = \mathsf{J}(\mathsf{X} - \mathsf{X}*) \tag{6.26}$$

Given the linear system of Eqs. 6.26 one can go to investigate the stability of the equilibrium state by calculating the eigenvalues of the Jacobian matrix. If all the real parts of the eigenvalues are negative then the fixed point is stable, but if the real part of any eigenvalue is positive then the fixed point is unstable. The stability defined by the eigenvalues is referred to as **linear stability** to distinguish it from structural stability .

There is a problem with the linearisation approach in that it does not say how to classify a centre for which the real part of the eigenvalues is zero. The underlying reason for this drawback is that the behaviour of a system with a centre is not structurally stable in that a small changes in the equations can lead to an abrupt switch from stable to unstable behaviour, or vice versa, at the fixed point. Such qualitative changes of behaviour are referred to as a **bifurcation**.

The possible bifurcations of system can be investigated by altering some parameter μ which affects the behaviour of the system. The system of n first order differential equations with a parameter μ can be written as:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= f_1(x_1, x_2, \ldots, x_\mathrm{n}, \mu) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= f_2(x_1, x_2, \ldots, x_\mathrm{n}, \mu) \\ &\cdots \\ \frac{\mathrm{d}x_\mathrm{n}}{\mathrm{d}t} &= f_\mathrm{n}(x_1, x_2, \ldots, x_\mathrm{n}, \mu) \end{aligned} \tag{6.27}$$

To illustrate this approach, take the two examples of slow-fast systems specified by Eqs. 5.4 and 5.5 and treat them as a single system with a bifurcation parameter μ:

$$\begin{aligned} \frac{dx_1}{dt} &= x_2 \\ \epsilon \frac{dx_2}{dt} &= -(x_2^3 + \mu x_2 + x_1) \end{aligned} \tag{6.28}$$

The system has a fixed point at the origin which is stable when μ is greater than 0 and unstable when μ is less than 0. When μ is 0 a centre is formed. The changes in the eigenvalues of the Jacobian matrix of the fixed point as the value of μ is altered from 0.25 to 0 to $-.25$ and the corresponding vector fields of the system are illustrated in Figs. 6.10, 6.11 and 6.12. This type of bifurcation is referred to as a **Hopf bifurcation**

The vector field close to the fixed point can be described as a vector field over the product of three manifolds. The stable and unstable manifolds are associated with the trajectories approaching and leaving the equilibrium state respectively and the **centre manifold** is defined by the collection of trajectories that neither approach or move away from the equilibrium state. Because the lengths of the vectors attached to the stable and unstable manifolds change exponentially with distance from the

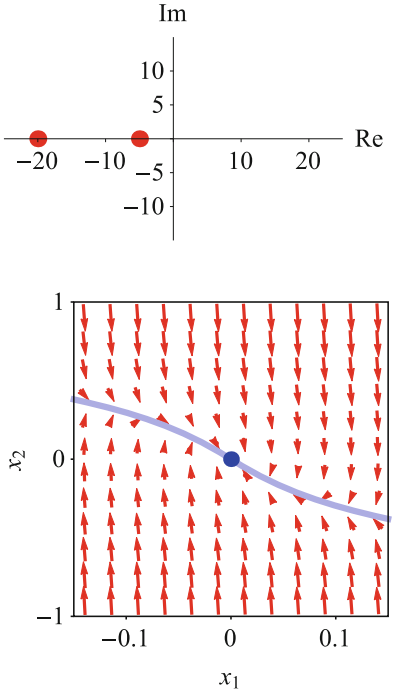


Fig. 6.10 Behaviour of Eq. 6.28 with the parameter μ equal to 0.25. **Upper Figure**: Eigenvalues of the Jacobian matrix of the system evaluated at the fixed point at the origin. In this and the subsequent two figures the eigenvalues are plotted in red. With the given parameter value, both eigenvalues are less than zero and the system is stable at the origin. **Lower Figure**: The vector field close to the fixed point. The state of the system moves rapidly onto the slow manifold, shown in blue, and then along the slow manifold to the stable fixed point at the origin, shown by a filled dot

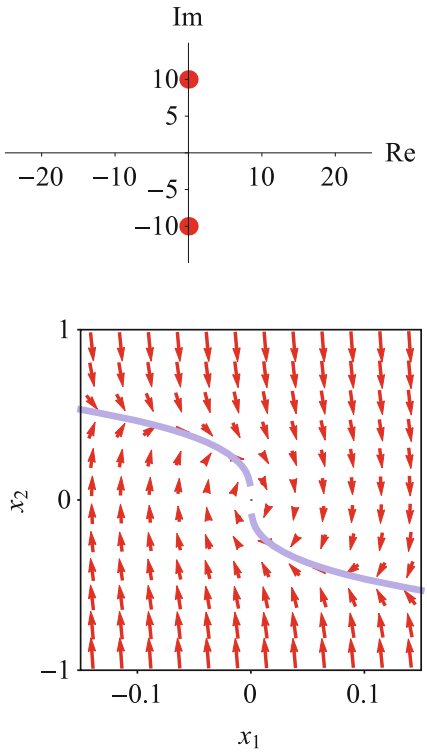


Fig. 6.11 Behaviour of Eq. 6.28 with the parameter equal to 0. **Upper Figure**: Eigenvalues of the Jacobian matrix, which consist of a pair of purely imaginary numbers. **Lower Figure**: Diagram of the vector field of the system. Away from the fixed point the state of the system moves rapidly onto the slow manifold, shown in blue. But close to the fixed point states follow an elliptical trajectory centred on the fixed point

equilibrium state, close to the equilibrium state the behaviour of the system is determined by the trajectories which make up the centre manifold.

6.4 Production of Voice Sounds

Voice sounds are produced by forcing air through the trachea resulting in oscillatory movements of the vocal folds of the larynx and a schematic illustration of a single cycle of the oscillation is shown in Fig. 6.13. The frequency of oscillation depends on both the stiffness of the muscles in the folds and the pressure with which air is forced through the folds so a quantitative description of the sound production has to include both of these components. To put such a description in context, it should be borne in mind that the pressure oscillations at the vocal folds are subsequently modified as they pass through the throat and mouth, where movement of the tongue can be used to produce the different vowel and consonant sounds of speech.

Each fold of the larynx can be modelled by a combination of mass, spring and viscous damping elements with the masses connected by a spring as illustrated in Fig. 6.14. This model captures the changes in the relative positions of the front

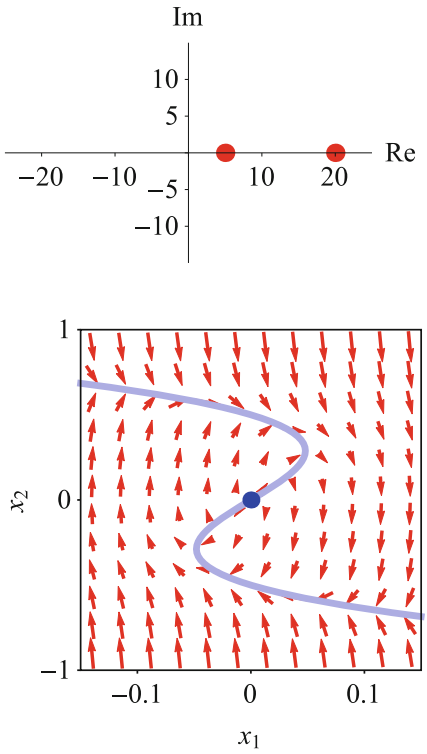


Fig. 6.12 Behaviour of Eq. 6.28 with the parameter μ equal to -0.25. **Upper Figure**: Eigenvalues of the Jacobian matrix, which are both greater than zero indicating that the system is unstable at the origin. **Lower Figure**: Diagram of the vector field of the system. The state of the system moves rapidly onto the attracting portions of the slow manifold, shown in blue, and then along the slow manifold to a limit cycle

and back of the folds as the air is forced through. However the linear model is not correct because it does not include the increased stiffness that occurs when the two folds collide. This increased stiffness can be included in the model by adding a function g to the linear equations specifying a threefold increase in stiffness during the collision. The pressure f due to the air flow through the larynx is assumed to be smooth until after it passes through the narrowest part of the larynx whereupon the flow forms a jet. The pressure of the jet is equal to that of the cavity above the vocal folds, which is assumed to be zero.

The behaviour of this system can be described by a system of four differential equations [15]:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= y_1 \\ \mathrm{m}_1 \frac{\mathrm{d}y_1}{\mathrm{d}t} &= -\mathrm{r}_1 y_1 - \mathrm{k}_1 x_1 - \mathrm{k}_\mathrm{c}(x_1 - x_2) - g_1 + f \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= y_2 \\ \mathrm{m}_2 \frac{\mathrm{d}y_2}{\mathrm{d}t} &= -\mathrm{r}_2 y_2 - \mathrm{k}_2 x_2 - \mathrm{k}_\mathrm{c}(x_2 - x_1) - g_2 \end{aligned} \tag{6.29}$$

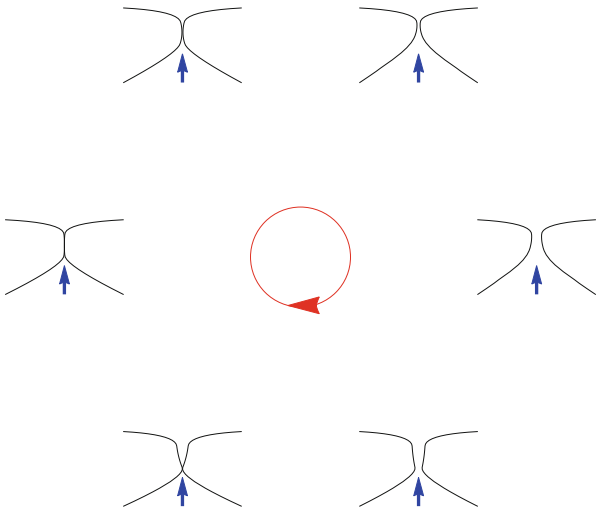


Fig. 6.13 A single cycle of the oscillatory movement of the larynx. A cross section of the larynx is drawn in black, with the direction of airflow illustrated by a blue arrow and the direction of the cycle indicated by the arrow on the red circle. In the position shown at the top left of the figure, the larynx is at rest. The lower portion of the larynx begins widening until the airway is fully open. The lower portion then begins closing until the two folds collide and then settle back into their resting state. A mathematical description has to capture the way in which the movement of the upper parts of the folds follows that of the lower parts

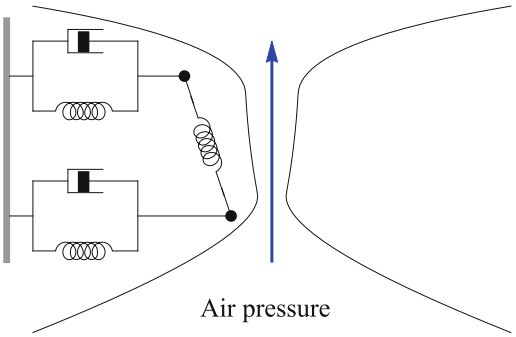


Fig. 6.14 Linear mechanical model of the larynx in which separate mass, spring and viscous damping elements are used to model the behaviour of the upper and lower parts of the folds. The mechanical link between the upper and lower parts is modelled by including a spring element joining the two masses

A fixed point of the system of equations is a solution to the system of equations:

$$\begin{aligned} 0 &= y_1 \\ 0 &= -\mathrm{k}_1 x_1 - \mathrm{k}_\mathrm{c}(x_1 - x_2) + f \\ 0 &= y_2 \\ 0 &= -\mathrm{k}_2 x_2 - \mathrm{k}_\mathrm{c}(x_2 - x_1) \end{aligned} \tag{6.30}$$

Although there is more than one solution to this equation, only the fixed point at $(x_1, x_2) = (0, 0)$ does not require negative pressure and so is the only possible solution given the airflow assumptions of the model. The fixed point corresponds to a position where the vocal folds are just separated in their mechanical resting position.

The importance of identifying the Hopf bifurcation that occurs at the onset of phonation is that one can go on to to find the parameter values that must be reached for phonation to occur. For the model larynx there are two parameters that can be varied by neural commands: the subglottal pressure and the stiffness of the muscles in the vocal folds. An exploration of the role of the air pressure is illustrated in Fig. 6.15. With increased airflow pressure the system undergoes a Hopf bifurcation giving rise to oscillatory movements of the vocal folds that increase in amplitude with the air pressure.

There are many other bifurcation mechanisms than just the Hopf bifurcation and a direct approach to uncovering the range of behaviours inherent in a model is to investigate the possible bifurcations of the system of equations. This knowledge

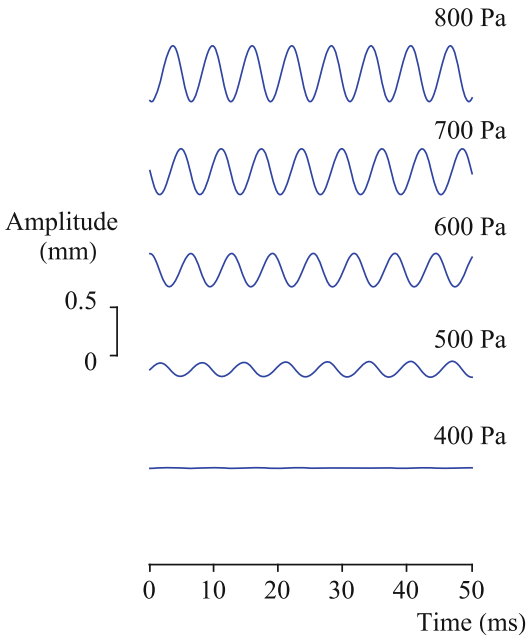


Fig. 6.15 Development of oscillatory behaviour of the lower vocal fold through a Hopf bifurcation with increasing airflow through the trachea

is a prerequisite for ensuring that the quantitative description can account for the entire range of behaviours found experimentally. In this context models involving the bilateral organisation of the brain are particularly apposite as even a slight imbalance in the symmetry often leads to a change of behaviour. Such analyses have been carried for models of the larynx in humans [15] and the corresponding syrinx in birds [2] and for models of the oculomotor system [1, 4]. In the case of models of the larynx and oculomotor system the predicted bifurcations of the models give rise to behaviours which are found clinically. However it is difficult to test if the bifurcation structures resulting from changes in the parameters of the models match those found clinically as the behaviour of each individual will typically correspond to the behaviour of a model with a fixed set of parameters so there is no direct way of finding out what bifurcation occurs in an individual when the parameters are changed. The situation is better for models of birdsong involving the syrinx, during which different sounds can be produced by varying the pressure of airflow through the syrinx and the stiffness of the muscles of the syrinx so that the bifurcation structure found in a given bird can be compared directly with that predicted experimentally [2]. One advantage of having a validated model is that it can then be used to produce synthetic birdsong for use in experimental investigations, where removal of the extraneous fluctuations of normal birdsong leads to more reliable experimental results [3].

