
Learning and Recognition by a Dynamical System with a Plastic Velocity Field

Author:

Daniel GASCOYNE

Supervisor:

Dr. Alexander BALANOV

A DOCTORAL THESIS

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR
THE AWARD OF DOCTOR OF PHILOSOPHY OF LOUGHBOROUGH
UNIVERSITY

September 2015

Contents

Abstract	vii
1 Introduction	1
1.1 Introduction to Learning in AI	2
1.1.1 Human vs Rational Approach	4
1.1.2 The Human Brain and ‘Brain’ Research	4
1.1.3 AI without the Limitations of Biology	6
1.2 Prominent AI Research Areas	8
1.2.1 Neural Network’s	8
1.2.2 Decision Tree Learning	17
1.2.3 Association Rule Learning	18
1.2.4 Genetic Algorithms	19
1.3 ‘Memory Foam’ Approach to Unsupervised Learning’	22
1.4 Elements of Non-Linear Dynamics	24
1.4.1 Differential Equations and Attractors in Phase Space	24
1.4.2 The van der Pol Oscillator	25
1.5 Numerical Approaches	32
1.5.1 Solving Differential Equations Numerically	32
1.5.2 Gaussian Noise	33
1.6 Summary and Conclusions	33
2 Classification and Recognition by a Dynamical System with Self-adapting Vector Field	35
2.1 ‘Memory Foam’ Approach to Unsupervised Learning’	35
2.1.1 Constraints on the Stimuli and Convergence of the Potential	37
2.2 Initial Simulations	39
2.2.1 Simulating Stimuli	40
2.2.2 Application to the Memory Foam Model	42
2.3 The Influence of Gaussian Width on Class Formation	42
2.4 Extending the Model to N-Dimensions	50
2.5 Summary and Conclusion	51
3 Non-Linear Dynamics of Learning and its Application to Colour Recognition	57
3.1 Colour Perception	58
3.1.1 Machine Vision	59

3.2	Colour Recognition and Classification by a Dynamical System with Self-adapting Vector Field	59
3.2.1	Assessing the accuracy of RGB value identification	60
3.3	Applying a Simple RGB Stimuli	62
3.4	Learning Colours by Observing Cartoons	68
3.4.1	Southpark	68
3.4.2	Futurama	71
3.5	A One-dimensional Approach to Colour Coding	73
3.5.1	Mapping RGB to Wavelength	73
3.5.2	The One-Dimensional Learning Model	81
3.6	Summary and Conclusion	84
4	Representation of Class Recognition via Limit Cycles	91
4.1	A Hierarchical Representation of Class Recognition	92
4.2	Utilizing the van der Pol System to Describe Recognition	93
4.2.1	Numerical Calculations	98
4.3	Modelling the Foam Profile with Parabolas	102
4.3.1	Parabola width and Frequency	105
4.4	Limit Cycles Recognition Scheme	106
4.4.1	One Dimensional Colour Recognition with Oscillatory Dynamics	109
4.5	Summary and Conclusion	110
5	Coupled Oscillators and higher order Phase Dynamics	115
5.1	Coupled Oscillators	116
5.2	Coupled van der Pol Oscillators	117
5.2.1	The Truncated Equations for a Pair of van der Pol Oscillators	118
5.2.2	Dissipative Coupling	119
5.3	Limit cycles for an two-dimensional Potential	121
5.3.1	Realizing Limit Cycles, Quasi-periodic orbits and Oscillation Death	125
5.3.2	Adjusting ε as a function of distance traversed	131
5.4	A System of 3 Diffusively Coupled van der Pol Oscillators	134
5.4.1	Poincaré approach to characterising the Dynamical regions	135
5.5	Oscillatory Recognition of Colour	144
5.6	Summary and Conclusion	146
6	Summary and Conclusion	151
A	Appendix	159
A.1	Runge-Kutta Numerical Integration	159
A.2	The Box-Muller Transformation	159
A.3	Realizations of Colour Learning	161
A.4	The Truncated Equations for a Pair of Coupled van der Pol Oscillators	163
A.5	The Truncated Equations for Three Coupled van der Pol Oscillators	165
	Paper	167

Abstract

Learning is a mechanism intrinsic to all sentient biological systems. Despite the diverse range of paradigms that exist, it appears that an artificial system has yet to be developed that can emulate learning with a comparable degree of accuracy or efficiency to the human brain. With the development of new approaches comes the opportunity to reduce this disparity in performance. A model presented by Janson and Marsden [arXiv:1107.0674 (2011)] (Memory foam model) redefines the critical features that an intelligent system should demonstrate. Rather than focussing on the topological constraints of the rigid neuron structure, the emphasis is placed on the on-line, unsupervised, classification, retention and recognition of stimuli. In contrast to traditional AI approaches, the systems memory is not plagued by spurious attractors or the curse of dimensionality. The ability to continuously learn, whilst simultaneously recognizing aspects of a stimuli ensures that this model more closely embodies the operations occurring in the brain than many other AI approaches. Here we consider the pertinent deficiencies of classical artificial learning models before introducing and developing this memory foam self-shaping system.

As this model is relatively new, its limitations are not yet apparent. These must be established by testing the model in various complex environments. Here we consider its ability to learn and recognize the RGB colours composing cartoons as observed via a web-camera. The self-shaping vector field of the system is shown to adjust its composition to reflect the distribution of three-dimensional inputs. The model builds a memory of its experiences and is shown to recognize unfamiliar colours by locating the most appropriate class with which to associate a stimuli. In addition, we discuss a method to map a three-dimensional RGB input onto a line spectrum of colours. The corresponding reduction of the models dimensions is shown to dramatically improve computational speed, however, the model is then restricted to a much smaller set of representable colours.

This models prototype offers a gradient description of recognition, it is evident that a more complex, non-linear alternative may be used to better characterize the classes of the system. It is postulated that non-linear attractors may be utilized to convey the concept of hierarchy that relates the different classes of the system. We relate the dynamics of the van der Pol oscillator to this plastic self-shaping system, first demonstrating the recognition of stimuli with limit cycle trajectories. The location and frequency of each cycle is dependent on the topology of the systems energy potential. For a one-dimensional stimuli the dynamics are restricted to the cycle, the extension of the model to an N-dimensional stimuli is approached via the coupling of N oscillators. Here we study systems of up to three mutually coupled oscillators and relate limit cycles, fixed points and quasi-periodic orbits to the recognition of stimuli.

KEYWORDS: Learning, Unsupervised, Self-shaping, Non-Linear Dynamics, van der Pol, Limit cycle, Quasi-periodic.

Chapter 1

Introduction

Learning is described as “the acquisition of knowledge or skills through experience, practice, or study, or by being taught” [145].

The development of artificial intelligence (AI) has occupied researchers for over half a century, however, the field remains far from concluded. In reality the notion of AI has fascinated mankind for millennia. Greek mythology describes Talos, an artificial man of bronze, and ancient Chinese texts describe human-shaped figures of mechanical handiwork [118]. Stories can be found from every corner of the world and throughout human history. However, the modern reality of AI is far less impressive than the portrayals of science fiction. There has been much speculation about the future of AI and the potential impact it may have. In popular media Stephen Hawking has recently cited AI as a real danger to mankind [26]. At present this may seem more like a summary of the 1984 film ‘Terminator’ than a genuine concern. Nevertheless, he is not the first scientist to express these views [38]. This predicted singularity event, if realistic, requires a substantial break-through in AI, which currently appears to pose little risk.

Winner of quiz show ‘Jeopardy!’ in 2011, IBM’s Watson may be considered the forerunner in AI [47]. Applying DeepQA [48] it is currently being utilized in specialized domains such as health care [49]. Watson searches a large database of knowledge and returns solutions based on human input. Although it exhibits natural language processing, and is arguably a sophisticated tool, it is undeniably algorithmic and does not ‘learn’ as we perceive a human to, functioning instead as a next generation search engine.

The development of any learning system presents significant challenges. Can a machine really learn and how should we define intelligence? These may appear questions better posed to philosophers and psychologists, however, we cannot hope to engineer intelligent systems without first discussing objectives. To define intelligence we cite the definition of Russell and Norvig: “An intelligent agent is a system that perceives its environment and takes actions that maximize its chances for success.” [139]. This definition remains intentionally vague. This is mirrored by the range of approaches taken to learning, intelligence, and AI, which are exceptionally diverse.

Unlike in other scientific fields, where new phenomena or results are theoretically predicted and then explored, a working intelligent system already exists and can be studied; the human brain. The human brain may be considered one of the most complex systems available for scientific study. Not only does it have in excess of 120 billion computational units, but the connections between these units are thought to number in excess of 100 trillion [168]. If we also consider that the brain

is known to exhibit temporal changes across a persons lifespan, and that neural processes can be altered as a result of neurological trauma or disease, then the magnitude of the modelling challenge can be put into perspective.

The consensus within psychology is that stimuli-action pairs are retained in memory by the brain [82]; this memory is clearly very different to that of traditional computers. Presented with a stimulus the brain is able to recall the appropriate response and make associations with other relevant information. On a basic level encoding new pairs to memory can be considered to encapsulate the essence of learning. At present, an understanding of how biology actually encodes this information remains elusive.

Multiple definitions of machine learning exist, however, how best these can be realized is far from apparent. The field of machine learning considers algorithms that build predictive/decision making models based on inputs prescribed to a system¹. These approaches should be contrasted with conventional computing where programmed instructions are explicitly executed. As may be expected there exists a plethora of machine learning techniques. Many have overlapping applications, as well as drawbacks and advantages when compared to their compatriots. Neural networks [135] have remained a relatively effective, popular approach since the inception of the field. Nevertheless, they are by no means the only currently studied paradigm of learning. Approaches such as Decision Tree Learning [139] and Support Vector Machines [31] have gained traction and may now be implemented into AI. Other means for representing learning, such as Genetic Algorithms [110], have had a profound effect on the way learning is conceptualized, but now receive less attention.

1.1 Introduction to Learning in AI

Learning within the domain of AI may be accomplished by training a system with stimuli-action pairs, or by allowing a system to adapt its parameters until it is able to effectively elicit appropriate responses to a stimuli [84]. When known stimuli-action pairs are used to educate a system, the learning is said to be supervised. This is not the only possible way to achieve learning. Alternatives to supervision include; reinforcement [33], semi-supervised [28] or fully unsupervised learning [55].

- Reinforcement learning is concerned with reward maximization. The system's responses to stimuli are compared to desired outcomes and feedback is provided in the form of a reward function. This allows the system to promote suitable learning rules and neglect those that result in a reduction of reward [12]. Reinforcement learning is a highly studied approach as it may be seen as a method that closely emulates one way in which a human may learn. Let us consider a simple culinary analogy to help explain the principal. When we eat something we decide whether it tastes good or bad. If it tastes bad this is equivalent to negative feedback. We are less likely to eat the same food again. In contrast, we are more likely to seek food that we determine tastes nice. Where feedback is positive the behaviour is more likely to be repeated, whereas negative feedback causes us to avoid such decisions in future. Various complications of reinforcement learning may be considered. These include the exploration vs. exploitation trade-off, caused by sub-optimal initial solutions, and the credit-assignment problem, resulting from delayed feedback [12].

¹Tom M. Mitchell postulated the definition of machine learning as "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E." [111]

- Goal based learning is closely related to reinforcement learning with the addition that it incorporates the influence of other learning machines [55]. This may infer competition between different elements, or conversely, the notion of teamwork. Rather than the system assessing success against its own experiences, it also considers a measure of the other competing elements of the system. Such approaches are typically utilized in a game based environments, goal based algorithms being utilized by game simulators such as those applied to chess or checkers [96].
- An unsupervised learning method does not rely on an external teacher or labelled data sets. Like a human, the system identifies characteristic properties of the input without direct instruction by finding patterns within data. Unsupervised learning is sometimes referred to as clustering or unsupervised pattern classification, it is essentially a pure form of data mining [87].
- Between the supervised and unsupervised classes lie semi-supervised approaches [28]. These include algorithms that learn from incomplete data sets. These realisations of learning may better emulate human learning as opposed to fully supervised techniques. Some feedback and instruction is utilized, but the system is able to generalize, find patterns in unfamiliar data, and draw comparisons to previously learnt information. This filtering and fitting of ambiguous examples still leads to effective learning.

It is highly unlikely that a human learner will ever experience the exact same environment multiple times. Relevant stimuli-action pairs are identified and retained, whilst large quantities of data may be considered superfluous. Retained information is then used in other inter-related environments. Given an unfamiliar circumstance, decisions are reached even when stimuli are noisy or fuzzy, this minimally increases computation time when recognizing familiar stimuli out of the complex myriad of inputs. The flexibility of the human brain to correctly deal with unfamiliar circumstances by making associations between retained information and stimuli, can be perceived as unsupervised learning. This type of learning has proven the hardest to mimic utilizing machines, however, it appears to be predominant in human learners. Currently, the most effective machine learning routines tend to fall into either the supervised or reinforcement categories. It should be noted that approaches to model learning and memory are extremely varied, there exists a vast number of learning rules as well as a multitude of memory models. We should also consider that the same model of memory may be taught by many different rules.

Modern AI has become a vibrant field of science and engineering by drawing on knowledge and expertise from a diverse range of related subjects. Philosophy, mathematics, economics, neuroscience, psychology, computer engineering, control theory, cybernetics and linguistics are a few of the fields that have inspired techniques and methods applied to AI. When considering research into learning systems, we should also reference computational intelligence and soft computing as areas that may be considered synonymous with AI [37]. Shared components of these three disciplines include neural networks, support vector machines, fuzzy logic [81], genetic algorithms, swarm intelligence [15] and probabilistic reasoning mechanisms, such as bayesian networks [53]. Computational intelligence generally utilizes a stricter definition of intelligence than the AI field from which it evolved, referring to the complexity of a problem as a key indicator of intelligence. Soft computing can be seen as a branch of computer science that seeks to emulate the parallel reasoning skills of the human mind to utilize unpredictable fuzzy inputs, forming solutions where

conventional computing fails. The distinction between these fields has little impact on the proceeding discussion and readers should consider the term AI to also incorporate the ideas of these sub-fields.

1.1.1 Human vs Rational Approach

The design strategies employed in AI development may be loosely divided into human and rational approaches. A human approach seeks to recreate some essence of the human learner. This may refer to the structure of the system, such is the case for a neural network, or it may reflect the decision making process (an example being the ‘General Problem Solver’, introduced by Newell and Simon (1959) [119]). A rational approach removes focus from the human and instead concentrates on finding optimal solutions to the problems of learning, classification, retention and recognition. Within the rational framework, the clash between the logicist viewpoint and the connectionist approach rages on [21, 22]; neither able to discredit the other. A viable direction of research may lie in the amalgamation of these two concepts [20]. Connectionism may be considered to relate to the structure of the brain. Despite this, connectionist models typically have a rational aim. It is also apparent that most deviate from biological reality in order to enhance performance.

There is clear evidence that a ‘human’ approach can provide a successful model of learning; we are living proof of this. However, currently achieving this artificially presents some seemingly insurmountable challenges. A human based approach requires complex knowledge of the human brain, this is an issue as our knowledge in this area remains limited.

1.1.2 The Human Brain and ‘Brain’ Research

The human brain is an incredibly complex system composed of around 120 billion distinct elements called neurons [59, 57]. Each neuron has multiple connections to others within the network, interacting at transmission points known as synapses. As many as 100 trillion connections are thought to exist [168]. It is the way in which these connections are modulated that is thought to determine the group dynamics of the system and hence dictates the response to a given stimuli. It has been postulated that the modulation of connections is controlled by the interaction of active neurons as summarized by C. Shatz as “Cells that fire together, wire together” [36]. The storage and access of information in the brain relies on highly distributed representations and transformations that operate in parallel. Information appears to be stored in a distributed form across the synapses of the system [142]. The transmission of a neural impulse along a neuron is unidirectional, a feature that is not always replicated in artificial neurons. Although there are various types of neuron, the basic structure remains constant. The synapse of one neuron meets the dendrites of another. Action potentials travel along the dendrites to a soma, here the action potentials from each input neuron are combined to determine the response of the neuron. Based on its inputs, the neuron may propagate an action potential to subsequent neurons in the network. This is achieved by passing an impulse along the axon to its many synapses. The configuration of the network, the processes occurring in the soma, and the behaviour of the synapses all play pivotal roles in the behaviour of the system. A change in the excitation of a single neuron may have dramatic repercussions on the dynamics of the system as a whole. This property may be used to explain the seemingly infinite capacity of the human memory.