6.5 Heteroclinic Cycles

Although the behaviour of simple systems often consists of trajectories leading to a stable equilibrium or a limit cycle in more complicated systems other types of stable behaviour become possible. In particular saddle points enable a trajectory to leave the point along an unstable direction and return to it along a stable direction and when several saddle points are present a trajectory can leave one saddle point along an unstable direction and approach another saddle point along a direction which is stable. Such trajectories are referred to as a **heteroclinic** trajectories. The importance of heteroclinic trajectories is that a stable heteroclinic cycle can be used to embody components of behaviour that go together.

A well-defined example of a stable heteroclinic cycle forms part of the behaviour of the Lotka-Volterra equations used to describe predator-prey population dynamics [8]. The example consists of a system of three equations:

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_1(1 - w_{1,1}x_1 - w_{1,2}x_2 - w_{1,3}x_3) \\
\frac{\mathrm{d}x_2}{\mathrm{d}t} &= x_2(1 - w_{2,1}x_1 - w_{2,2}x_2 - w_{2,2}x_3) \\
\frac{\mathrm{d}x_3}{\mathrm{d}t} &= x_3(1 - w_{3,2}x_1 - w_{3,2}x_2 - w_{3,1}x_3)
\end{aligned} \tag{6.31}$$

where the $w_{i,j}$ are elements of the 3 x 3 matrix of constants W:

$$\mathsf{W} = \begin{pmatrix} 1 & 1.2 & 0.8 \\ 0.8 & 1 & 1.2 \\ 1.2 & 0.8 & 1 \end{pmatrix} \tag{6.32}$$

The key feature of these equations is that the constant amount of inhibition from the other variables is asymmetric and this makes it possible for the system to cycle through states where one variable inhibits the others. The system of equations has eight equilibrium states; one at the origin, three where two variables are zero, three where one variable is zero and one where the three variables are equal. None of these fixed points is stable. The inhibition constants have been chosen to allow a heteroclinic trajectory that cycles past each of the saddle points at (1,0,0), (0,1,0) and (0,0,1). This heteroclinic trajectory is illustrated in Fig. 6.16 and the plots of the variables over time show them characteristically activating in sequence.

An example of the role of heteroclinic cycles is provided by the olfactory pathway of the locust. Sensors on the antennae connect to local circuits referred to as glomeruli in the antennal lobe. The glomeruli are interconnected by a class of neurons referred to as intrinsic neurons and the output generated by the interconnected glomeruli is carried to the mushroom body by projection neurons where they synapse on the Kenyon cells. An olfactory stimulus causes a subset of the

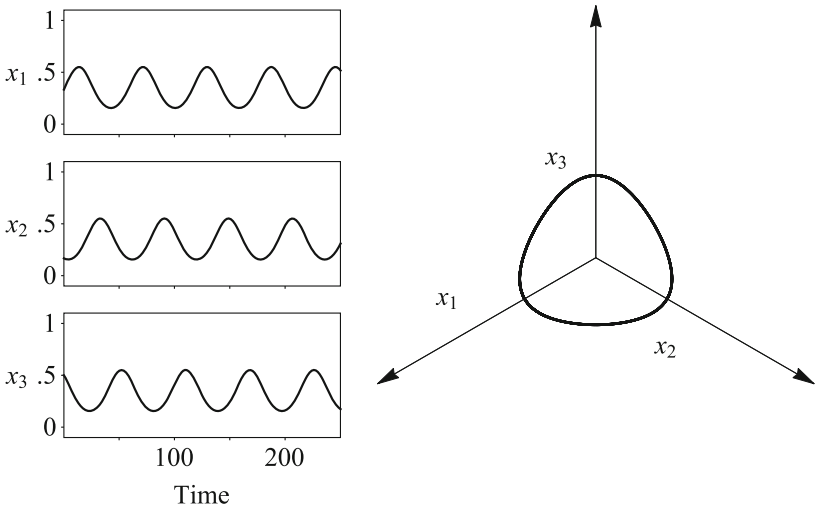


Fig. 6.16 Example of a heteroclinic cycle. In this and the subsequent two figures, both the time series of the system variables and a state space representation have been plotted. **Right Figure**: Illustration of how the trajectory of the system follows the stable direction towards and the unstable direction away from a succession of saddle points to form a closed loop. **Left Figure**: Associated with a heteroclinic cycle is a sequential activation of the variables, which is apparent in their plots against time

projection neurons to fire repeatedly in sequence, and this can be assumed to be a heteroclinic cycle. For a different olfactory stimulus there will be a different heteroclinic cycle so each heteroclinic cycle acts a signature of a particular odour [9].

The switching between different cycles with different inputs can be illustrated by adding an input mechanism that depends on the general level of activity to the Lotka-Volterrra system of Eqs. 6.31:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i(\rho(4(x_1 + x_2 + x_3) + s_i) - w_{i,1}x_1 - w_{i,2}x_2 - w_{i,3}x_3 \tag{6.33}$$

where S is a set of inputs to each of the three neurons and ρ is the nonlinear function:

$$\rho(x) = 1 - \frac{2}{1 + \exp(10(x - 0.4))} \tag{6.34}$$

In this case the numbers used in the matrix W were 1, 5 and 0.2. Figures 6.17 and 6.18 and illustrate how the repeating patterns of firing of the model neurons alter with a change in the input [13].

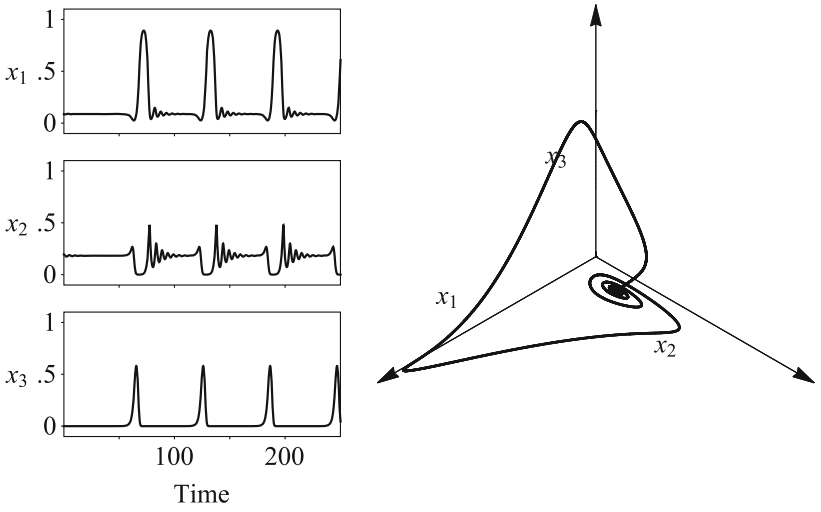


Fig. 6.17 Example of a heteroclinic cycle produced by the system of Eqs. 6.33 with the input $\mathsf{S} = (0.72, 0.089, 0.737)$. This figure should be viewed in conjunction with Fig. 6.18 where an entirely different heteroclinic cycle is produced when the input is changed, illustrating that different inputs can be characterised by different heteroclinic cycles

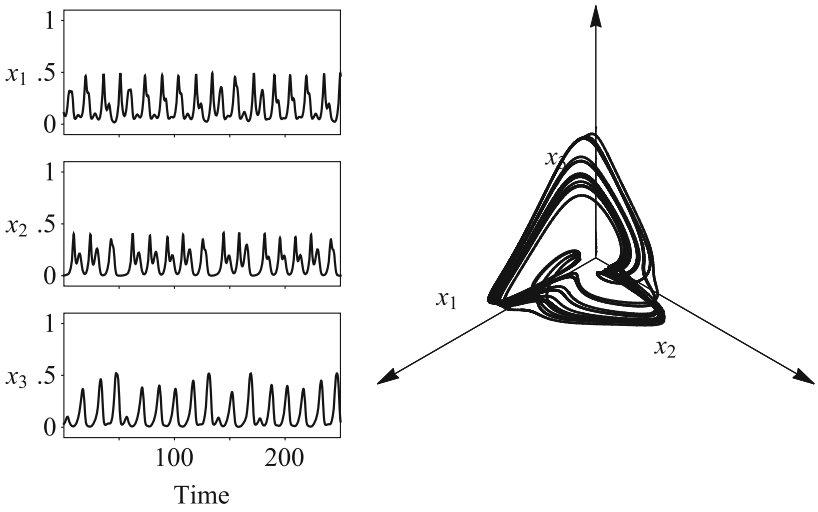


Fig. 6.18 Example of a heteroclinic cycle produced by the system of Eqs. 6.33 with the input $\mathsf{S} = (0.189, 0.037, 0.342)$

References

1. Akman, O.E., Broomhead, D.S., Abadi, R.V., Clement, R.A.: Eye movement instabilities and congenital nystagmus can be predicted by a nonlinear model of the saccade system. J. Math. Biol. **51**, 661–694 (2005). https://doi.org/10.1007/s00285-005-0336-4
2. Alonso, R.G., Amador, A., Mindlin, G.B.: An integrated model for motor control of song in *Serinus canaria*. J. Physiol. Paris **110**, 127–139 (2016). https://doi.org/10.1016/j.jphysparis.2016.12.003
3. Amador, A., Mindlin, G.B.: Synthetic birdsongs as a tool to induce, and listen to, replay activity in sleeping birds. Front. Neurosci. **15** (2021). https://doi.org/10.3389/fnins.2021.647978
4. Barreiro, A.K., Bronski, J.C., Anastasio, T.J.: Bifurcation theory explains waveform variability in a congenital eye movement disorder. J. Comput. Neurosci. **26**, 321–329 (2009). https://doi.org/10.1007/s10827-008-0113-7
5. Cannon, S.C., Robinson, D.A., Shamma, S.C.: A proposed neural network for the integrator of the oculomotor system. Biol. Cybern. **49**, 127–136 (1983). https://doi.org/10.1007/BF00320393
6. Goldstein, H.: Modelling post-saccadic drift: dynamic overshoot may be passive. In: Proceedings of the Northeast Bioengineering Conference, pp. 245–248. IEEE, Philadelphia (1987)
7. Joshua, M., Lisberger, A.: A tale of two species: neural integration in zebrafish and monkeys. Neuroscience **296**, 80–91 (2015). https://doi.org/10.1016/j.neuroscience.2014.04.048
8. May, R.M., Leonard, W.J.: Nonlinear aspects of competition between three species. Siam J. Appl. Math. **29**, 243–253 (1975). https://doi.org/10.1137/0129022
9. Mazor,O., Laurent, G.: Transient dynamics versus fixed points in odor representation by locust antennal lobe projection neurons. Neuron **48**, 661–673 (2005). https://doi.org/10.1016/j.neuron.2005.09.032
10. McMahon, T.A.: Muscles, Reflexes and Locomotion. Translated by B. Bridgeman. Princeton University Press, Princeton (1984).
11. Miri, A., Daie, K., Arrenberg, A.B., Baier, H., Aksay, E., Tank, D.W.: Spatial gradients and multidimensional dynamics in a neural integrator circuit. Nat. Neurosci. **14**, 1150–1159 (2011). https://doi.org/10.1038/nn.2888

12. Optican, L.M., Miles, F.A.: Visually induced adaptive changes in primate saccadic oculomotor control signals. J. Neurophysiol. **54**, 940–958 (1985). https://doi.org/10.1152/jn.1985.54.4.940
13. Rabinovich, M.I., Huerta, R., Volkovskii, A., Abarbanel, H.D.I., Stopfer, M.: Dynamical coding of sensory information with competitive networks. J. Physiol. Paris **94**, 465–471 (2000). https://doi.org/10.1016/s0928-4257(00)01092-5
14. Robinson, D.A.: The mechanics of human saccadic eye movement. J. Physiol. **174**, 245–264 (1964). https://doi.org/10.1113/jphysiol.1964.sp007485
15. Steinecke, I., Herzel, H.: Bifurcations in an asymmetric vocal-fold model. J. Acoust. Soc. Am. **97**, 1874–1884 (1985). https://doi.org/10.1121/1.412061

Chapter 7
Analysis of Experimental Measurements

How does a geometric approach help us investigate the brain? The technique of delay space embedding can be used to transform a sequence of measurements of a neural system's behaviour into a corresponding dynamical system. The system copy can be analysed to obtain the local dimensionality and eigenvalues at the fixed points of the system, just as if the equations for a model of the system were available. The eye movement disorder of nystagmus is used as an example of this approach. In addition to the local analysis, topological data analysis can be used to obtain a description of the overall structure of the manifold of the dynamical system.

7.1 Delay Embedding

In practice one is not usually given the equations describing the behaviour one is interested in and local linear analysis has to be extended to be applicable directly to the measurement data. The key to applying the linear approach directly to data is that measurements taken over time can be used to reconstruct the trajectory of the system through its state space and such reconstructed trajectories can be used to compute the properties of the local linear behaviour.

These techniques can be illustrated by an example system provided by the Euler's equations, which describe the behaviour of a rotating rigid body:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= x_2 x_3 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= -2x_1 x_3 \\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= x_1 x_2 \end{aligned} \tag{7.1}$$

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_7

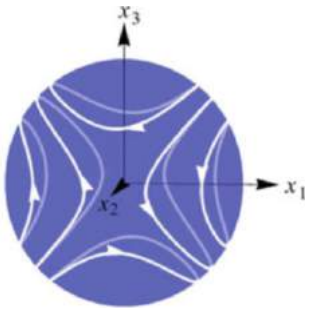


Fig. 7.1 Some example trajectories of the Euler system of Eqs. 7.1. The figure illustrates that the manifold defined by the collection of trajectories is a sphere

In this case the collection of possible trajectories of the system define a manifold consisting of a sphere embedded in three-dimensional Euclidean space as illustrated in Fig. 7.1.

The first question to investigate is the dimension of the manifold. There is no straightforward way of doing this directly but it can be done indirectly by calculating the dimension of the tangent spaces at a sample of points on the manifold. In the case of the example system specified by the system of Eqs. 7.1 the manifold is a surface for which the curvature becomes less and less apparent as one selects a neighbourhood closer and closer to the point. So if one calculates the principal components of the covariance matrix of the points in the neighbourhood, then the largest eigenvectors will lie in the tangent plane. By repeating the principal components calculation with a range of neighbourhood sizes one obtains a range of values for the eigenvalues, but the key point is that only the eigenvalues of the eigenvectors that lie in the tangent plane will scale directly with neighbourhood size [5]. Hence the number of such eigenvalues corresponds to the dimension of the tangent plane, in this case two. This calculation is illustrated in Fig. 7.2.