The myelin sheath, a dielectric material that encapsulates the axon dramatically increases the action potential transmission rate. Transmission speed is fibre type dependent with speeds

ranging from 2 mph up to 200 mph. Even at the upper limit, this is approximately 3 million times slower than the speed of electricity through a wire [115]. Therefore, in theory, an electronic brain should be able to process information at a much higher rate. Unfortunately there remains a significant gap between what exists in nature and what we can manufacture. The advantage of an electronic system is significantly faster linear processing and signal propagation than seen in biology, however, even the most advanced neural chips cannot be manufactured with comparable parallel processing power. Microchip manufacturing continues to push the frontiers of scale and performance. The forerunner, the IBM TrueNorth chip, is capable of modelling 1 million neurons and 256 million synapses. It consumes a mere 70 milliwatts of power, and is capable of 46 billion synaptic operations per second, per watt [105]. This still pales into insignificance if we scale these figures against the brain.

Unlike other areas of biology where processes are relatively well understood, there are still substantial gaps in our understanding of the human brain. It appears that certain regions of the brain can be associated with different functions, however, this mapping is not straight forward. Vision alone is thought to involve billions of neurons and more than 300 separate but interrelated pathways in the cerebral cortex [100]. Identifying which pathways are associated with certain tasks is not simple. In addition to the complexity of the system, there are also moral factors to consider when designing methods to obtain information. With so many unanswered questions it seems essential that extensive research takes place. New research may revolutionize not only medicine, but also inspire a new generation of computing based on parallel computation and distributed memory architectures, advancing the current neural network paradigm.

Stimuli cause action potentials to spread across relatively large distances within the brain, even across several quadrants. As such it is hard to connect certain regions to specific computational tasks. The myth that we only use 10% of our brains is still widely accepted, however it has been categorically disproved². The fact that such a myth remains popular highlights the failings of brain research to provide a clear interpretation of how the brain actually works. It is unclear what level of observation is required to adequately model functionality. Is a consideration of regions adequate, or do the interactions played at a neuronal level have wider implications to the behaviour of the system as a whole? Alivisatos et al suggests that research should focus on uncovering the emergent behaviours of the brain that arises from complex dynamic interactions [5]. The focus should be on identifying dynamical attractors rather than recording the receptive field responses of individual cells. This means that classical approaches such as the use of electrodes to monitor electrical activity are unlikely to yield much useful information. The problems facing brain research are wide ranging, however, the development of new techniques seems to present the opportunity to begin to answer some fundamental questions. There is significant activity within the field which may give us cause for optimism. Our ability to observe the brain at finer scales is improving as new techniques are developed. For instance, images of up to 80% of all neurons at a single-cell resolution can now be captured and used to compile an evolving image of brain activity [4].

Brain research was initially advanced through the efforts of individuals or small groups of researchers focussing on highly specific areas. In this environment extensive quantities of data was obtained. Despite this, utilizing the data to build a global picture remained beyond the scope of each individual project [100]. The task of integrating findings and deducing a global model of brain

²This myth is generally attributed to American psychologist W. James [73] and was even popularized by Albert Einstein when explaining his intellect. In reality we utilize almost all of our brain most of the time. Consider for instance that the brain accounts for roughly 3% of our body weight and yet consumes approximately 20 % of our energy [17]

activity would require the manipulation of large databases of information. Hence, expertise from outside neuroscience, including computer science, mathematics and physics are all called upon. In 2013, the Brain Research through Advancing Innovative Neurotechnologies Initiative (BRAIN Initiative), also referred to as the Brain Activity Map Project, commenced. Subsuming the National Neural Circuitry Database and receiving over \$300 million per year for ten years in funding, it is now tasked with overseeing the directed study of the brain in the US.

“By accelerating the development and application of innovative technologies, researchers will be able to produce a revolutionary new dynamic picture of the brain that, for the first time, shows how individual cells and complex neural circuits interact in both time and space.” [155].

This US project is replicated in Europe by an independent research framework referred to as the ‘The Human Brain Project’ (HBP). Set up by the European Commission in 2013, and now involving 112 partners in 24 countries across Europe, it shares many of the aims of the BRAIN Initiative. Following a pilot study [97] the remit of the HBP was established as:

“The Human Brain Project should lay the technical foundations for a new model of ICT-based brain research, driving integration between data and knowledge from different disciplines, and catalysing a community effort to achieve a new understanding of the brain, new treatments for brain disease and new brain-like computing technologies.” [1]

Despite the intentions of the HBP to catalyse the scientific community, the project met with some early scepticism, prompting a petition in 2014 to the European Commission to disband portions of the project. This discontent pertained to the €1.2 billion price-tag that was felt could be better used funding investigator-driven neuroscience grants. It was also claimed that the project took a too narrow approach. The view of Peter Dayan, director of the Gatsby Computational Neuroscience Unit at University College London is particularly damning:

“The project was sold on the idea of building an infrastructure to simulate the human brain. That simply is a nonsense. We can’t even simulate a nematode”. “We don’t know what questions to ask to build something like a simulation of the human brain. It’s not a sensible goal, it never was a sensible goal.” [160].

Despite these concerns, both the European and American projects are now in full operation and have begun to publish their findings. The HBP published 56 research papers in its first year.

1.1.3 AI without the Limitations of Biology

With biology not yet able to yield sufficient evidence to definitively define the way we learn, store, and process information, it does not seem unreasonable to explore other methods of modelling learning. Certainly there should be alternative methods to realize our aims. Consider the pioneering engineers of the 1900’s attempting to create a ‘flying machine’. Nature would indicate that such a machine behave like a bird, the success of the Wright Brothers alternative strategy serves to remind us that there is often another way to approach a problem. If we neglect the requirement for a neural-like architecture, and instead consider the composition of our learning system

to be irrelevant to our goal, then the approaches that may be taken to the problem are diverse. AI may be considered from a human behaviour perspective. Such an approach would focus on selecting a response comparable with a human, rather than fixating on the internal workings of the intelligent system. This requires an intimate knowledge of the way a human would respond in certain circumstances. There is therefore an emphasis on the behavioural and social sciences when considering the performance of a system. Alan Turing (1912-1954) proposed the Turing Test [154] as a means to gauge the intelligence of an artificial system. A system that passes such a test would be indiscernible from a human when interrogated. This would require natural language processing, knowledge representation, automated reasoning, machine learning, computer vision and robotics. Such a system is still a distance from being realized, however, certain requirements have been accomplished in isolation. A review of 50 years worth of accolades in this area is presented in [140]. Recently the AI platform ‘Cleverbot’ convinced 59.3% of users at a technique festival at the Indian Institute of Technology, Guwahati, that it was in fact human [6].

We may alternatively pursue a rational approach. Rational AI does not compare features to human characteristics, rather it searches for optimal solutions. This removes a number of obstacles and constraints. As should be apparent, there appears to be no discernible route to create an intelligent system, biology offers some interesting insights, but at our current level of engineering expertise and biological understanding the brain cannot be replicated. It is also evident that evolution does not guarantee an optimal solution to a problem. Systems may be devised with superior processing power via radically different routines to those observed in nature.

In what follows we shall review several approaches to AI. The reader should note that although several begin from a biological premise (neural networks, genetic algorithms), many biological concepts are either omitted or ignored in the search for optimal performance. It may be considered that biology is a useful guide until its application becomes a hindrance. Touching upon a few important models of AI we consider their shortcomings. We shall later introduce a new paradigm of learning that has some important characteristics which are discernible from many of the ideas within the AI field. We hope to demonstrate how this method may be developed to achieve a new AI that does not suffer from the same limitations as many of the more classical ideologies. This new concept takes the guise of a non-linear dynamical system that classifies stimuli without the need for supervision and without a distinction between training phase and standard operation. A rigid unit architecture is abandoned in preference of a flexible potential. This system presents the opportunity to circumvent issues such as the ‘curse of dimensionality’ and the requirement for microchips with connectivity and capacity comparable to the brain.