For the example system defined by Eqs. 7.1 there are six fixed points with coordinates $(0, 0, \pm 1)$, $(0, \pm 1, 0)$ and $(0, 0, \pm 1)$. Following this approach with the example system defined by Eq. 7.1, the Jacobian matrix is:

$$\mathsf{J} = \begin{pmatrix} 0 & x_3 & x_2 \\ -2x_3 & 0 & -2x_1 \\ x_2 & x_1 & 0 \end{pmatrix} \tag{7.2}$$

which has eigenvalues $\pm\mathrm{i}\sqrt{2}$ and 0, ± 1 and 0 and $\pm\mathrm{i}\sqrt{2}$ and 0 for the fixed points $(0, 0, \pm 1)$, $(0, \pm 1, 0)$ and $(0, 0, \pm 1)$ respectively.

The calculations of the dimensionality of the tangent space and the eigenvalues of the fixed points have all been done using solutions to the known Eqs. 7.1, but usually the system of equations is not known. The starting point for analysing measurements from an unknown dynamical system involves embedding the dynamical system in a Euclidean space, with the proviso that the embedding is a smooth map ϕ with a smooth inverse, in order to preserve the properties of the dynamical system. In

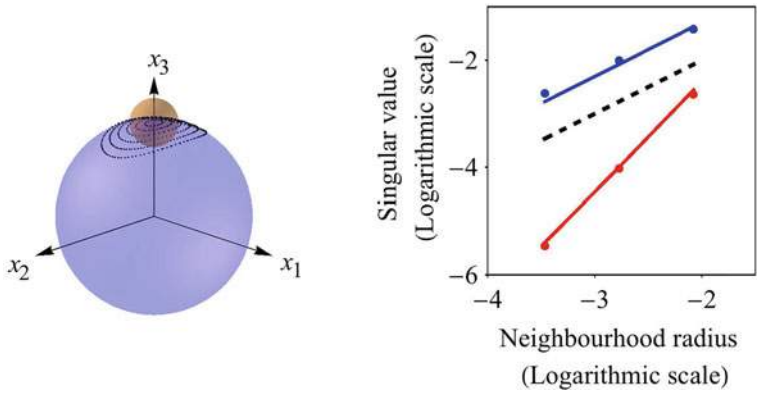


Fig. 7.2 Estimation of the local dimensionality of the example system specified by Eqs. 7.1. The figure on the left illustrates how points on the manifold in the neighbourhood of the point with coordinates (0,0,1) are selected by choosing those within a sphere with a radius r and centre at (0,0,1). The graph on the right illustrates two patterns of scaling of the singular values of the trajectory matrix, which is formed by writing the coordinates of each point in the neighbourhood into a row of the matrix. The singular values of the trajectory matrix correspond to the square roots of the eigenvalues of the trajectory covariance matrix. Two of the singular values, plotted in blue, are identical and scale directly with neighbourhood radius. These correspond to the eigenvectors lying in the tangent plane. One singular value, plotted in red, scales with the square of the radius. This correspond to the eigenvector in the direction (0,0,1). The results of the numerical procedures are plotted as points and these lie on the continuous lines which were calculated analytically [5]

particular, a **delay embedding** consists of a sequence of p measurements s(x) taken at successive time intervals τ:

$$\mathsf{Y} = \phi(\mathsf{X}) = (s(x(t)), s(x(t+\tau)), \ldots, s(x(t+(\mathrm{p}-1)\tau))) \tag{7.3}$$

In the context of delay embedding the p-dimensional Euclidean space in which the delay vectors lie is referred to as the **delay space**.

In the neighbourhood of the a fixed point X^* a linear approximation to the behaviour of the unknown system will hold, so one can describe the behaviour by a system of equations of the form:

$$\frac{\mathrm{d}(\mathsf{X}-\mathsf{X}^*)}{\mathrm{d}t} = \mathsf{J}_\mathrm{F}(\mathsf{X}-\mathsf{X}^*) \tag{7.4}$$

although the functions $\mathrm{F} = (f_1, f_2, \ldots, f_\mathrm{p})$ that define the behaviour of the system are not known. By using delay embedding the eigenvalues of the matrix J_F can be found from the measured behaviour of the system [4].

The copy of the dynamical system in delay space consists of a vector field G defined on a the embedded manifold $\phi(\mathrm{M})$ and the behaviour of the embedded system is described by the system of equations:

$$\begin{aligned}
\frac{\mathrm{d}y_1}{\mathrm{d}t} &= g_1(y_1, y_2, \ldots, y_\mathrm{p}) \\
\frac{\mathrm{d}y_2}{\mathrm{d}t} &= g_2(y_1, y_2, \ldots, y_\mathrm{p}) \\
&\cdots \\
\frac{\mathrm{d}y_\mathrm{p}}{\mathrm{d}t} &= g_\mathrm{p}(y_1, y_2, \ldots, y_\mathrm{p})
\end{aligned} \tag{7.5}$$

In the neighbourhood of a fixed point Y^* the equation describing the behaviour of the embedded system simplifies to the linear equation:

$$\mathsf{J}_\mathrm{G} = \begin{pmatrix}
\left.\frac{\partial g_1}{\partial y_1}\right|_{\mathsf{Y}^*} & \left.\frac{\partial g_1}{\partial y_2}\right|_{\mathsf{Y}^*} & \cdots & \left.\frac{\partial g_1}{\partial y_\mathrm{p}}\right|_{\mathsf{Y}^*} \\
\left.\frac{\partial g_2}{\partial y_1}\right|_{\mathsf{Y}^*} & \left.\frac{\partial g_2}{\partial y_2}\right|_{\mathsf{Y}^*} & \cdots & \left.\frac{\partial g_2}{\partial y_\mathrm{p}}\right|_{\mathsf{Y}^*} \\
\cdots & & & \\
\left.\frac{\partial g_\mathrm{p}}{\partial y_1}\right|_{\mathsf{Y}^*} & \left.\frac{\partial g_\mathrm{p}}{\partial y_2}\right|_{\mathsf{Y}^*} & \cdots & \left.\frac{\partial g_\mathrm{p}}{\partial y_\mathrm{p}}\right|_{\mathsf{Y}^*}
\end{pmatrix} \tag{7.6}$$

So in the neighbourhood of the fixed point the behaviour of the embedded system is described by the linear equation:

$$\frac{\mathrm{d}(\mathsf{Y} - \mathsf{Y}^*)}{\mathrm{d}t} = \mathsf{J}_\mathrm{G}(\mathsf{Y} - \mathsf{Y}^*) \tag{7.7}$$

Furthermore, from the definition of the embedding given in Eq. 7.3 it follows that:

$$(\mathsf{Y} - \mathsf{Y}^*) = \mathsf{J}_\phi(\mathsf{X} - \mathsf{X}^*) \tag{7.8}$$

where J_ϕ is the Jacobian of the delay embedding evaluated at X^*:

$$\mathsf{J}_\phi = \begin{pmatrix}
\left.\frac{\partial \phi_1}{\partial x_1}\right|_{\mathsf{X}^*} & \left.\frac{\partial \phi_1}{\partial x_2}\right|_{\mathsf{X}^*} & \cdots & \left.\frac{\partial \phi_1}{\partial x_\mathrm{p}}\right|_{\mathsf{X}^*} \\
\left.\frac{\partial \phi_2}{\partial x_2}\right|_{\mathsf{X}^*} & \left.\frac{\partial \phi_2}{\partial x_2}\right|_{\mathsf{X}^*} & \cdots & \left.\frac{\partial \phi_2}{\partial x_\mathrm{p}}\right|_{\mathsf{X}^*} \\
\cdots & & & \\
\left.\frac{\partial \phi_\mathrm{p}}{\partial x_1}\right|_{\mathsf{X}^*} & \left.\frac{\partial \phi_\mathrm{p}}{\partial x_2}\right|_{\mathsf{X}^*} & \cdots & \left.\frac{\partial \phi_\mathrm{p}}{\partial x_\mathrm{p}}\right|_{\mathsf{X}^*}
\end{pmatrix} \tag{7.9}$$

Applying the equivalence specified by Eq. 7.8 to 7.7 gives the equation:

$$\mathsf{J}_\phi \frac{\mathrm{d}(\mathsf{X} - \mathsf{X}^*)}{\mathrm{d}t} = \mathsf{J}_\mathrm{G}\mathsf{J}_\phi(\mathsf{X} - \mathsf{X}^*) \tag{7.10}$$

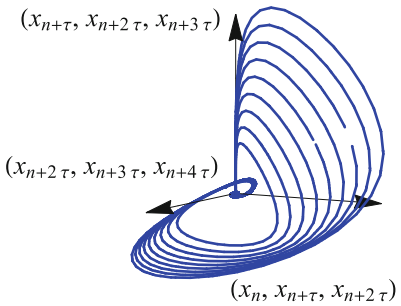


Fig. 7.3 Trajectories reconstructed from the x_1 variable of the saccade generation model specified by Eqs. 5.12. Saccades were simulated in sizes ranging from 2.5 to 25° in 2.5° steps. The sampling interval τ was 50 ms and triples of successive samples were used to create a three-dimensional delay space reconstruction of the state space trajectories. For comparison with the trajectories shown in Fig. 5.7 it should be noted that in the this figure the state moves clockwise round the trajectories

and joining this equation with Eq. 7.4 for the linear behaviour of the original system gives the equivalence:

$$\frac{\mathrm{d}(\mathsf{X} - \mathsf{X}^*)}{\mathrm{d}t} = \mathsf{J}_\mathrm{F}(\mathsf{X} - \mathsf{X}^*) = \mathsf{J}_\phi^{-1}\mathsf{J}_\mathrm{G}\mathsf{J}_\phi(\mathsf{X} - \mathsf{X}^*) \tag{7.11}$$

From which it follows that:

$$\mathsf{J}_\mathrm{F} = \mathsf{J}_\phi^{-1}\mathsf{J}_\mathrm{G}\mathsf{J}_\phi \tag{7.12}$$

The matrices J_F and J_G are similar matrices, which have the property that their eigenvalues are the same, so it is possible to recover the eigenvalues for the linear behaviour of an unknown system from experimental measurements.

For a neuroscience example of the application of this procedure consider the slow-fast model of saccadic eye movements. The pair of complex eigenvalues that characterise the spiral towards the origin of the trajectories can be calculated to be -13.8 ± 36.7 i with the parameters for human eye movements. The trajectories reconstructed from the x_1 variable are shown in Fig. 7.3, and the eigenvalues calculated numerically from these trajectories are -14.1 ± 35.6 i.

7.2 An Unstable Neural Pathway

Unstable movements are unfortunately not uncommon. An example of an eye movement disorder is **nystagmus**, which consists of the eye being unable to remain steady, but instead making repetitive movements away from and back towards the target gaze direction. In particular, one form of nystagmus, that starts during

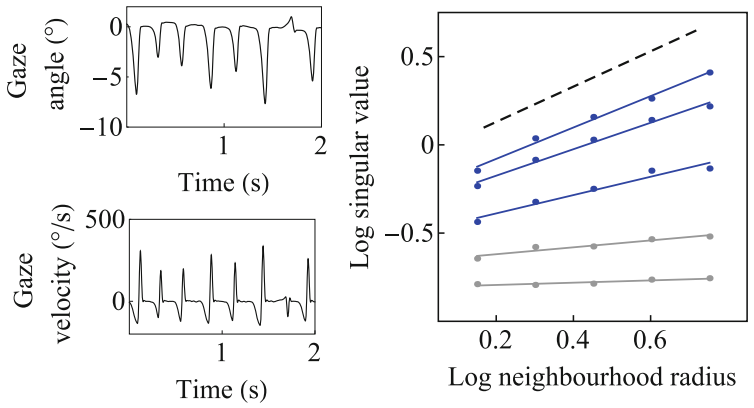


Fig. 7.4 Local dimensionality derived from experimental measurements of eye movements recorded from a person with nystagmus. The velocity data was embedded in five-dimensional Euclidean space using the method of delays. The slope of the plot of a singular value that scales linearly with neighbourhood radius is indicated by the dashed line and the three singular values that have slopes between 0.5 and 1.5 are plotted in blue. The scaling of three of the singular values indicates that the system is three-dimensional close to the origin, which corresponds to the target direction

infancy, involves both eyes moving together and predominantly horizontally, with the movement away having an increasing velocity. Although the nystagmus is approximately periodic over a few cycles it can vary considerably over time, especially with changes in attention.

The eye movement instability of nystagmus provides an example of where one can apply geometric analysis techniques directly to measurements without the guidance of a model. A typical nystagmus waveform can have either one of two equilibrium positions. An example of local linear analysis of a waveform with one equilibrium position is illustrated in Fig. 7.4. In this waveform the eye repeatedly drifts with increasing velocity away from the fixation target and is returned towards the target by a saccade.

The velocity of the waveform, which reflects the neural control signal, was embedded in five-dimensional Euclidean space and it was found that three singular values scale with the radius of the neighbourhood centred on the origin. The three eigenvalues associated with this three-dimensional system were calculated from the Jacobian matrix, calculated at the origin, of a three-dimensional trajectory reconstruction. Two of the eigenvalues formed a complex pair that describe the stable spiral in towards the origin at the end of the saccades, and the third eigenvalue characterises the unstable drift of the eye away from the fixation target. In this case the eigenvalue corresponds to a time constant for the drift of 30 ms which is similar to the values found in other recordings [1, 2].

This result provides a test for models of nystagmus in that whilst many models, both animal and computational, produce waveforms that look like nystagmus, none of them have been shown to have fixed points with the eigenvalues found

in the experimental data. Given that learning how to use and understand such models requires a considerable investment of time, local linear analysis provides a simple method for screening out models which do not have the behaviour found experimentally.

7.3 Topological Data Analysis

Clearly, analysis of measurements made close to a fixed point only reveals a limited amount about the behaviour of a neural system. What can one do with the measurements made away from the fixed point? One promising approach involves moving away from looking at continuously varying properties such as eigenvalues to investigating discrete properties. One can still look for topological structure in the data, in particular the presence of holes in the geometric structure of the data. This can be done by finding closed circuits in the data space that enclose the holes.

The first step in looking for closed circuits involves specifying how the data points can be connected up. The portion of space enclosed by a collection of $k+1$ points is referred to as a **simplex** $\sigma[p_1, p_2, \ldots, p_{k+1}]$. The low number simplexes are illustrated in Fig. 7.5. The 0-, and 1- simplexes often referred to by name, which are vertex and edge respectively. Each k-simplex is made up of a collection of $(k-1)$- simplexes and these components are referred to as faces. Simplexes can be organised into a **simplicial complex**. These are collections of simplexes in which each simplex has a face in common with another. The degree of the complex is given by the simplex with the highest degree.

An orientation can be assigned to a simplex by numbering each vertex in turn. Assigning an orientation to a simplex allows one to keep track of the direction of a path through a simplex. The direction of the path can be indicated by the ordering of the the vertices so:

$$[1, 2] = -[2, 1] \tag{7.13}$$

The boundary of a k-simplex can be specified by adding all the $k-1$ simplexes that make up the larger simplex. These can be found by removing each vertex of the simplex in turn as:

$$[1, 2, 3] = [1, 2] + [2, 3] - [1, 3] \tag{7.14}$$

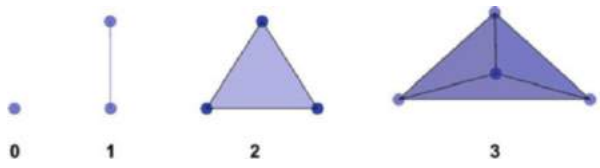


Fig. 7.5 Illustrations of the basic simplexes

The map from a k-simplex to a $(k-1)$-simplex is referred to as the **boundary matrix** and is denoted by ∂_{k}

$$\partial_k([1,2,\ldots,k+1]]) = \sum_{i=0}^{k}(-1)^i[1,2,\ldots x_i\ldots,k+1] \tag{7.15}$$

where $\mathrm{x_i}$ is the vertex removed.

A path through a simplicial complex can be defined as consisting of a collection of linked k-dimensional simplices $(\sigma_1, \sigma_2, \ldots \sigma_{\mathrm{n}})$ selected from the complex. Such collections are referred to as a **chain** and are denoted by $\mathbf{c}_{\mathrm{k}}$. Each chain can by described by a vector:

$$\mathbf{c}_{\mathrm{k}} = a_1\sigma_1 + a_2\sigma_2 + \ldots + a_{\mathrm{n}}\sigma_{\mathrm{n}} \tag{7.16}$$

where the values of the coefficients $a_1, a_2, \ldots, a_{\mathrm{n}}$ are 1 or 0. The collection of these vectors forms a vector space $\mathbf{C}_k$.

A chain complex C is a sequence of vector spaces $\mathrm{C_k}$ in which successive spaces are related by the boundary matrix $\mathbf{C}_k - \partial_{\mathrm{k}} \rightarrow \mathbf{C}_{k-1}$

$$\mathbf{0} - \partial_{k+1} \rightarrow \mathbf{C}_k - \partial_k \rightarrow \mathbf{C}_{k-1} - \cdots \rightarrow \mathbf{C}_1 - \partial_1 \rightarrow \mathbf{C}_0 - \partial_0 \rightarrow \mathbf{0} \tag{7.17}$$

So now one can look for paths that end up back where they start from. Such a path should enclose a part of the simplicial complex and if it does not then there is a hole in the complex. The type of hole will depend the dimension of the simplexes in the path in that a loop of k-dimensional simplexes will enclose a k-dimensional simplex.

The kernel of the boundary matrix ∂_{k}, which is comprised of all vectors transformed to the zero vector by the matrix, consists of k-cycles, and the image of the boundary matrix $\partial_{\mathrm{k+1}}$, which is comprised of all vectors not transformed to the zero vector by the matrix, consists of boundaries. The **Betti number** is defined by:

$$\beta_k = \text{dimension of the kernel of } \partial_k - \text{dimension of the image of } \partial_{k+1} \tag{7.18}$$

The zeroth Betti number is the number of connected spaces. The first Betti number is the number of two-dimensional holes, the second Betti number is the number of three-dimensional holes and so on.

For example, a simplicial complex consisting of a tetrahedron is topologically equivalent to the sphere. The collections of 0, 1 and 2 simplexes that make up the tetrahedron form vector spaces $\mathrm{C_0}$, $\mathrm{C_1}$, and $\mathrm{C_2}$. These vector spaces have 4, 6 and 4 basis vectors as illustrated in Fig. 7.6, and the boundary matrices describing the transformations between successive vector spaces are given in Tables 7.1 and 7.2.

The dimension of the kernel of ∂_2 is 1 and the dimension of the image of ∂_3 is 0 so β_2 equals 1. The dimension of the kernel of ∂_1 is 3 and the dimension of the

Table 7.1 Boundary map ∂_2 from triangles to edges

	[1,2,3]	[1,2,4]	[1,3,4]	[2,3,4]
[1,2]	1	1	0	0
[1,3]	−1	0	1	0
[1,4]	0	−1	−1	0
[2,3]	1	0	0	1
[2,4]	0	1	0	−1
[3,4]	0	0	1	1

Table 7.2 Boundary map ∂_1 from edges to vertices

	[1,2]	[1,3]	[1,4]	[2,3]	[2,4]	[3,4]
[1]	−1	−1	−1	0	0	0
[2]	1	0	0	−1	−1	0
[3]	0	1	0	1	0	−1
[4]	0	0	1	0	1	1

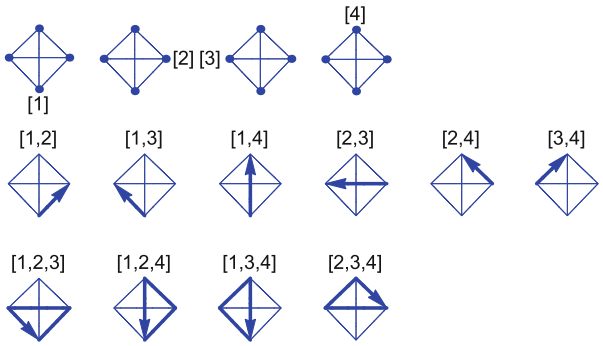


Fig. 7.6 Example of basis vectors associated with the vector spaces associated with chains

image of ∂_2 is 3 so β_1 equals 0. Finally, the dimension of the kernel of ∂_0 is 4 and the dimension of the image of ∂_1 is 3 so β_1 equals 1. In effect, Betti number zero tells you that there is only 1 simplicial complex, Betti number 1 tells you there are no holes bounded by edges and Betti number 2 tells you there is one hole bounded by triangular faces, corresponding to the inside of the tetrahedron.

Simultaneous recordings from many electrodes has led to an abundance of high-dimensional data sets in neuroscience and so a technique for revealing the geometric structure of such data sets will prove very useful. For a simple example of how topological data analysis can be applied to high-dimensional data take the six-dimensional data set of length changes of the extraocular muscles illustrated in Fig. 3.14 which was used in the dimensionality reduction example of Chap. 3. Such a data set can be turned into a graph by placing an edge between any point that lies within a radius r of another point. With an appropriate set of choice of radius a graph is formed that has a zeroth Betti of 1 and the first and second Betti numbers are zero as illustrated in Fig. 7.7. This result signifies that the arrangement of data points is topologically equivalent to points lying on a portion of a sphere. But the form of the

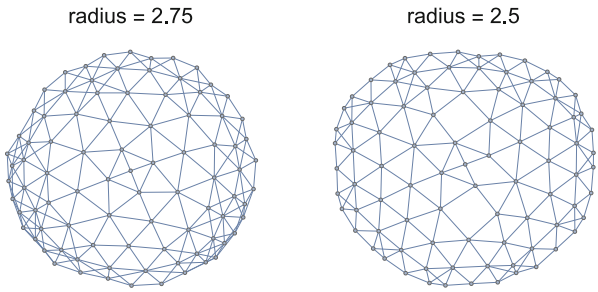


Fig. 7.7 Two examples of a graph derived from a set of data points. The data points were the length changes of the six extraocular muscles used to illustrate the use of principal components analysis in dimensionality reduction, and the layout of the data points follows the form of the associated visual directions which lie within 25° of the straight ahead direction. With a threshold radius of 2.75 for connecting the data points a graph is obtained for which the first three Betti numbers are 1,0 and 0 respectively. But with a smaller threshold radius of 2.5 for connecting the data points a graph is obtained for which the first three Betti numbers are 1,2 and 0 respectively. The two holes associated with the second Betti number can be seen above and below the centre of the graph on the right

graph depends on the choice of radius and with a different radius different values for the Betti numbers are obtained. So reliable results from the technique require its repeated use with a range of radii, so that consistent sets of Betti numbers can be found.

A less direct method of applying topological data analysis has proved pertinent to neuroscience applications. The odours produced by animals involve numerous chemical reactions so that recognised by the pattern of molecules present. But the relative levels of the molecules fluctuate with environmental conditions and so the correlational structure recovered by principal components analysis is unreliable. However, one can still gain insight into the manifold in which the data points lie by forming the correlation matrix between the measured levels of the molecules and then setting a threshold at which a link is created. So one can form a simplicial complex with the vertices corresponding to the different molecules. One can then compute the first few Betti numbers of the complex and the relationship between the number of edges and the Betti numbers as the threshold varies gives an indication of the form of the manifold in which the data lies.

In particular, with data from a Euclidean space the curves of Betti number as a function of number of edges have lower and lower peaks with successive Betti numbers, whereas with data from a hyperbolic space the curves all peak at approximately the same level. In the case of odor molecules, this type of analysis indicates that the data lies in a hyperbolic space. In such a space the distance between points matches that of a tree structure in that the shortest path between points often involves backing up a branch to get to a point on a neighbouring branch and such a tree structure represents an arrangement of the odours corresponding to an optimal classification of them [3, 6].

References

1. Abadi, R.V., Broomhead, D.S., Clement, R.A., Whittle, J., Worfolk, R.: Dynamical systems analysis: a new method of analysing congenital nystagmus waveforms. Exp. Brain Res. **117**, 355–361(1997). https://doi.org/10.1007/s002210050229
2. Akman, O.E., Broomhead, D.S., Abadi, R.V., Clement, R.A.: Components of the neural signal underlying nystagmus. Exp. Brain Res. **220**, 213–221 (2012). https://doi.org/10.1007/s00221-012-3130-8
3. Giusti, C., Pastalkova, E., Curto, C., Itskov, V.: Clique topology reveals intrinsic geometric structure in neural correlations. PNAS **112**, 13455–13460 (2015). https://doi.org/10.1073/pnas.1506407112
4. Healey, J.J., Broomhead, D.S., Cliffe, K.A., Jones, R., Mullin, T.: The origins of chaos in a modified Van der Pol oscillator. Physica D **49**, 322–339(1991). https://doi.org/10.1016/0167-2789(91)90091-M
5. Kirby, M.: Geometric Data Analysis. John Wiley and Sons, Inc., New York (2001). ISBN: 978-0-471-23929-1
6. Zhou, Y., Smith, B.H., Sharpee, T.O.: Hyperbolic geometry of the olfactory space. Sci. Adv. **4**, eaaq1458 (2018). https://doi.org/10.1126/sciadv.aaq1458

Chapter 8
Where Are We Going With All This?

What are the implications of using geometric techniques for understanding the brain? After assessing the current limitations, the relationship with complementary techniques are outlined and it is argued that neuroscience is in need of a mathematics more suited to the behaviour of neurons as opposed to mathematics suited to our thought processes.

8.1 Review of the Geometric Approach

The starting point of this book was the argument that the brain is adept at handling objects which have few degrees of freedom and predictable behaviour. Of especial importance are biological objects, whose movement behaviour reflects the constraints of a skeleton. Examples of such biological objects were provided by the eye and arm. Underlying movements of the eye is a ball and socket joint and underlying movements of the arm in the horizontal plane is a hinge joint. The configuration space of the ball and socket movement was found to consist of a projective plane and that of the hinge movement was a Euclidean plane.

It was further argued that neurons are adept at computing properties in Euclidean space and that a first step in dealing with biological objects involves embedding them in Euclidean space. Both the Veronese embedding of the projective plane and the Euclidean space representation of a plane can be distinguished from the majority of embeddings by the property that the region surrounding every point the surface maintains, not just the topological properties, but the same angles and distances as a local Euclidean space. Manifolds with this property are referred to as **flat manifolds** [7]. It may be that such flat manifolds make computation of the properties of the objects simpler for the brain, by locally preserving the metric properties of Euclidean space. Certainly, other embeddings such as that of the

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_8

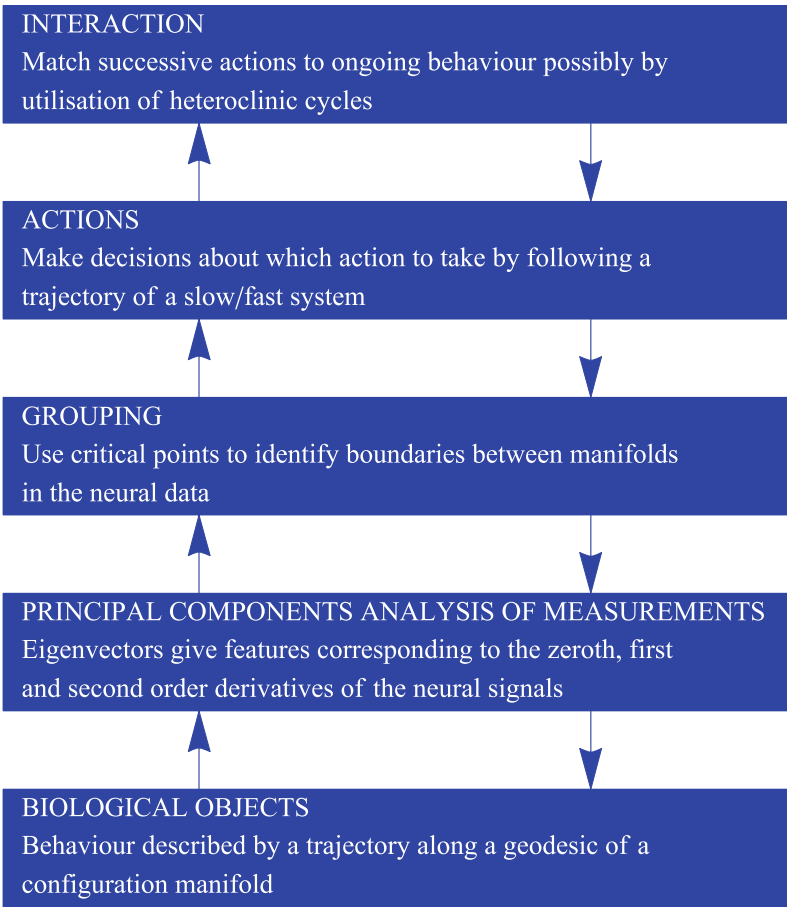


Fig. 8.1 Summary diagram of the geometric approach argued for in this book

righting movement of the cat or of head movements, stay close to the flat manifold requirement.