In the proceeding overview the models discussed have been chosen as it is felt they provide an insight into some of the more notable, relevant contributions to the field. That is not to say that alternative methods have not had a greater impact, but rather that the topic is vast and cannot be covered in intricate detail here. Although we draw attention to back-propagation [142], Hopfield’s model [52], genetic algorithms [110] and decision tree learning [139] and also touch upon extreme learning machines [66] and association rule learning [46], methods such as inductive logic programming [114], support vector machines [31], clustering [87] and bayesian networks [117] should not be ignored from a overview of the AI family.

1.2 Prominent AI Research Areas

1.2.1 Neural Network's

Artificial neural networks (nn)'s are mathematical models of brain activity that are utilized within AI for tasks such as pattern recognition and machine learning. What differentiates nn's from standard rule based programming is the highly parallelized structure of the system. The architecture is designed to emulate the structure of the human brain, allowing nn's to perform tasks such as computer vision and speech recognition for which traditional linear computing is ill equipped. Neurons are represented by artificial units within the network whilst their interactions are controlled by synaptic weights. There exists a plethora of nn models with many variations in network topology, learning rules (synaptic weight updating), unit activation functions, and stimuli/response representations. Despite this diversity, most nn's share some principal constituents [142]:

- Parallel, distributed processing structure.
- Processing elements that possess a local memory and are capable of localized information processing operations.
- Processing elements that are interconnected, each element has a single output which branches into many collateral connections.
- The output of each processing element can be of any mathematical type desired.
- All processing in an element depends only on the current input signal and values stored in the elements local memory.

Computation within an nn happens at the localized scale of individual units. Figure 1.1 demonstrates a typical computational unit j . Each unit receives a number of activity inputs from either the environment or other units within the network. These inputs form a vector $\mathbf{A} = (a_1, \dots, a_n)$. Associated with each input is an adjustable variable referred to as a weight. The ensemble of weights w_{1j}, \dots, w_{nj} comprise a vector \mathbf{W}_j . The weight w_{ij} governs the connection from unit i to unit j . Learning within a nn is typically achieved by adjusting these weights in accordance with a learning rule. A threshold function θ_j may also be modulated by a weight w_{0j} . Generally the activity of the unit b_j is determined by some function of the dot product of \mathbf{A} with \mathbf{W}_j . This function is commonly referred to as a threshold or activation function. Equation (1.1) provides a typical example.

$$b_j = f(\mathbf{A} \cdot \mathbf{W}_j - W_{0j}\theta_j) = f\left(\sum_{i=1}^n a_i w_{ij} - w_{0j}\theta_j\right) \quad (1.1)$$

Activation functions used for traditional nn models include linear, ramp, step and sigmoid functions. These choices are made mainly for convenience and often have few implications on the performance of the network [141]. However, one exception is the step function. In this circumstance a unit's output is discrete and typically binary, the number of states that the system can represent is therefore restricted. We shall later consider an important model that utilizes this activation function - the Hopfield network.

Research on network topology suggests that the choice of architecture is highly situation specific. There appears to be no general purpose nn topology that works well in all environments. Barring a few exceptional cases, most studied networks contain some form of layering [50]. Layers often consists of input neurons (these receive stimuli from the environment), hidden neurons (that have no contact with the outside world but play an important role in information processing) and output

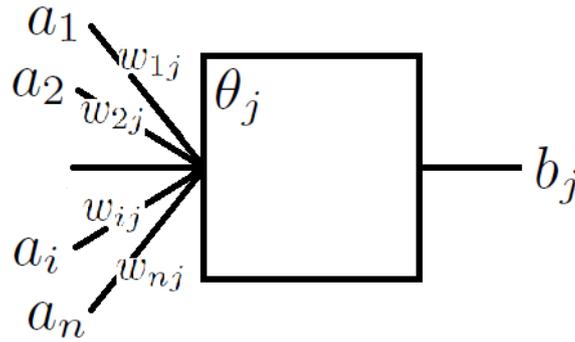


Figure 1.1: The Artificial neuron j which receives inputs a_1, \dots, a_n . These inputs are modulated by the set of weights $\omega_{1j}, \dots, \omega_{nj}$ and compared to the threshold θ_j . The output b_j of the unit is determined by an activation rule such as that given in equation (1.1).

neurons (which elicit the response of the system). Unlike the input and output layers where the number of units is defined by the constraints of the stimuli/responses, the hidden layer may be composed of any number of units. It is apparent that there exists an optimal number for a given task. Too few or too many and the system's performance will suffer, nevertheless, networks tend to work well over a range. As there is currently no definitive rule dictating how many hidden units a given network requires, the size of the hidden layer is typically determined via a trial and error approach. This may be effective when the system operates in a supervised fashion and responses can be compared to desired results, but such an approach may fail for unsupervised learning. One of the characteristics of 'deep neural networks' [120], a class of nn that has become increasingly studied, is the high number of layers. Networks containing ten hidden layers are not uncommon. Their hierarchical concept structure allows such networks to perform better on many tasks in comparison to shallow nn's [120].

In addition to layering, network topology also includes the connections between units. Connections may be intra-field, inter-field or recurrent. An intra-field connection joins two units in the same layer whilst inter-field connections connect units in different layers. These connections may be feed-forward or feed-back. A feed-forward signal only propagates in one direction whereas feed-back signals are bi-directional. Recurrent connections loop and connect a unit back to itself. These topological considerations can have a significant impact on the performance of the system and the types of task a network can be used for. From a biological perspective, action-potentials only propagate in one direction, however, certain tasks applied to artificial nn's cannot be achieved with feed-forward connections alone.

Learning in a nn is defined as any change made to the set of weights \mathbf{W} . Teaching strategies include reinforcement learning [12], stochastic learning, hebbian learning [58], back-propagation [138], competitive learning [83] and cooperative Learning. Furthermore, some nn's may have hard-wired weights. This is the case for finite state automata such as those constructed to represent the semantics of language. In what follows we shall consider a couple of the more eminent nn models. We shall describe their key features and discuss some of the fundamental drawbacks of each architecture.

Back-Propagation Multi-layered Feed Forward Network

W. McCulloch and W. Pitts defined the first artificial neuron in 1943 [104]. This inspired a cascade of research into nn theory. New ideas such as hebbian learning emerged that stimulated models

such as the ‘Adaline’ (B. Widrow 1962) and ‘Perceptron’ (F. Rosenblatt 1962). Unfortunately, nn research experienced a exodus of researchers and a large cut in funding after 1969 when Minsky and Papert published their book ‘Perceptrons’ [109]. Here they proved that although Perceptrons could learn anything that they could represent, their ability to represent information was severely restricted. This exodus was exacerbated in 1973 by the publishing of the ‘Lighthill Report’ [92] which declared the failings of the paradigm. This disappointment was overcome by ‘Back-Propagation’, a method first conceived by Bryson and Ho [25], but only popularized in the 1980’s by a number of researchers including Werbos [162], Rumelhart, Hinton and Williams [138]. Since then nn’s have regained popularity and lie at the forefront of AI research. The creation of the back-propagation algorithm, still widely utilized today, can be considered a pivotal development in the history of the field.

Back-propagation is commonly employed to train multi-layered feed forward networks, such as the very simple example shown in figure 1.2. Learning is achieved by searching the space of all weights for the minimum of an error function [135]. This typically occurs off-line, meaning that there is a dedicated learning/training phase. During this training phase the network is provided with training patterns. It compares its performance to the desired output and adjusts its weights to better express the function that the network is intended to replicate. This separation of learning and utilization may be considered a primary drawback of this method.

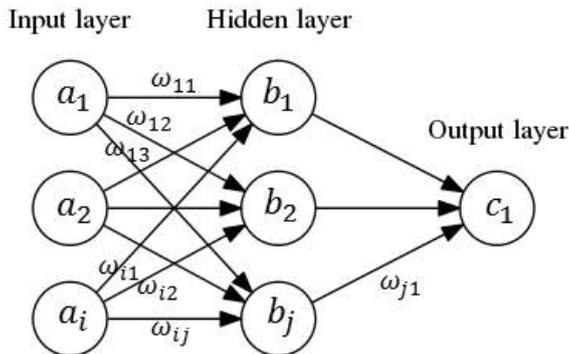


Figure 1.2: An example of a neural network structure that may employ a back-propagation technique to modulate the connection strengths ω_{ij} between units. The hidden layer allows the system to process more complex functions than a simple 2-layer network. Although we have only demonstrated a single hidden layer, such systems may possess many. Each unit b_j of the hidden layer receives an input from every unit of the input layer a_i . The outputs of the hidden layer all feed to a single output layer unit c_1 . Note that this is a simple case, NN’s typically include many more input, hidden and output units.