A schematic outline of the geometric approach is shown in Fig. 8.1 and to be realistic, the examples presented in this book are concerned with data taken close to the muscles on the motor side and the eye on the sensory side and it remains to be demonstrated that the geometric approach can be extended to more complex data structures, such as facial expressions, or to a broader range of data structures, such as those involved in speech and language.

But maybe one cannot tackle problems such as recognition, speech and language without knowing the limitations of the representations that the brain uses for its low-level processing. For instance, one approach to making pictures of the projective plane is to plot the projection of a higher–dimensional embedding onto the Euclidean plane. Two projections of a five-dimensional embedding of the

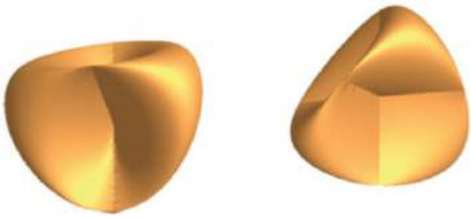


Fig. 8.2 Two projections of an embedding of the real projective plane into three-dimensional Euclidean space. The projection on the left is known as the cross-cap and that on the right is referred to as the Steiner surface

Fig. 8.3 Plastic bag as a cat lying beneath a car

projective plane are shown in Fig. 8.2. The drawback of this approach is that because of the way the visual system works, exploiting the fact than the outline corresponds to a fold of the surface when the projection is from three-dimensional Euclidean space to two, the projections are always seen as two different three-dimensional objects.

At the very least, the approach should allow one to understand, in the sense of what happens next, the interactions of other brains with the world. For instance, a human can look at the scene shown in Fig. 8.3 and automatically register that one is looking at a plastic bag stuck under a car. But my dog does not seem to be able to do this and acts as though there is cat there, especially if the wind moves the bag.

If the sequence of processes illustrated in the summary Fig. 8.1 does reflect the way the brain operates, then it is natural to question why it works like this. It may simply be that processes required for this way of working match those available for individual neurons so that primitive brains set the ground plan for future developments. The model neuron illustrated in Fig. 1.6 is ideal for carrying out linear filtering because it calculates the Euclidean inner product. But more

than that, because neurons receive many synaptic inputs they can work with high-dimensional Euclidean spaces. Similarly, synaptic plasticity provides a basis for Hebbian learning and a range of depolarising mechanisms enable slow-fast behaviour.

For a single nerve cell the equation describing the depolarising mechanism has the form:

$$\begin{aligned} \mathrm{C}\frac{\mathrm{d}v}{\mathrm{d}t} &= f(v, r_1, r_2, \ldots, r\mathrm{n}) \\ \frac{\mathrm{d}r_s}{\mathrm{d}t} &= g_s(v, r_s) \end{aligned} \tag{8.1}$$

where v is the membrane potential and C is its capacitance. The remaining s variables $r_1, r_2, \ldots, r_\mathrm{S}$ are variables that describe the rates at which ion channels open and close. Because of the time course of the rates, a simplified model of the spiking neuron can be made using only 1 rate variable for the combined effects of the sodium channel closing and potassium channel opening, and the membrane voltage equation of this model defines a cubic slow manifold curve, as illustrated in Fig. 8.4 [19].

A possible exploitation of the slow-fast behaviour of the nerve cell may occur in the fish escape circuit. Given a sudden mechanical stimulus a goldfish makes a c-start response in which the front and back halves of its body bend away from the stimulus into a c shape, with its head pointing away from the stimulus. Subsequently the tail is pushed back to propel the animal away from the stimulus. The response is controlled by a pair of **Mauthner cells** which are excited by the VIIIth nerve and inhibited by the commisural passive hyperpolarising potential neurons, in both cases by a combination of fast electrical and slower chemical synapses. The electrical inhibition is reciprocal so that firing of the Mauthner cells also hyperpolarises the hyperpolarising potential neurons, hence their description. Mauthner cells also

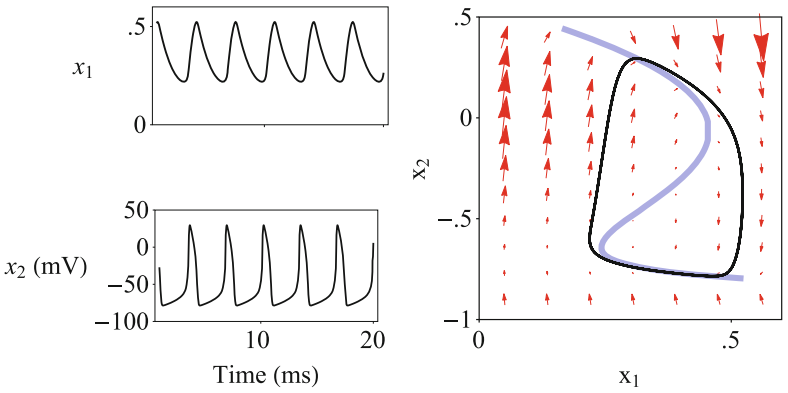


Fig. 8.4 Behaviour of a simplified version of the Hodgkin-Huxley model of the nerve impulse consisting of a membrane voltage x_2 equation and an ion channel x_1 rate equation. The model illustrates that slow-fast behaviour is inherent in the mechanism of excitable nerve cells

excite spinal motorneurons and cranial relay neurons which in turn excite another group of hyperpolarising potential neurons that electrically inhibit the Mauthner cells [16]. The Mauthner cells have the very short membrane time constant required for them to act as the fast variable, and the chemical and electrical synapses could supply the pair of inhibitory inputs required for the behaviour of the slow variables, in Eq. 5.7, although this proposal is currently no more than speculation.

The control of an action provides an example of how successive elaborations of a ground plan might be effected. The action could be controlled by a single neuron, if the slow-fast description of the Mauthner cell is correct, or it could be controlled by a circuit of neurons such as that involving burst and pause cells in the brainstem saccadic circuitry or the behaviour of the basic circuit could be subsumed into that of interacting regions such as the basal ganglia and the motor cortex.

Birdsong provides another elaboration of an action circuit. The songs consist of very brief vocalisations, several of which occur in sequence to form a syllable lasting up to several hundred milliseconds. The syllables are in turn joined in sequence to form short segments referred to as phrases or motifs lasting around a second. The motor neurons of the muscles in the syrinx and surrounding area and of the respiratory muscles are both activated by signals from the robust nucleus of the archistriatum (RA for short). Individual RA neurons produce a characteristic sequence of bursts for a given phrase. The RA is itself driven by neurons in the High Vocal Centre (HVC), each of which produce a single burst at a specific moment of a particular phrase. One possibility is that the RA and HVC correspond to the motor cortex and premotor frontal cortex responsible for voluntary movements in humans [2].

8.2 Alternative Approaches

The brain can be understood, in the sense of knowing what will happen if you make a change to it, at many levels ranging from molecular concentrations at the cellular level through to thought processes at the cognitive level and it is reasonable to ask if there is a useful role for the geometric approach. The choice of level of explanation often depends upon what you want from the investigation. If you need to know how a drug affects synaptic transmission then you have to find out what is happening at the synapses, whereas if you want to investigate object recognition then you need to measure what is happening with millions of neurons. But the behaviour of the brain, in terms of whether or not it responds to a change and how strongly, must be the same at all levels, given the assumption that mental activity is entirely down to brain activity.

The advantage of the behavioural explanation is that it is not tied to a particular chemical, cellular or anatomical mechanism. Given that corresponding parts of the brain vary from species to species, another advantage is that the behavioural explanation is not tied to a particular species. The differences between species are perplexing. For example in the cat retina a distinction can be made between X cells,

in which the different parts of the receptive field summate linearly, and Y cells in which the different areas do not summate linearly. These X and Y cells have been identified with the anatomically distinct beta and alpha cells respectively. In the primate, a distinction can be made between P (parvocellular) and M (magnocellular) cells, which have been identified with midget and parasol cells respectively. But whereas the X and Y cells send axonal terminations to both the main layers of the cat lateral geniculate nucleus, the P and M cells terminate in separate layers of the monkey lateral geniculate nucleus [14]. This is part of a wider problem in that the number of layers of the nucleus varies from three in the rat to six in the human and whereas the sequence of ipsilateral and contralateral layers is contra, ipsi, contra, ipsi, ipsi and contra in humans, the corresponding sequence in tree shrews is ipsi, contra, contra, contra, ipsi and contra [4]. As far as I am aware, we have no cogent explanation for these differences, but an analysis of the behaviour of these neurons can be used to capture the common functionality of these markedly different circuits.

But there are many insightful approaches to understanding behaviour. In the context of evolution, it is natural to think that animal behaviour results from optimising some characteristic in order to give a species evolutionary advantage. Unfortunately, specifying exactly what is being optimised in all the broad expanse of features is not easy. For example, take the superior rectus muscle of the swordfish eye which has a large swelling with a different composition from that of the other extraocular muscles. It turns out that this part of the muscle acts as a heater for the brain of the animal [1]. There is no straightforward method for calculating how the loss of strength of the muscle is balanced by an increased temperature of the brain in a way that is optimal for the swordfish.

But a useful consequence of the evolutionary approach is that it naturally pushes one into to looking at many different species in order to check that a particular property is indeed being optimised. Exploring the behaviour of different species has proved to be an effective strategy for investigating neural mechanisms, in particular when it is possible to experimentally manipulate a neural circuit with relatively few neurons such as that of the lobster stomatogastric ganglion and to obtain results that are easier to understand than those from circuits with the enormous number of neurons usually encountered [12]. The drawback of using a smaller animal is that it is harder to measure accurately the physical properties of the sensory and motor components to which the circuit is attached and this makes it that much harder to relate the neural signals to the behaviour of the animal [6].

This is an important point because it turns out the processing that the brain has to carry out is simplified by judicious construction of the sensory and motor components external to the brain. In the case of the visual system, the initial processing is directed towards isolating the outlines of individual biological objects in preparation for interaction with them. The projection from everyday three-dimensional space to the two-dimensional retinal image, which can be taken to be a loss of data that has to be corrected for, makes identification of the outline easier. This feature of the projection is illustrated in Fig. 8.5. The figure shows the projection of a sphere covered with a uniform distribution of dots. Because of the projection, the density of dots in the image is much higher at the edges of the image

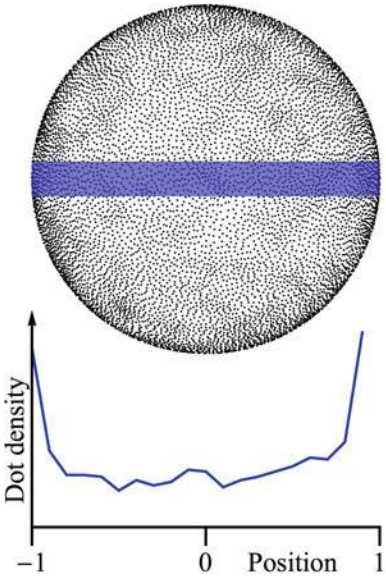


Fig. 8.5 Illustration of how the projection in the eye from three dimensions to two makes the task of recovering the outlines of objects easier. The example object is a sphere covered with a uniform distribution of dots. In the two-dimensional projection the dot density increases markedly at the edges. This argument is portrayed graphically in the lower portion of the figure which shows the dot density as a function of position, for the dots within the blue horizontal band in the image. Dots where chosen for this picture, but the argument holds with patches of colour or texture such as might be found on an object covered allover in the same material

as illustrated by the histogram of dots in the horizontal band across the image, so that half the work in identifying the outline is already done. Figure 8.5 also indicates that there are many viable alternative functions to local contrast that can be used in the analysis of the retinal image. One alternative is the density of values as a function of height [13]. Similar changes occur in the derivative a function defined by the area of the surface at a given height [9].

On the motor side, the anatomical arrangement of the muscles ensures that the changes in the axis of rotation of each muscle that occur when the initial position of the eye is away from the straight ahead direction are aligned with the changes specified by the half-angle rule of Listing's law. In part this is due to the broad insertions of the muscles into the eye and in part due to the orbital tissue attached to the muscle. To take the superior rectus as an example, when the eye is turned nasally, the fibres of the muscles furthest away from the nose are stretched the most. Hence the effective point of insertion of the muscle moves away from the nose. The opposite effect occurs with movements away from the nose [8]. Furthermore, the check ligaments in the orbit act like pulleys which ensure that also help to ensure that the changes in the axis of rotation are in accordance with Listing's law [10].

Failure to take into account the role of sensory and motor components external to the brain can be avoided by following the unquestionably logical approach of investigating the transformations of energy as it is processed by the sense organs and used to generate movements, given that these transducers have to obey the laws of physics like any other object. For neuroscience applications, the language of control theory has proved especially useful for understanding how a brain interacts with its world. Following the control theory approach, the description of the transformation of a neural signal to a physical movement is referred to as the forward model and the description of the transformation of the movement made into the corresponding neural signal is referred to as the inverse model. For example, Eq. 6.6 which describes how tension developed in the extraocular muscles moves the eye is a forward model and the combination of the burst, step and slide signals is an inverse model. For a given eye movement, applying the inverse model gives the required neural signal. This approach still leaves open the question of how the movement should be made, and a solution to this problem is to apply a function, referred to as the cost function, that specifies how good or bad the movement is for the animal.

In the control theory framework, Eqs. 6.24 which were introduced in the Chap. 6 to describe how a system changes from one set to the next, are enhanced by including functions $g_1, g_2, \cdots, g_\mathrm{m}$ of additional control variables $c_1, c_2, \ldots, c_\mathrm{m}$ that can be used to alter the behaviour of the system:

$$\begin{aligned}
\frac{\mathrm{d}_1}{\mathrm{d}t} &= f_1(x_1, x_2, \ldots, x_\mathrm{n}, g_1(c_1, c_2, \ldots, c_\mathrm{m}), \ldots, g_\mathrm{m}(c_1, c_2, \ldots, c_\mathrm{m})) \\
\frac{\mathrm{d}_2}{\mathrm{d}t} &= f_2(x_1, x_2, \ldots, x_\mathrm{n}, g_1(c_1, c_2, \ldots, c_\mathrm{m}), \ldots, g_\mathrm{m}(c_1, c_2, \ldots, c_\mathrm{m})) \\
&\ldots \\
\frac{\mathrm{d}_\mathrm{n}}{\mathrm{d}t} &= f_\mathrm{n}(x_1, x_2, \ldots, x_\mathrm{n}, g_1(c_1, c_2, \ldots, c_\mathrm{m}), \ldots, g_\mathrm{m}(c_1, c_2, \ldots, c_\mathrm{m}))
\end{aligned} \tag{8.2}$$

But Eqs. 8.2 for a controlled dynamical system can also be interpreted geometrically. For example, for the falling cat equation 2.9 can be considered to have two control variables ψ and θ which specify the shape of the cat. This specification corresponds to a point in the projective plane and the entire righting movement forms a closed loop in the space of controlled variables, as illustrated in Fig. 2.10. Alternative ways of making the righting movement are all associated with different loops, some examples of which were shown in Fig. 2.10. The required righting movement is given by the shortest of the possible closed loops, so now the problem of choosing the right way to make a movement becomes one of defining a metric on the space of control variables.

In the case of the falling cat the space of control variables can be given a metric derived from the kinetic energy associated with the shape changes. This metric takes its simplest form with the most symmetrical form of the cat, which consists of two spheres so that the transverse and axial moments of inertia are identical. This version of the metric is identical to that of Listing's law and similarly, the paths following geodesic curves in the projective plane consist of direction circles. So rather than optimising some cost function, a strategy that will guarantee that the appropriate

movement is made consists of finding the shortest path in the most symmetrical version of the surface, then translating the solution back to the original problem [11].

But also important for understanding the brain is the fact that it does not behave entirely deterministically. Instead of having access to the coordinates $(x_1, x_2, \ldots, x_n)$ of a point X on a manifold the brain maybe only has access to a representation of the coordinates $\mathsf{Y} = (x_1 + \nu_1, x_2 + \nu_2, \ldots, x_n + \nu_n)$ of a point X where the coordinates are contaminated by the addition of noise $(\nu_1, \nu_2, \ldots, \nu_n)$. One approach to overcoming this misinformation is to estimate the probability of the coordinates x by applying **Bayes's theorem**:

$$\mathrm{P}(\mathsf{X}|\mathsf{Y}) = \frac{\mathrm{P}(\mathsf{Y}|\mathsf{X})\mathrm{P}(\mathsf{X})}{\mathrm{P}(\mathsf{Y})} \tag{8.3}$$

where $\mathrm{P}(\mathsf{X}|\mathsf{Y})$is the posterior probability of the coordinates of the point being X given the measured coordinates Y, $\mathrm{P}(\mathsf{Y}|\mathsf{X})$ is the likelihood of measuring the coordinates of Y given a set of coordinates x, X is the prior probability of the point X and $\mathsf{Y}$is the probability of measuring the coordinates of Y. Different senses give different estimates of the shapes of surfaces and these are combined in the optimal manner prescribed by Bayes's theorem, although this finding gives no insight into the neural mechanism and how we should expect the combination to change when the brain is damaged.

The Bayesian explanation of behaviour can only work if the brain is able to specify the required probabilities correctly. But this may not require them to be made explicit. Rather, the probabilities may be implicit in the structure of the brain. To take an example from the visual system, estimation of the probability that two edges are part of the same outline can be built into the connections of the visual cortex [15]. A geometric approach to the problem of calculating the probability of co-occurrence of edges begins with the description of the configuration space of the possible edge orientations given in Chap. 4. Ordinarily, the curvature value κ at each point is determined by the contour in the image and the cortical representation is given by the solution of the system of differential equations:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{d}t} &= \cos(\theta) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} &= \sin(\theta) \\ \frac{\mathrm{d}x_3}{\mathrm{d}t} &= \kappa \end{aligned} \tag{8.4}$$

All the variables in this equation are functions of a parameter t, corresponding to the distance moved along the contour, so $\mathrm{x_i}$, θ and κ should be understood as $\mathrm{x_i(t)}$, $\theta(\mathrm{t})$ and $\kappa(\mathrm{t})$. If instead, it is assumed that the probability of the orientation at successive points along a retinal contour are distributed at random, then cortical contours can

also be generated by solving the system of stochastic differential equations:

$$\begin{aligned} \frac{dx_1}{dt} &= \cos(\theta) \\ \frac{dx_2}{dt} &= \sin(\theta) \\ \frac{dx_3}{dt} &= \mathrm{N}(0, \sigma) \end{aligned} \tag{8.5}$$

In this system of equations it has been assumed that successive orientations are distributed according to a normal distribution $\mathrm{N}(0, \sigma)$ with a mean of zero and variance of σ^2. On can go on to compute many solutions to this equation, always starting from the same horizontal retinal orientation. One can then project these three-dimensional trajectories back down onto the retina to find out what this assumption implies about the retinal contours. The findings are simplest to understand if one makes a histogram of how many cortical trajectories pass through each small portion of the retina and such a histogram is illustrated in Fig. 8.6, where the number of trajectories passing through each region is indicated by the intensity of the blue colour. The histogram has the bow-tie shape of the association fields of the visual cortex, so it may that rather than having a population of neurons

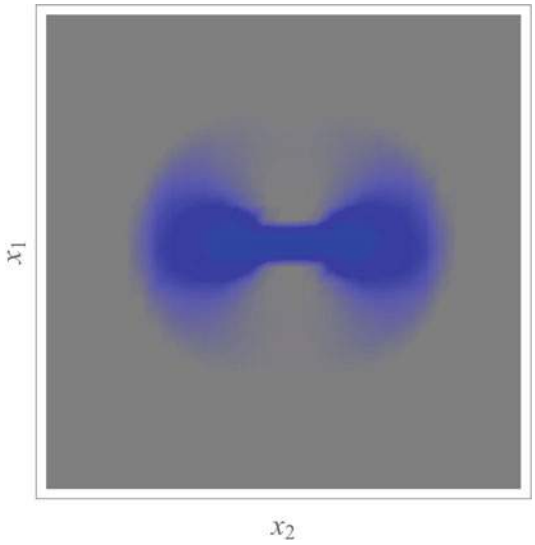


Fig. 8.6 Illustration of the histogram produced by taking many solutions to the system of stochastic differential equations 8.5 and then counting how many trajectories pass through each region of the retina. This number is represented by the intensity of the blue colour at each point. It can be seen that this histogram has the characteristic bow-tie shape also shown by the strength of the connections between orientation columns in the visual cortex. So these lateral connections may embody the same estimate of the probability distribution for the co-occurrence of orientations that is assumed in the stochastic equations

whose firing explicitly represents the probability distribution of edge orientations, the probability distribution may be represented implicitly by the distribution of horizontal connections between the orientation columns in the cortex.

For many tasks, such as facial recognition or language comprehension, there is no obvious extension of the existing approaches to understanding the behaviour. Nevertheless, artificial neural networks can be trained on data sets, such as those obtained from different views of a face, and can learn to recognise different faces and so provide working models for many tasks for which there had not previously been any working models. The great advantage of artificial neural networks is that all the connection strengths are known, and so are easier to analyse than real nervous systems for which only a small fraction of the connection strengths are known. One straightforward technique for analysing the behaviour of such a network is to select **slow points** on the trajectory where the state of the system is changing least [17]. A measure of the changes in the state variables is given by a function analogous to the kinetic energy of mechanical systems:

$$k(\mathsf{X}) = \frac{1}{2}\|\frac{\mathrm{d}x_1}{\mathrm{d}t}, \frac{\mathrm{d}x_2}{\mathrm{d}t}, \cdots, \frac{\mathrm{d}x_\mathrm{n}}{\mathrm{d}t}\|^2 \tag{8.6}$$

and an example application of the technique is shown in Fig. 8.7. The technique picks out the parts of a trajectory close to a fixed point and by starting trajectories

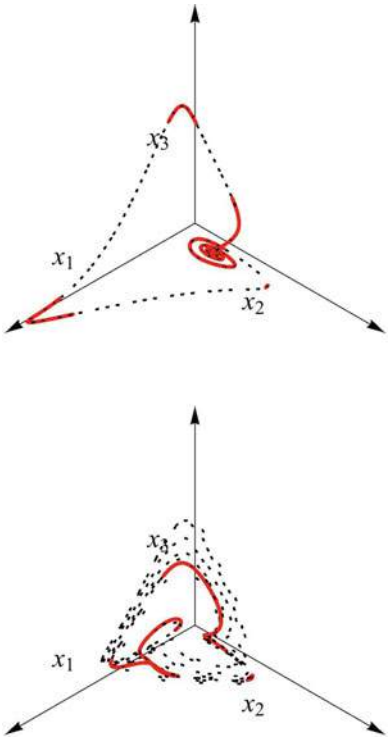


Fig. 8.7 Example of the use of slow points, plotted in red, to gain insight into the behaviour of an artificial neural net. By running simulations with different initial conditions a range of trajectories is generated, which can be used to uncover the regions in state space where the behaviour of the network is approximately linear. In this case the trajectories consist of heteroclinic cycles produced by the system of Eqs. 6.33 described in Chap. 6

from different locations all the fixed points can be identified and their behaviour characterised by applying the numerical techniques described in Chap. 7.

Understanding of what makes a brain decide on this or that behaviour is going to continue to require all the above approaches and undoubtedly others in the future. But whatever the approach, sooner or later neuroscientists are going to have to come up with the mathematics that is appropriate for their brains rather than their minds.

8.3 Mathematics for Neurons

A mathematical approach enables exploration of the properties of any system, irrespective of whether or not it has a known physical basis. Very often the successful outcome of the exploration is a classification of the possible behaviours of the system and identification of the archetypal instances of each class. An early example of the success of this approach was the enumeration of the possible sections of a conic surface and the derivation of their basic equations. A similar example is provided by the identification of the possible behaviours of a two-dimensional linear system around a fixed point.

However, understanding the brain requires more than just a classification of the possible behaviours of neural networks. Brains are so complicated that some part of some brain will probably provide an example of even the most esoteric behaviour discovered by physics or mathematics, but the behaviour may be extraneous to the successful interaction of an animal with its world. In order to understand the importance of such unexpected behaviours we have to go further and discover the significance for the animal's interaction with the world.

The brain can only effect interactions with objects which it can embody. What sort of mathematics best describes such an embodiment? The argument put forward in this book is that an object, especially a biological one, is best described by a vector field over a manifold.

To adapt a saying about a chess, n-dimensional Euclidean space is a sea in which a gnat may drink and an elephant may bathe. From a geometric perspective the point is that n-dimensional Euclidean space is a place where descriptions of the gnat and elephant can be accommodated simultaneously. The advantage of embedding data from biological objects in Euclidean space is that geometric relationships can easily be computed in Euclidean space and used to determine the manifolds on which the data lie.

The use of high-dimensional Euclidean spaces changes the way in which computations need to be done. For example, a straightforward method for telling if an object has moved or the eye has moved is to convert the retinal direction of the object to a head-based frame of reference in which the object does not change direction with movement of the eyes. Following this approach one can go on to ask where in the brain the conversion to a head-based frame of reference takes place and where the signal specifying the eye movement comes from. But an alternative approach is to embed both the retinal position of the object and the position of

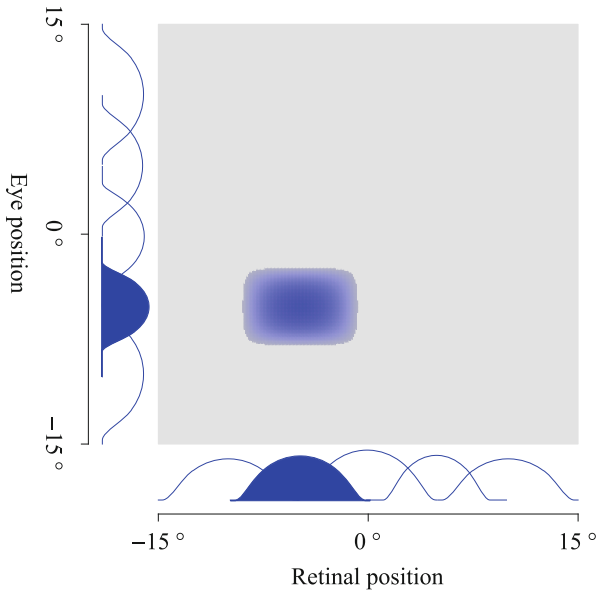


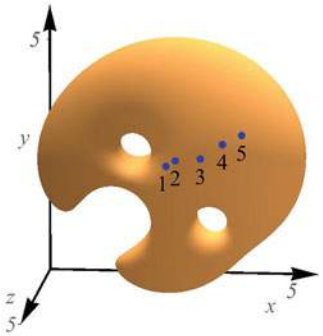
Fig. 8.8 Illustration of how the horizontal position of a target could be represented in a product space of neurons selective for retinal position and eye position with respect to the head

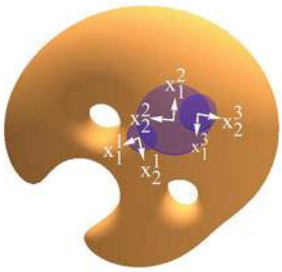
the eye with respect to the head in Euclidean spaces and then form the product space. One then has a Euclidean space big enough to hold every possible object/eye position coordinates as illustrated in Fig. 8.8.

Allied to the change in the nature of the computation is a change in the type of question that has to be asked to discover what is going on. The difference is illustrated in Fig. 8.9 which compares a description in Cartesian coordinates with a description in n-dimensional Euclidian space. In the Cartesian system of coordinates there is a frame of reference provided by the axes which are fixed at the origin. Using this description immediately leads to the question of the frame of reference of the brain. In the case of the oculomotor system the question might be whether the system uses an eye-based or a head-based frame of reference.

The over-arching goal of the program of identifying these archetypal behaviours is to arrive at a classification of the possible data structures that can be processed by neurons. For example, one possibility is that brains are adept at handling invariants of projections from three-dimensional space onto two-dimensional space.