To explain how back-propagation may be applied we must first define the training set, hence let us consider a set of p input-output vectors $(\mathbf{I}_1, \mathbf{D}_1), \dots, (\mathbf{I}_p, \mathbf{D}_p)$. Here \mathbf{I}_i is a vector defining the state of the input units and \mathbf{D}_i is a vector defining the desired state of the output units. Firstly let us consider just 1 of these training patterns $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$. The subscript in this case refers to a unit in the input layer (N is the number of units in the layer). $\mathbf{D} = \{D_1, D_2, \dots, D_L\}$ refers to the desired state of each unit in the output layer (L is the number of units in the output layer). Allowing our input to propagate through the network in accordance with randomly chosen weights and a prescribed activation function we obtain an output $\mathbf{C} = \{C_1, C_2, \dots, C_L\}$. The aim of back-propagation is to minimize the error between the outputs \mathbf{C} and desired output \mathbf{D} for the entire training set. The choice of activation function may vary although it is common to apply a sigmoid

function. As we utilize a gradient descent technique, our only restriction is that the function must be smooth and differentiable. Continuing to consider only the single training pattern, we can compute the error in the output layer by taking:

$$E_j = C_j (1 - C_j) (D_j - C_j) \quad (1.2)$$

Here E_j is the error of the j^{th} unit, where $j = 1, 2, \dots, L$. Each unit in the output layer will have an assigned level of error. This error can be used to update the weight w_{ij} . The weight w_{ij} connects the unit i of the hidden layer to the unit j of the output layer. We shall denote the updated weight as w_{ij}^* .

$$w_{ij}^* = w_{ij} + \eta E_j B_i \quad (1.3)$$

Here B_i is the output of the unit i within the hidden layer. We can define the state of the hidden layer as $\mathbf{B} = \{B_1, B_2, \dots, B_M\}$ where M is the number of units comprising the layer. Also note that η is a learning rate parameter. We may now consider the error of the units comprising the hidden layer. The error on the output units can be seen to propagate backwards through the network with the portion of the error attributed to the former layer being prescribed by the weights of the system. As such, the error in a hidden layer unit, which we shall define as H_j is given by:

$$H_j = B_j (1 - B_j) \sum_{i=1}^L E_i w_{ji}^* \quad (1.4)$$

This is computed for each unit $j = 1, \dots, M$ of the hidden layer. The weight w_{ji}^* refers to the updated weight between the hidden layer unit and the output unit that it feeds (the output unit has an error of E_i). The sum over all L infers that the unit feeds forward to all output units but this does not necessarily have to be the case. The network topology dictates which units are connected. The error in the hidden units is utilized to update the error in the weights between input layer and hidden layer. In complex networks featuring back propagation, there may be many layers of units; the same principal of error flowing backwards through the system can be applied. In this simplified case of a 3-layer network, the new weights between input and hidden layer are given by:

$$w_{ij}^* = w_{ij} + \eta H_j I_i \quad (1.5)$$

As before, w_{ij}^* is the updated weight. This time, i is a unit within the input layer and j is a unit in the hidden layer. Updates should be applied for all $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$ as dictated by the network topology. If new outputs are computed using this same training pattern and the new system weights, then the error value will be reduced. This process should be applied for each training pattern $(\mathbf{I}_k, \mathbf{D}_k)$ in turn. This process should be repeated until the overall error is reduced to a suitable level.

The Back propagation algorithm reinigorated the nn field as it presented a means to represent information that is not linearly separable [137]. We must, however, consider the draw-backs. Two of the main limitations of this system are the requirement for learning to be supervised and the problem of over training (there are a small branch of back-propagating systems that learn in an unsupervised manner, referred to as auto encoders, however their applications within AI are limited). Supervision refers to the manipulation of unit weights by comparing outputs to

prescribed training examples. We have previously discussed the problems with supervised learning and explained how such a process is mostly redundant in human learners. We have not however touched on the issue of over-training. Over training relates to the number of times that the learning algorithm is iterated. Its influence can impair the performance of nn models and can be difficult to guard against without significant intervention. Too few iterations and the system will not perform well at recognizing known inputs, too many and the system becomes over trained. Over training means that the system will not be able to recognize stimuli in the presence of small quantities of noise. This is a problem as we desire the system to be robust. To overcome this, a measure of the error can be compared to a validation set of data; this is in contrast to ceasing training at a finite level of error. A validation set comprises the same inputs as the training set, plus a small amount of noise: we therefore require the outputs of the nn to remain the same. Comparison between inputs/outputs and a validation set ensures that the network does not become over-specialized. Generating this validation set is not a simple task and requires a high degree of human intervention. It is difficult and subjective to determine an adequate level of noise. Although the system becomes resilient, it also becomes more dependent on the training method. Further, the training phase may be significantly extended. An additional problem relates to the local minima of the error function. Ideally we would like to identify the global minima of the error function and find the set of weights that correspond to this. Unfortunately, gradient descent only guarantees that we find a local minimum. This local minimum may be a long way from the global minimum in the weight space resulting in sub-optimal performance. A typical approach to overcome this is to randomize weights after a prescribed number of iterations if an appropriate error margin is not achieved. Such a procedure still fails to guarantee that the global minimum is located, however, it does prevent the system remaining trapped in an undesirable local well. This means that the time taken for an appropriate set of weights to be located is undefined and can be extensive. Systems employing this methodology can therefore suffer from slow convergence times.

Extreme Learning Machines

Variants of extreme learning machines (ELM) [65, 67] are currently perceived as candidates for attaining plausible AI. These networks typically consist of a single hidden layer encapsulated between input and output layers. The network topology closely resembles the single layer feed forward network (SLFN) shown in figure 1.2. ELM is therefore an alternative way to train a network instead of using back-propagation. In ELM systems the hidden layer is not required to be neuron-like. The key feature is that the weights and biases between the input and hidden layers are randomly assigned. Only the weights between hidden and output units are manipulated to train the system. In order to classify unfamiliar stimuli, information propagates forward through the network in the same way as for any other SLFN. Compared to traditional NN's, ELM's are remarkably efficient whilst retaining the ability to reach the almost optimal generalization bound of feed-forward neural networks. In [68], Huang et al shows that this method is able to train a network approximately 170 times faster than back-propagation, and 190 times faster than a support vector method, for certain training sets. ELM does not require a gradient approach or iterative process to manipulate the set of weights, this property is responsible for the considerable speed up in training. The state of a standard SLFN that has learnt N training sets may be defined as:

$$\sum_{i=1}^{\tilde{N}} \mathbf{B}_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{o}_j \text{ for } j = 1, \dots, N \quad (1.6)$$

Here $\mathbf{o}_j = [o_{j1}, o_{j2}, \dots, o_{jm}]^T$ is the output of the SLFN, where $1, \dots, m$ refers to the m nodes of the output layer and j corresponds to the training set. We may refer to the entire training set as: $(\mathbf{x}_i, \mathbf{t}_i)$, where $i = 1, \dots, N$. For the training pair i , $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ is the input to the input layer which consists of n nodes and $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ is the target output of the output layer consisting of m nodes. The bias of each of the hidden units $i = 1, \dots, \tilde{N}$ is described by b_i whilst the weights between inputs and a hidden node i are described by $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$. Likewise, the weights between the hidden nodes $i = 1, \dots, \tilde{N}$ and the m output nodes are described by $\mathbf{B}_i = [B_{i1}, B_{i2}, \dots, B_{im}]^T$. The function $g()$ is the activation function and can be any infinitely differentiable function.

A network that has been trained to represent a set of sample input-output relations with zero error is characterized by the relation $\|\sum_{j=1}^N \mathbf{o}_j - \mathbf{t}_j\| = 0$, the state of the output units match the state of the desired outputs for every sample input. This means that there must be \mathbf{w}_i , \mathbf{B}_i and b_i such that:

$$\sum_{j=1}^{\tilde{N}} \mathbf{B}_i \mathbf{g}_i(\mathbf{x}_j) = \mathbf{t}_j \text{ for } j = 1, \dots, N \quad (1.7)$$

This may be written in the form $\mathbf{H}\mathbf{B} = \mathbf{T}$ where

$$\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \quad (1.8)$$

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1^T \\ \vdots \\ \mathbf{B}_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m}$$

For a back-propagation learning model the hidden layer output matrix \mathbf{H} is tuned to minimize the cost function that compares the state of the network to a desired state. This is achieved via iteratively adjusting the weights of the system via gradient descent. For ELM learning, the hidden layer output matrix is randomly defined and fixed. Learning is only associated with the adjustment of the matrix \mathbf{B} and so there is only one layer of tunable parameters. To correctly train the network the least square solutions for the matrix \mathbf{B} can be shown to correspond to:

$$\mathbf{B} = \mathbf{H}^\dagger \mathbf{T} \quad (1.9)$$

Here \mathbf{H}^\dagger is the Moore-Penrose generalized inverse matrix of \mathbf{H} [11]. This can be calculated via several methods including orthogonal projection and single value decomposition [68]. These solutions are not iterative and so training of the network occurs in a single time step. The principal aim of the ELM model is to limit human intervention whilst maximizing learning speed and accuracy [64]. We would argue that the requirement for training sets is contrary to this objective but do recognize that learning rates are much faster than for networks taught by back-propagation. Traditional feed-forward NN's suffer from slow learning rates due to the requirements of gradient descent techniques and iterative processes. By avoiding these algorithms ELM represents a signif-

icantly faster approach [66]. We should also point out that utilizing a random matrix \mathbf{H} allows us to use aspects of the same memory for very different learning sets. We may consider the modelling of two very different functions. For traditional SLFN we would require two networks, ELM allows us to use the same random matrix \mathbf{H} for both systems and so significantly reduces the amount of memory utilized by the system.

A primary drawback of the ELM method is again the requirement for training to be isolated from the subsequent application of the system. Further to this is the limitation on the number of training examples/hidden neurons. [68] shows that in order to guarantee effective training, the number of hidden neurons must be greater than the number of distinct training samples. This requirement is arguably less restrictive than the neuron population requirements for some other neural networks (i.e Hopfield model [63]), but may still present an engineering problem. In order to train a ELM model all sample sets must be stored in memory and utilized in a single step to create the weight matrix \mathbf{B} , this is not an issue when the sample set is relatively small, but can cause major problems when trying to learn a vast amount of complex information. Where larger training sets are necessary, alternative AI architectures such as Deep neural nets are preferred as they are able to assimilate samples iteratively, rather than composing the weight matrix in a single calculation. Due to the random features of the hidden layer, larger network sizes are required to match the performance of networks with tuned parameters. This larger size results in longer run time during the testing phase. Comparing the classification of inputs between a back-propagation taught system and a ELM system, the back-propagation system will typically compute associated outputs at a much faster rate. Finally, we should highlight that there will always be a degree of uncertainty about the operations occurring in an ELM network, as a significant proportion of the connections are randomly defined. This situation seems far from optimal, as we can never be certain that the system is evenly remotely close to offering the simplest solution to a problem.

The Hopfield Neural Network

The Hopfield model [63] refers to an associative memory model where the updating of synaptic weights is performed asynchronously in accordance with the Hebb rule [58]. In contrast to many nn models, the units of the system are not arranged in layers. Instead, each unit is coupled to every other unit and the weights between them are symmetric: by this we mean $\omega_{ij} = \omega_{ji}$. Computational units are based on the classical McCulloch-Pitts neuron [104], with a step-function representing the activation of the unit. As such, the state of each unit is binary. Despite the utilization of binary units, the nn is capable of storing a vast number of patterns. The number of different patterns that can be represented by 10 units is $2^{10} = 1024$.

The McCulloch-Pitts Neuron: The state of each unit is dependent upon the relevant synaptic weights and the inputs from connecting units. The classical McCulloch-Pitts neuron also includes a threshold parameter. In the original definition a unit i 's activation is described by equation (1.10). The use of the Heaviside step function ensures the state of the neuron is either 0 or 1. However, it is computationally more convenient to use a sigma function, this is shown in equation (1.11).

$$n_i(t+1) = \Theta \left(\sum_j w_{ij} n_j(t) - \mu_i \right) \quad (1.10)$$

$$\Theta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$S_i = \text{sgn} \left(\sum_j w_{ij} S_j - \theta_i \right) \quad (1.11)$$

$$\text{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0; \\ -1 & \text{if } x < 0; \end{cases}$$

Parallels may be drawn between the use of binary units and the actual ‘all or none’ rule that appears to govern neuron firing in biological systems.

Associative Memory: Memory can be represented as a set of stored patterns ξ_i^μ , where i refers to the unit of the network and μ is the identifier of the pattern. A stored pattern in this case is simply a sequence of 1’s and -1’s for each of the units $i = 1, \dots, N$. If we consider a test pattern, then for an associate memory model the system would configure itself to the stored pattern that most closely resembles this input. In order to determine which stored pattern this is, we need to know the set of weights (w_{ij}) that will configure the network from its current state to the state closest to the test pattern. This is referred to as a content-addressable memory. Such a system will always retrieve the nearest pattern related to a stimuli. Issues arise when an input pattern does not closely match any stored pattern, the information then retrieved may bear little resemblance to the input. The configuration space that contains all possible states of the network is split into basins of attraction. The system converges to a stable attractors by updating units in accordance with equation (1.11). There are two plausible methods via which such an update can be applied; synchronously or asynchronously. For the synchronous case, all units are updated simultaneously at each time step. This would require a central clock, which is not biologically realistic and presents the possibility of timing errors. Alternatively, the units may be updated one at a time, selected randomly at each time step. Equivalently, each unit may independently update itself reliant on a prescribed probability distribution.

It can be shown that for an initial configuration close to a stored pattern, the state of the system will quickly relax to this stored state. For configurations further away a series of unit updates are required. If less than half of the units are in an unknown configuration, then they will be overwhelmed in the sum $h_i = \sum_j w_{ij} S_j$ used to update each unit i . Here S_j is the activation of the unit j that synapses to i . If greater than half the units are initially in an unknown configuration, then the unit being updated will not be updated correctly and the system will fall into a spurious state known as the reversed state.

Determining synaptic weights: The synaptic weights are determined during training via the Hebb rule, this takes the form:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \epsilon_i^\mu \epsilon_j^\mu \quad (1.12)$$

In this formulation ϵ_i^μ represents the i^{th} bit of the pattern μ and ϵ_j^μ represents the j^{th} bit. Recall that these take values 1 or -1 and that each bit is represented by a corresponding neuron from the ensemble N . This algorithm is applied iteratively during a distinct training phase. Iteration

continues over the ensemble of training patterns until performance for the entire set meets a desired level of accuracy. After this, synaptic weights are fixed and the network will evolve to the closest pattern when an input is presented.