The sensory and motor cortices are anatomically similar so it seems likely that they also handle data structures in the same way and consideration of how manifolds are processed in the cortices gives an idea of where the flexibility of the manipulations come from. In Fig. 8.10 two examples of two-dimensional manifolds are shown together with their outlines. The folded surface on the left could be an example of a visual stimulus which could be reconstructed from its outline by the visual cortex. On the right is another example of folded surface which could occur in





	x	y	z
1	0.52	1.30	−2.86
2	0.91	1.56	2.91
3	1.95	1.69	3.04
4	2.86	2.34	3.06
5	3.64	2.73	2.96

x_1^1	x_2^1	x_1^2	x_2^2	x_1^3	x_2^3
0.33	0.02	0	0	0	0
−0.01	0.05	−0.25	1.12	0	0
0	0	−0.23	0.27	0	0
0	0	0.46	0.67	0.01	−0.12
0	0	0	0	−0.32	0.60

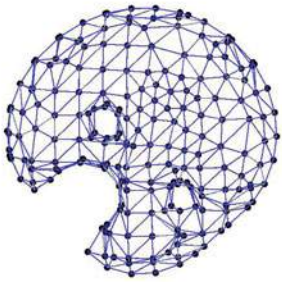


Fig. 8.9 Comparison of the use of embedding in three-dimensional Euclidean space and n-dimensional Euclidean space. The upper row of figures illustrates the coordinate systems, the middle row gives a table of the coordinates of the 5 points numbered in the top left figure and the lowest row shows how the surface might be represented. **Left Column**: Given a three-dimensional system of coordinates, one can use it as a frame of reference within which the position of every point on a surface can be specified. We are familiar with the effectiveness of this description from everyday interaction with computers. **Right Column**: Alternatively, one can describe the position and shape of a surface using coordinates in an n-dimensional space. In this approach the outline of the object is specified and properties of the surface are described locally by population coding. You might reasonably comment that I have taken a unified description you do understand and turned it into multiple little descriptions that you do not understand. But whilst our minds prefer the three-dimensional description, it may be that the neurons below the level of consciousness find working in n-dimensional space easier, because they can get by with using only local interactions

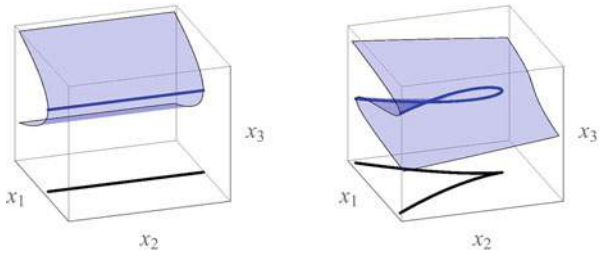


Fig. 8.10 Possible examples of manifolds in the cortex and their corresponding outlines. The figure on the left illustrates a folded surface reconstructed from an outline in the visual cortex. The figure on the right illustrates a slow manifold reconstructed from an outline in the motor cortex. The idea here is that just as the visual cortex interprets outlines the motor cortex could be passed an outline and could reconstructed the slow manifold required to specify a particular action

the context of control of eye movements. Comparison of these two examples leads to the idea that rather than specifying the slow manifold directly when an action is to be made, the motor cortex could have the outline passed to it and reconstruct the required slow manifold in the same way as the visual cortex. Once established the behaviour of the cortical neurons would follow that of a slow-fast system for an action and recordings from neurons in the motor cortex have provided evidence for the spiral attractor that forms part of such a system [5].

For a single view of smooth surface, the only invariants of the projection of the outline onto the plane are the fold and the cusp. But when successive views are allowed then additional invariants arise, such as an outline that looks like that of a swallow's tail [3] and such invariants may be especially relevant when movement is involved, although this remains to be demonstrated experimentally.

As a final point, whilst the brain may make use of embedding in Euclidean space, Hebbian learning, critical points and slow-fast control to embody dynamical systems, the brain itself is probably best not treated as a single dynamical system. Many areas of the brain contain topographic arrays of neurons that show a steady increase with position in the array of the spatial range over which they operate, of which the visual cortical areas are the most obvious example. But such ranges also occur in time, as in the neural integrator, where a range of time constants are computed. This gives the brain enormous flexibility in that the appropriate one can be selected as it is required [18]. So if the brain has to interact with an object, one idea is that it builds a model of the behaviour of the object and uses it to calculate how to get the target outcome. But this solution may be misleading in that the brain may only ever work with temporary models, and rather than making a complex set of calculations with a model in which the parameters are fixed, it repeatedly creates a new model for each step of the task. This is why we have to be very cautious in analysing neural data, because it may only lie on a given manifold for a very short duration.

To finish, I should point out that neuroscientists more than anybody should know that our understanding of the world depends on our individual brains and

the geometric approach to tackling how brains work will not be for everyone. My aim in writing this book has been to introduce enough background and examples of the approach for the reader to be able decide for themselves whether or not to take it further.

8.4 Further Reading

8.4.1 Brain

Glickstein M Neuroscience: a historical introduction. The MIT Press, Cambridge, Massachusetts, 2014. It is easy to lose one's way when faced with the myriad of experimental findings about the brain. This book is exceptional in that the author easily describes the different approaches that have led to successful insights into the workings of the brain, so that the reader comes away with a clear understanding of how to prioritise experimental findings in neuroscience.

Nassim C Lessons from the Lobster. The MIT Press, Cambridge, Massachusetts, 2018. This books gives the reader insight into what working in a multi-disciplinary neuroscience laboratory is like, as well as very clearly explaining the significance of experimental findings from the stomatogastric ganglion of the lobster.

8.4.2 Geometry

Katok, A and Climenhaga V Lectures on Surfaces: (almost) everything you wanted to know about them. American Mathematical Society, Student Mathematical Library Volume 46, 2008. It is easy to lose one's way when faced with the myriad of mathematical techniques that one could apply to the workings of the brain. This extensively illustrated book gently guides the reader through the different geometric approaches, many of which have already been applied to neural mechanisms. The authors concentrate on surfaces which enables them to give relatively simple explanations and proofs so that the reader gains a clear impression of what would be involved in taking it further with a particular technique. More than that, the authors have unified the approaches by using each of them to reveal an alternative formulation of the Euler characteristic of the surface. In its simplest form the Euler characteristic relates the number of faces, edges and vertices of a polyhedron. But that is just number one of the seven alternative formulations the authors arrive at by the end of the book. There are some of the out-of-sequence parts and the index is minimalist so that you have to read it more than once to become familiar with where all the definitions are, but maybe that is no bad thing.

Kendig K Never a dull moment. American Mathematical Society/Mathematical Association of America Spectrum Volume 93, MAA Press, Providence, Rhode

Island. 2018. Many mathematical ideas appear perplexing at first, mainly because they seem to have come from nowhere. In this instructive biography, the author explains the background to the discoveries of the pioneering topologist Hassler Whitney. In particular, he uses many diagrams to make the point that Whitney began exploring topics such as embedding and projections by drawing figures of specific cases, although these were not included in his subsequent published papers.

References

1. Block, B.A.: Evolutionary novelties: how fish have built a heater out of muscle. Am. Zool. **31**, 726–742 (1991)
2. Brainard, M.S., Doupe, A.J.: Translating birdsong: songbirds as a model for basic and applied medical research. Annu. Rev. Neurosci. **36**, 726–742 (2013). https://doi.org/10.1146/annurev-neuro-060909-152826
3. Bruce, J.W.: Seeing-the mathematical viewpoint. Math. Intell. **6**, 18–25 (1984)
4. Butler, A.B., Hodos, W.: Comparative Vertebrate Anatomy. John Wiley & Sons, Inc., Hoboken, New Jersey (2005)
5. Churchland, M., Cunningham, J., Kaufman, M., Foster, J., Nuyujukian, P., Ryu, S., Shenoy, K.: Neural population dynamics during reaching. Nature **487**, 1–56 (2012)
6. Doya, K., Boyle, M., Selverston, A.: Mapping between neural and physical activities of the lobster gastric mill. Adv. Neural Inf. Process. Syst. **5**, 913–920 (1993)
7. Han, Q., Hong, J.X.: Isometric embedding of Riemannian manifolds in Euclidean spaces. In: Mathematical Surveys and Monographs, vol. 130. American Mathematical Society, Providence (2006). https://doi.org/10.1090/surv/130
8. Helmholtz, H.: Treatise on Physiological Optics. Translated by J.P.C. Southall. Reprinted by Dover, New York (1910)
9. Marzban, C., Yurtsever, U.: Baby morse theory in data analysis. In: Proceedings of the 2011 Workshop on Knowledge Discovery, Modelling and Simulation, pp. 15–21. San Diego, Association for Computing Machinery, New York (2011). https://doi.org/10.1145/2023568.2023576
10. Miller, J.: Understanding and misunderstanding extraocular muscle pulleys. J. Vis. **7**, 1–15 (2007). https://doi.org/10.1167/7.11.10
11. Montgomery, R.: Gauge theory of the falling cat. In: Enos, M.J. (ed.) Fields Inst. Commun. Dynamics and Control of Mechanical Systems: The Falling Cat and Related Problems, pp. 195–281. Fields Institute Communications (1993). https://doi.org/10.1090/fic/001
12. Nassim, C.: Lessons from the Lobster. MIT Press, Cambridge (2018)
13. Noakes, L., Mees, A.: Dynamical signatures. Physica D **58**, 243–250 (1992). https://doi.org/10.1016/0167-2789(92)90112-Z
14. Rodieck, R.: The First Steps in Seeing. Sinauer Associates, Inc., Sunderland (1998)
15. Sanguinetti, G., Citti, G., Sarti, A.: A model of natural image edge co-occurrence in the rototranslation group. J. Vis. **10**, 1–16 (2010). https://doi.org/10.1167/10.14.37
16. Sillar, K., Picton, L., Heitler, W.: The Neuroethology of Predation and Escape. John Wiley & Sons, Ltd., Chichester (2016)
17. Sussillo, D., Barak, O.: Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput. **25**, 626–649 (2013)
18. Wang, J., Narain, D., Hosseini, E., Jazayeri, M.: Flexible timing by temporal scaling of cortical responses. Nat. Neurosci. **21**, 102–110 (2018). https://doi.org/10.1038/s41593-017-0028-6
19. Wilson, H.R.: Spikes, Decision and Actions. Oxford University Press, Oxford (1999)

Chapter 9
Appendix: Background Material

9.1 Brains

Brains act to ensure that the state of the body is maintained at a stable, functioning level—a property referred to as homeostasis. The brain uses the same type of mechanism to ensure that the environment which the individual lives in remains stable. Imagine a cat sitting in front of a fire. It moves so that it so not too hot and not too cold. Now imagine a lot of cats sitting in front of the fire. Every cat wants to sit in the same place. Brains developed to sort this type of problem out. Of course brains developed to handle more than just the problem of who sits where in front of the fire. The top cat, in terms of the one who gets what it wants, will have the largest repertoire of behaviours so it can counter the greatest range of situations that the world throws at it.

From the homeostasis perspective it follows that a brain will require several components:

(1) A mechanism for recognising situations that have happened before, so that an appropriate counteraction can be taken.
(2) A mechanism for controlling and executing the counteraction.
(3) A mechanism for prioritising the strength and direction of behaviour, an aspect of behaviour referred to as motivation.
(4) A mechanism, referred to as learning, for altering the behaviour of all three of the above components to adapt them to a changing world.

Although brains require all four mechanisms there is no straightforward mapping of these functions onto the most prominent structures of the brain, some of which are illustrated schematically in Fig. 9.1.

The cortex is where sensory inputs and motor outputs are used to form a coherent understanding of the world and what to do about it. There are a number of specialised regions, such as the visual areas at the back of the head, and the somatosensory and motor areas down the side of the head.

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3_9

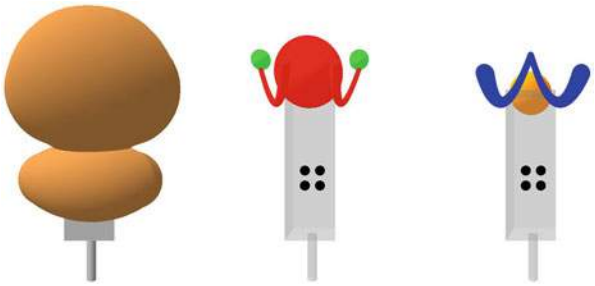


Fig. 9.1 A schematic diagram of some of the main anatomically identifiable of the brain. **Left Figure**: External appearance showing the cerebral cortex and cerebellum. **Middle Figure**: Some of the structures lying deeper in the brain: basal ganglia (red), amygdala (green) and brainstem (grey). The four black dots represent the paired superior and inferior colliculi. **Right Figure**: Other structures lying deeper in the brain: hippocampus (blue) and thalamus (orange)

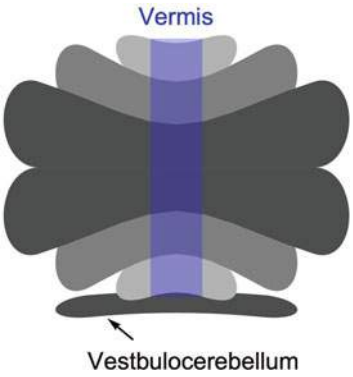


Fig. 9.2 Schematic illustration of the lobular structure of the cerebellum. The cerebellum is viewed from above and has been unrolled so the vestibulocerebellum is visible

Damage to a specialised area causes long term problems with the corresponding function. The primary motor area, also referred to as M1, provides the main input to the pathway from the cortex to the skeletal muscles. Damage to M1 results in a paralysis, from which the person usually recovers, and long-term difficulty in coordinating movements.

The next most prominent external structure of the brain after the cortex is the cerebellum. Damage to the cerebellum results in uncoordinated movements. The cerebellum is made up of a number of small lobes, as illustrated highly schematically in Fig. 9.2. Like the cortex, the cerebellum is comprised of two hemispheres but in this case the two hemispheres are joined by an unpaired portion referred to as the vermis. At the base of the cerebellum are three further lobes. The central one of these lobes is referred to as the nodule and each of the outside ones is referred to as the flocculus. The flocculus and nodule receive inputs directly from the vestibular nucleus and project back to them and so are referred together as

the vestibulocerebellum. The main portion of the cerebellum projects to the dentate nucleus and on to the cortex via the dorsal thalamus and also to the red nucleus, which is a source of descending spinal pathways. The vermis and surrounding region projects to the fastigial nucleus and on to the spinal neurons. The three attachments of the cerebellum to the brainstem are referred to as peduncles and the outputs from the cerebellum leave by the superior peduncle. Input from the somatosensory and motor cortices arrives via the middle peduncular lobe. Sensory signals from the spinal cord and vestibular system enter the cerebellum through the inferior peduncle. Neural signals from the nucleus referred to as the inferior olive also enter via the inferior peduncle.

In the middle of the brain are the thalamus and basal ganglia, which are comprised of three major nuclei, namely the caudate nucleus, putamen and globus pallidus and two smaller nuclei, the substantia nigra and the subthalamic nucleus. Brain damage studies have shown that the basal ganglia are primarily involved in movement selection and control.

Also forming part of the midbrain is the amygdala, which takes part in linking motivation with action, and the hippocampus. Damage to the hippocampus in humans results in difficulties with long term memory formation. The name hippocampus was adopted because of a perceived resemblance of the shape of the structure to that of a seahorse, but another perceived resemblance to a ram's horn lead the alternative name of Ammon's horn. The three main subdivisions of the hippocampus are designated CA1, CA2 and CA3 where CA is an abbreviation of the Latin name Cornu Ammonis. In between CA1 and the surrounding entorhinal cortex, with which the hippocampus has extensive interconnections, lies a region referred to as the subiculum.

The insect brain, illustrated in Fig. 9.3, has a different ground plan from the vertebrate brain. The major part of the insect brain is comprised of the optic lobe. The output of the sensory cells of the compound eye passes to the lamina, then the medulla and then on to the lobula. Signals from olfactory sensors on the antennae pass to the antennal lobe and on to the mushroom bodies, which are involved with memory and learning. Between the mushroom bodies is the central complex that integrates sensory inputs and processed signals from the mushroom bodies.

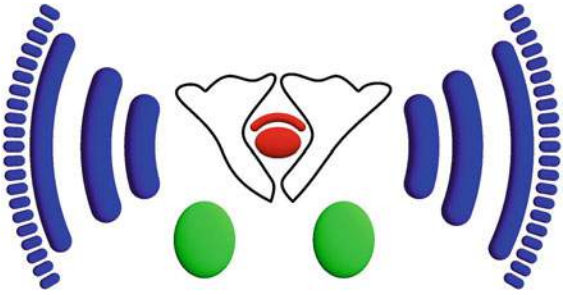


Fig. 9.3 A schematic diagram of the insect brain. The components of the optic lobe is shown in blue. The antennal lobe is shown in green and the mushroom bodies are shown in outline. The central complex is drawn in red

9.2 Calculus

Description of the effect of an action on an object requires specification both of the magnitude of the action and of the new situation of the object. A typical example could involve the contraction of a muscle changing the position of an object. More generally, any quantity which can change is referred to as a **variable** and the relation between a pair of variables $(x, f(x))$ is referred to as a **function** if each value of x is associated with a unique value of $f(x)$. A visual representation of a function can be made by plotting x along a horizontal axis and y along a vertical axis. The plot of the (x, y) pairs generated by the function form a **graph** of the function.

The effect of a change by an amount h in a variable from x to $x + h$ depends on the ratio of h to $f(x + h) - f(x)$. This ratio will depend on the size of the step h and the value at x can be defined as the limiting value of the ratio as h approaches zero. This value is referred to as the **derivative** of f(x) at x, and is denoted by $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$ and is formally defined as:

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{h \to 0} \frac{f(x+h) + f(x)}{h} \tag{9.1}$$

The graphical representation extends to the process of differentiation. One can draw a straight line between the points $(x, f(x)) and (x + h, f(x + h))$, and as h is made smaller the line will align with the tangent to the graph at the point $(x, f(x))$. The derivative corresponds to the slope of the tangent as illustrated in Fig. 9.4.

The graphical approach can also be used to define a further quantity referred to as the **integral** of $f(x)$, also illustrated in Fig. 9.4, which works in the opposite way to the derivative. The definite integral of a function $f(x)$ over an interval from a value a of x to a value b of x corresponds to the area under the graph of the function.

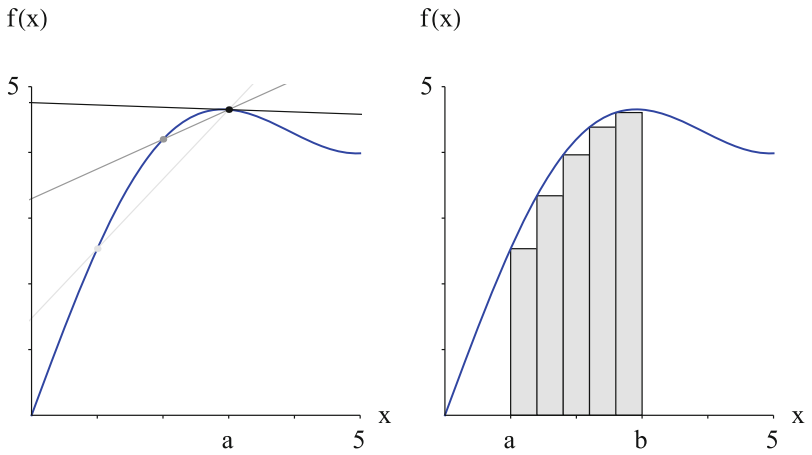


Fig. 9.4 Illustration of the derivative $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$ evaluated at x = a and the integral $\int_a^b f(x)dx$ as limiting processes. The graph of the function f(x) is plotted in blue

To compute this area one can begin by dividing the interval between a and b into n equal parts letting the width of each part be denoted by h. The sum S_n of the areas of the n rectangles which can be fitted under the graph of $f(x)$ then gives an approximation to the area under the graph between a and b. The integral of $f(x)$ between a and b is given by the limiting value of S_n as the width of the vertical strips is made vanishingly small.

$$\int_a^b f(x)dx = \lim_{h\to 0} S_n \tag{9.2}$$

A simple example of a function is $f(x) = kx$ where k is a constant. The graph of this function is a straight line and any function which has a graph that is a straight line is referred to as a **linear** function. Given a value for x, there is a corresponding value for $y(= kx)$ and this pair of values is represented by a point a distance x along the horizontal axis and y along the vertical axis. Also shown is a graph of the derivative of $f(x)$. Since the slope of the graph of $f(x)$ is constant at k, the derivative is also constant at this value, as illustrated in Fig. 9.5.

One of the simplest nonlinear functions is $f(x) = x^2$ and the graph of this equation, which is drawn in Fig. 9.6, is referred to as a parabola. This nonlinear function is an example of a power function which has the form $f(x) = x^r$ where r specifies the number of times that x is multiplied by itself.

Another useful class of functions is provided by functions of the form $f(x) = r^x$. In particular the nonlinear function referred to as the **exponential** function has the form $f(x) = \mathrm{e}^x$ where e is a constant which is approximately 2.7183. The value of this function is the same as that of its derivative, for all values of x, as illustrated in Fig. 9.7.

Another commonly occurring nonlinear relationship occurs with cyclical behaviour. If one takes a point moving anticlockwise round a circle of unit radius at a unit velocity then one can define two new functions. The sine function is given by the ratio of the x coordinate of the point to the radius and the cosine function is given by the ratio of the y coordinate of the point to the radius, as illustrated in Fig. 9.8.

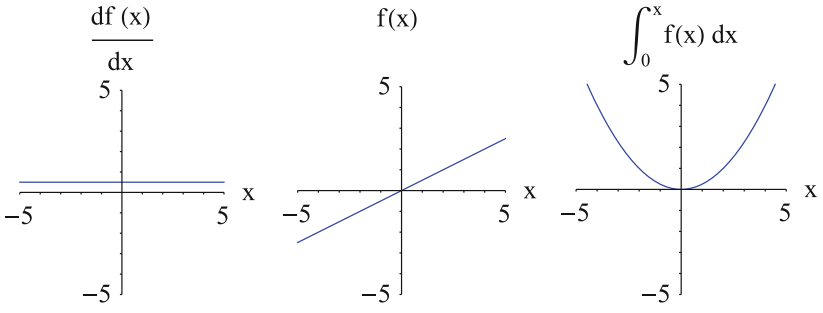


Fig. 9.5 Example of a graph of the function f(x) = k x with k = $\frac{1}{2}$

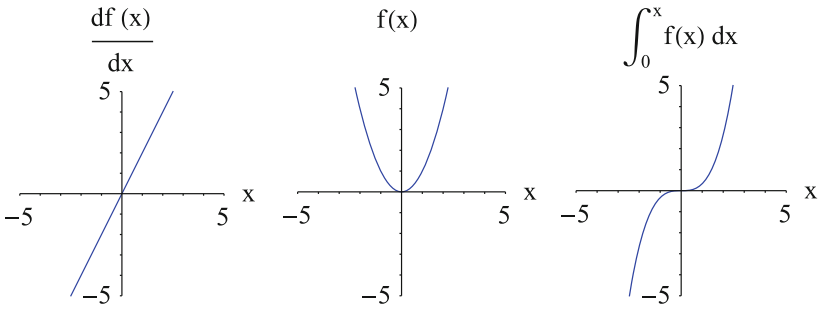


Fig. 9.6 Example of a graph of the function $f(x) = x^r$ with $r = 2$

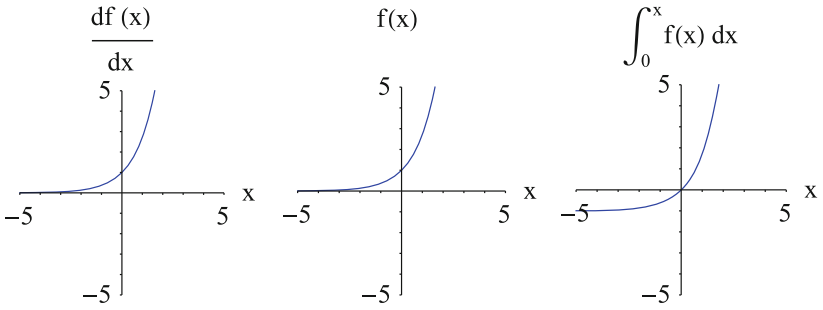


Fig. 9.7 Example of a graph of the exponential function

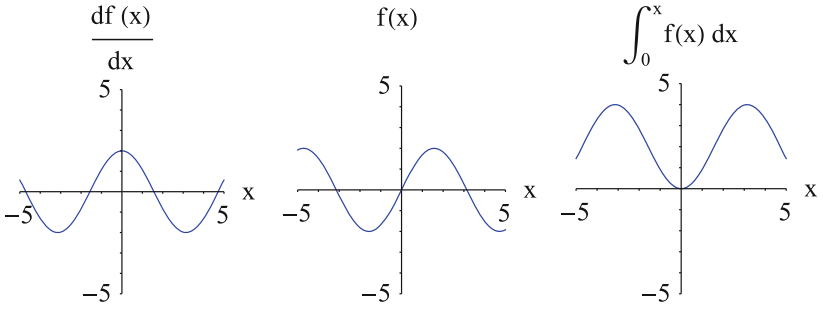


Fig. 9.8 Example of a graph of the sine function. The derivative of a sine function is a cosine function and the integral is a negative cosine function

For the example functions plotted in the figures it can be seen that differentiation of a function is the inverse of integration of the function. This relationship is made use of explicitly in Chap. 5 so I include now a justification of the statement but you can save this part for later without impeding your reading of this book.

To show that the inverse relationship follows from the definition of differentiation and integration, begin by expressing a definite integral in terms of indefinite integrals, which give the value of a definite integral from a fixed starting value u:

$$\int_a^b f(x)dx = \int_u^b f(x)dx - \int_u^a f(x)dx = F(b) - F(a) \tag{9.3}$$

where $F(a)$ and $F(b)$ are indefinite integrals. Given this notation one can specify the relationship between integration and differentiation as:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x) \tag{9.4}$$

and this equivalence is referred to as the **fundamental theorem of calculus**. To understand why this relationship holds begin by noting that from the definition of differentiation it holds that:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \tag{9.5}$$

As the graph of $f(x)$ does not have any breaks in the examples shown, for any arbitrarily small number ϵ one can define a variable y such that $x - h \leq y \leq x + h$ and find a value of h small enough to ensure that that $f(x) - \epsilon \leq f(y) \leq f(x) + \epsilon$. Furthermore if m and M are the minimum and maximum values of $f(y)$ on the interval then $f(x) - \epsilon \leq m \leq M \leq f(x) + \epsilon$. From the definition of an integral it follows that:

$$h\mathrm{m} \leq \int_x^{x+h} f(x)dx \leq h\mathrm{M} \tag{9.6}$$

and

$$f(x) - \epsilon \leq \frac{\int_x^{x+h} f(x)dx}{h} \leq f(x) + \epsilon \tag{9.7}$$

Since ϵ can be made arbitrarily small it must hold that:

$$\lim_{h\to 0} \frac{\int_x^{x+h} f(x)dx}{h} = f(x) \tag{9.8}$$

as required to show that the fundamental theorem holds.

Glossary of Terms with Hidden Assumptions

Attractor An attractor is a region of state space which captures trajectories that arrive in its surroundings. But what shape is the region? This question is usually avoided by considering simple cases where the shape is well defined. See the entries for point attractor and continuous attractor.

Conductance-based model A conductance-based model of a neural network gives a description of the membrane potential of each neuron. Such models are useful for investigating the effects of the level of a particular transmitter. The drawback is that they require many parameters to specify the characteristics of the ion channels. Also, they do not usually incorporate the effects of the close proximity of a population of neurons firing together, which generates electric field currents that enable neurons to affect the behaviour of other neurons with which they have no direct connection.

Continuous attractor The term continuous attractor has been introduced to refer to a collector of adjacent point attractors in order to model memory for continuous variables. For example, the horizontal eye position controller could be modelled by a line attractor consisting of a line of point attractors each corresponding to a particular position. Another example is the circuit that holds the heading direction in a fly. Since the activity in the neural network consists of a localised hill of activity representing orientation or heading direction such neural embodiment of continuous attractors are also referred to as bump attractors. But given the definition of an attractor as a region of space that captures all trajectories that arrive in its surroundings, two adjacent points cannot both be point attractors. What you can have is a line or a surface, such as the slow manifold used in the action model, in state space to which all trajectories are attracted. The dynamics within the attracting region remain to be specified. In the case of horizontal eye position, the state of the oculomotor integrator moves slowly to a point attractor at the origin. In the case of an action, the state follows a trajectory towards a spiral attractor at the equilibrium position.

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3

Dimensionality reduction At first glance, throwing away unimportant parts of the data to obtain a simpler system whose behaviour is effectively the same as the original system seems like a good idea. The problem lies in the definition of unimportant. The most widely used technique of dimensionality reduction in neuroscience is principal components analysis. But principal components analysis of the cone responses to the Munsell chip set gives that 99.7 % of the variance is carried by the light-dark channel. So following the dimensionality reduction approach one would ignore how the nervous system deals with hue and colour saturation data entirely.

Firing rate model A firing model of a neural network is one in which the output of each neuron can only be positive, corresponding to the average of the outputs of a population of biological neurons whereas in a conductance-based neuron model the output of each neuron corresponds to its membrane potential. The firing rate approach was based on the assumption that a neuron can be either excitatory or inhibitory. Subsequently it was found that a single neuron can release more than one type of transmitter and so may excite or inhibit other neurons, depending on their ongoing acitivity.

Point attractor A point attractor consists of a single point in state space that nearby trajectories all converge to. In the case of a linear system such a point is also referred to sink. The idea of a point attractor has proved especially useful for thinking about memory. If it is assumed that separate objects are represented by separate attractors, then an input to a neural network from measurements of part of an object could force the state of the network onto a trajectory that passes near to the attractor corresponding to the object, whereupon the state drops onto the attractor and the object is recognised.

Index

Page numbers in bold indicate definitions

R. A. Clement, *Mathematical Tools for Neuroscience*, Lecture Notes in Morphogenesis, https://doi.org/10.1007/978-3-030-98495-3