Storage capacity of the Hopfield Model: A given pattern v will be stable (i.e units defining the pattern will not flip into the reverse state) if:

$$\text{sgn}(h_i^v) = \epsilon_i^v \quad (1.13)$$

Here h_i^v is the net input to unit i as a result of pattern v .

$$h_i^v \equiv \sum_j w_{ij} \epsilon_j^v = \frac{1}{N} \sum_j \sum_{\mu} \epsilon_i^{\mu} \epsilon_j^{\mu} \epsilon_j^v \quad (1.14)$$

This can be separated to present the input to the unit i in terms of the bit i of the pattern v and what is referred to as the crosstalk term.

$$h_i^v = \epsilon_i^v + \frac{1}{N} \sum_j \sum_{\mu \neq v} \epsilon_i^{\mu} \epsilon_j^{\mu} \epsilon_j^v \quad (1.15)$$

If the cross talk term is zero or less than 1, it cannot change the sign of h_i^v . If the system configuration pattern is stable then the system will remain there. Configurations close to the stable pattern will also relax to this pattern. The storage capacity of the network is determined by the crosstalk term. Let us define another function:

$$C_i^v \equiv -\epsilon_i^v \frac{1}{N} \sum_j \sum_{\mu \neq v} \epsilon_i^{\mu} \epsilon_j^{\mu} \epsilon_j^v \quad (1.16)$$

If the crosstalk term has the same sign as the desired term ϵ_i^v , then C_i^v will be negative; the crosstalk has no effect. If $C_i^v > 1$, the sign of h_i^v will be changed, hence the pattern v is unstable. The probability that any unit is unstable is described by $P_{error} = Prob(C_i^v > 1)$. As the number of patterns that we attempt to store increases, so too does the probability of an error. As such an acceptable error tolerance must be defined. With this in mind we can determine the storage capacity (p_{max}) of the network. The probability of an error (P_{error}) is related to a Gaussian function with a mean of zero and a variance $\sigma^2 = \sqrt{p/N}$. We find that:

$$P_{error} = \frac{1}{\sqrt{2\pi}\sigma} \int_1^{\infty} e^{-x^2/2\sigma^2} dx = \frac{1}{2} \left[1 - \text{erf} \left(1/\sqrt{2\sigma^2} \right) \right] = \frac{1}{2} \left[1 - \text{erf} \left(\sqrt{N/2p} \right) \right] \quad (1.17)$$

where

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du$$

Utilizing equation (1.17) the initial stability of a configuration can be determined for a chosen error tolerance. The capacity of the network for a given tolerance can therefore be inferred. It should be stressed that this is just the error in the initial configuration. If these units are updated to incorrect values then this can cause a cascade, resulting in the entire memory being unusable. It can be shown that such a cascade will occur if $p > 0.138N$ [60].

Limitations of the Hopfield Model In order for the network to function effectively, the system requires a relatively high degree of redundancy. Coupled with the curse of dimensionality, this means such models do not present an efficient method of storing large data sets. Regardless of error tolerance thresholds, spurious attractors may also plague the performance of such systems. We have already made reference to one type of spurious state, the reserved state. In addition, there are also mixed states and spin glass states. Mixed states are not the result of any single pattern; they reflect linear combinations of an odd number of patterns [167]. The mixed state scenario is easiest to observe from an energy function perspective. Defining an energy function H is one of the major break-throughs of the hopfield model, and is why weights must be defined symmetrically. The energy function is described by:

$$H = -\frac{1}{2} \sum_{ij} w_{ij} s_i s_j \quad (1.18)$$

Here s_i is the activation of the unit i , s_j is the activation of the unit j and w_{ij} is the corresponding weight between them. The energy function H always decreases or remains constant. In accordance with this function stored patterns are represented by local minima. One problem is that some local minima may not represent a stored, desired pattern, rather, it may be the result of a mixed state. Stochastic units present a solution to this problem, however, they also increase convergence times.

Beyond the nn approaches to AI lie a vast array of alternative strategies. Decision Tree learning and Genetic Algorithms are two of the more studied non-nn paradigms, whilst associative learning rules are perhaps the most widely utilized. We have tried to present a general overview of the nn field, however, in reality, the variability and scope of the ideas emerging from this research area are far too vast and eclectic to summarize effectively and contrast to our new learning paradigm. It is hoped that providing a general background on back-propagation and hebbian learning informs readers of the necessity for alternative strategies to meet the expectations of AI. Before we move on to developing these new strategies, let us consider the aforementioned non-nn approaches and consider why these methodologies may not yield solutions conducive to solving the problem of effective AI.

1.2.2 Decision Tree Learning

A selected response is rarely a function of a single decision making factor. We may consider a stimuli as an N-dimensional vector representing a multitude of relevant decision making considerations. Take for example the decision to get a haircut, the response is binary-yes/no, but leading to this there are multiple relevant attributes that must be considered, these may include: ‘length’, ‘Cost’, ‘Do I have money?’, ‘Distance to Barber Shop’, ‘Time of day’ etc, there may be 100’s of components for even the most simple of decisions.

As the name suggests, this learning algorithm can be visualized as a tree or flow diagram. Inputs to the system are vectors of attributes. The system separates these attributes into classes and subclasses, until all members of a class have the same outcome. An example structure for a boolean decision is shown in figure 1.3. To build such a system training knowledge must be exploited; as such supervised learning is typically required. There are some classes of decision tree that operate in an unsupervised or semi-supervised manner [78]; these are typically classifier systems that separate databases of information in relevant classes and subclasses. Both inputs and outputs can be continuous or discrete, however, a continuous approach introduces considerably more complexity.

Multiple decision tree algorithms exist, each with different functionality. Amongst the most notable are Iterative Dichotomiser 3 [130], C4.5 [131], CART (Classification And Regression Tree)[19] and CHAID (CHi-squared Automatic Interaction Detector)[79].

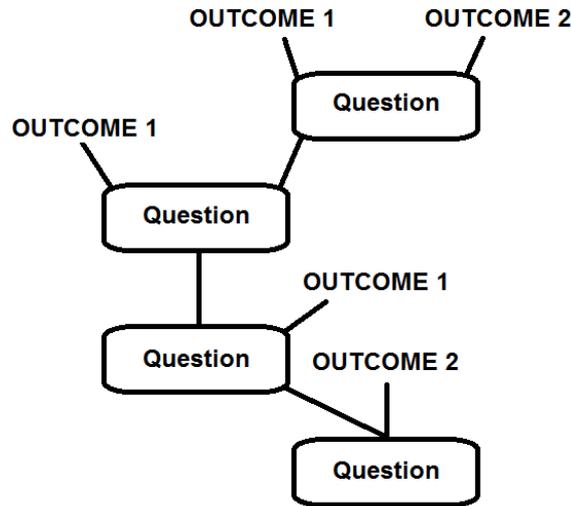


Figure 1.3: Decision Tree Learning for a boolean decision. The attributes of the vector input are divided into classes until all members of a class share a common outcome.

It can be considered that the aim of a decision tree learning system is to create a tree like architecture with outputs at its extremities. These are reached by answering a series of ‘questions’. This concept is demonstrated pictorially in figure 1.3. The system does not initially know which attributes should be prioritized when forming classes. As a broad flat tree is less computationally expensive, it is important to minimize the number of layers of reasoning. The system seeks to achieve this autonomously by implementing what is referred to as an importance function or greedy algorithm. This function calculates the entropy of the system and uses the information gain, calculated from a test on each attribute, to reduce the entropy. The attribute yielding the greatest reduction in entropy is then utilized to form classes. This approach is simple to interpret and can utilize and represent large sets of data. Unfortunately, it is not without drawbacks. Principal amongst these is the requirement for large sets of training data and the need for supervision. Over fitting can also cause problems where the system perceives an attribute as a suitable classifier when the relation is really an anomaly. Furthermore, over-fitting may result in an excessive level of complexity. This problem is partly overcome by a technique called decision tree pruning where statistical significance tests are applied to attributes to test their relevance. The greedy algorithm finds the locally-optimal solution, however, the combination of these local solutions is not guaranteed to prescribe the global optimum. As such, decisions may not be made in the most effective/shortest reasoning hierarchy. The structure of these systems is also poorly equipped to deal with certain classes of problem such as XOR. Although it is possible to represent XOR relationships, the size of the tree becomes prohibitively large [139].

1.2.3 Association Rule Learning

Association rule learning has strong applications in data mining and is primarily utilized to identify hidden patterns in large data sets [46]. The applications of such techniques are extensive and wide ranging, from studying the visiting patterns of tourists [158] to identifying patterns in

herbal formulae [90] and accessing the vulnerability of cities to earthquakes [133]. Association rule learning was initially proposed as a data mining method for making associations between supermarket product transactions. Such associations may be embodied by statements such as ‘90% of transactions that include bread and butter also include milk.’ [3]. Deriving such information is useful as it may be applied to improve store layouts or for selecting special offer discounts.

The learning system searches a database of transactions and makes associations between items. Taking into account the confidence in the association rule and the statistical significance of the measure, the system can follow instructions such as: ‘identify all rules that meet a prescribed criterion’, or, ‘find the rule with the highest confidence, that meets a prescribed criterion’. Omitted from association rule learning is the notion of feedback. Typically, outcome goals are assessed after a response - learners then adjust accordingly. An associative learning machine is dependent on the initial database of information defined. As such, such systems are highly supervised and are devoid of the flexibility to respond to short term stimuli changes.

1.2.4 Genetic Algorithms

A genetic algorithm (GA) [62] is a special type of evolutionary algorithm designed to demonstrate learning. Evolutionary algorithms apply an iterative, stochastic learning rule that utilizes a ‘survival of the fittest’ principle to search for optimized solutions within a finite population of retained individuals (information stored in memory) [143]. Within this framework one can also consider genetic programming, evolutionary strategies and evolutionary programming. The method employed by a GA is based on an analogy to biology and the coding of chromosomes with specific genes. In this context a gene can be seen as variable and the combination of multiple variables a chromosome.

In biology a gene is thought to define a single feature of a person, for instance, we may consider a gene indicating eye colour. The possible gene expressions are green, blue or brown. Another gene may represent hair colour and another may represent whether a person can roll their tongue. Combining these genes together forms a chromosome, which defines the overall features of the person. Clearly there are many different combinations that a chromosome can express. A GA model considers a population of these chromosomes as a type of memory. These chromosomes are initially randomly defined. They are then trained by determining fitness, triggering crossovers and instigating mutations. These are key ideas behind this ‘survival of the fittest’ methodology. Let us continue to avoid the rigours of GA methodology and consider what is meant by ‘survival of the fittest’ in evolutionary terms.

The concept of ‘Survival of the fittest’ as a driving force behind the advancement of a species was first postulated by Charles Darwin in 1859. Pivotal to this ideology is the concept of mutation, genetic crossover and survival. Within an animal or plant population, there exists individuals that possess genetic traits that provide an advantage over their counterparts. These traits may include things like longer beaks or more acute hearing. Qualities that are the result of small genetic mutations. What classifies an advantage depends on environmental factors. For instance, more acute hearing may be advantageous in an ecosystem with many predators. This individual would have a higher probability of survival and hence more opportunity to pass on its genetic traits to offspring. The genetics of any offspring share the crossover of genes from both its parents. This presents the possibility that advantageous traits are passed onto further generations. Over many generations environmental pressures coupled with genetic mutations cause the continual

development of a species, optimizing its features based on environmental stimuli. This process can be viewed as an optimization problem where the systems features are adjusted due to the application of a stimuli. We should point out that this is essentially a definition of learning. As such, the GA approach can be considered to restate Darwinian evolution in a different context.

In any evolutionary algorithm a representation scheme is selected to define the set of solutions that form a search space. In the case of GA's, a binary string is typically utilized. Solutions are initially created randomly and form an initial population. The fitness, crossover and mutation of members of the population are then iteratively compared to a stimuli, until a solution matching a termination criterion is discovered. This process is summarized in figure 1.4. The fitness function is a probabilistic measure that assigns each solution a probability of being selected for crossover. Fitness is assigned based on a measure of a solutions similarity to some prescribed target. Two parent solutions are selected from the population and a crossover of binary bits takes place. Of importance is the selection method used to determine parents. Many methods have been proposed in the literature [143, 110] to describe selection. These include, but are not limited to, roulette wheel, tournament and Boltzmann selection. The aim is to maintain a population size whilst favouring chromosomes that are more closely aligned to the target. Parents are typically selected stochastically, but preference is shown to those with higher fitness assignments. In this way, the system performs faster than a random search algorithm. The cross-over point for the exchange of bits between parents is also generally random. The new solution is considered to be an offspring. Due to the fitness criteria and random bit exchange there is a chance that this new solution is closer to optimal than the member of the previous generation. This offspring is also susceptible to mutation, this should be considered as a random flipping of a random number of bits on its string. This may have beneficial effects.

The use of binary units for GA's is advantageous as it captures a principle aspect of genetics known as the laws of Mendelian inheritance. The gene inherited by the offspring is not a blend of the two parents, instead each gene is inherited from one of the solutions alone. This hypothesis was first attributed to Gregor Mendel (1866), who studied the cross-fertilization of pea plants. He is now considered the founder of genetics. The representation of solutions to GA models are predominantly based on binary units, although octal, hexadecimal and real number encoding can all provide viable alternatives in certain situations. The flexibility of the model is further emphasised if we consider that the fitness function can feasibly be any relevant function. There also exists a range of methods to select parents from a population. These methods may be as simple as taking a random choice but most utilize a function of the fitness, weighting choices based on the selection pressures. The selection pressure should be considered as the degree to which better individuals are favoured. Where selection pressure is higher, we would therefore expect a faster convergence towards an acceptable solution. One of the more conventional selection methods employed by GAs is roulette wheel. This selects parents from the mating pool with a probability proportional to each individual's fitness.

Advantages and Limitations of GAs

The GA approach offers several advantages over alternative AI and conventional computing. The biological basis for the model aids in directing further research ideas as more is learnt regarding genetics. Furthermore, the relatively simple mathematical constructs allow multi-disciplinary use of the model. This aids in the generation of new ideas. Since GA's can be applied to any opti-

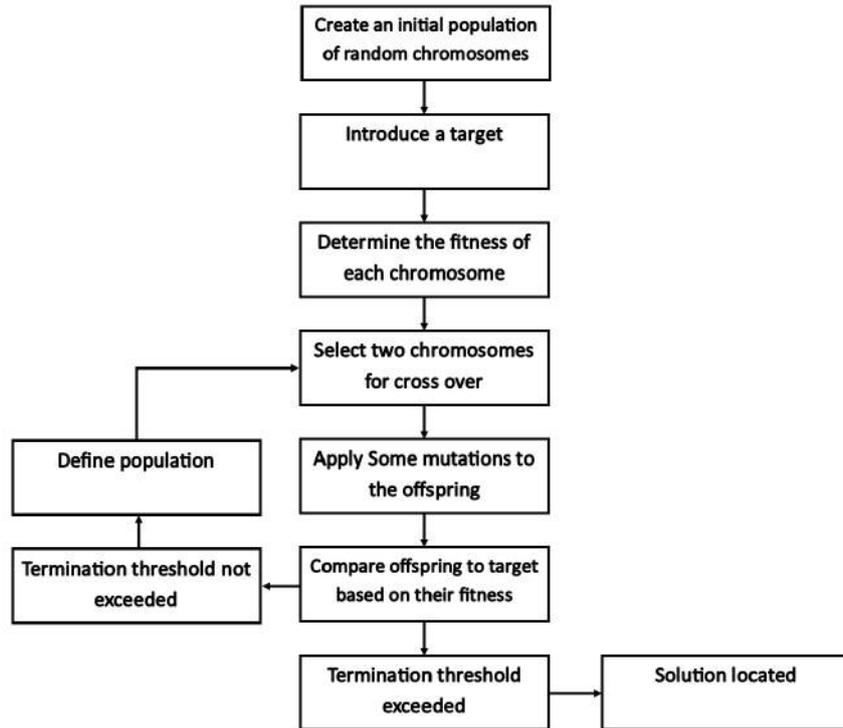


Figure 1.4: Demonstration of the stages involved in updating a Genetic algorithm to replicate a given target. Via this method, the GA learns to replicate a prescribed input. After learning the GA will then be able to select appropriate responses by comparing unknown stimuli to the chromosomes within the population and eliciting the relevant response.

mization problem that can be represented as binary strings, the applicability is also broad. GA approaches remain robust to noisy stimuli, a quality imperative in any AI architecture and eminent in nature. In addition, the concepts of GAs should not be isolated from the rest of AI, they can be incorporated with traditional optimization techniques and utilized to improve the performance of other paradigms. Seeing as several combinations of solutions can be computed simultaneously the speed of GA approaches can be significantly improved on distributed processing machines. The key advantage of a GA approach over traditional computing is that it is able to address problems that cannot be solved linearly. It is also able to interpret patterns within larger sets of data, well beyond what is comprehensible for a human.

Like most AI paradigms, GAs have their own share of shortcomings. We may consider that the GA approach to optimization is relatively general, it is therefore likely to be out performed by techniques tailored to specific problems. This downside may be reflected by both the speed and accuracy of a chosen algorithm. GAs search a restricted population of solutions that form a subclass of the search space. The search space in complex optimization problems can be too large for conventional computational search techniques such as stochastic hill climbing, random search, gradient descent and simulated annealing to effectively explore within a reasonable time-scale. GAs have the advantage that the restricted population they utilize evolves and can eventually represent any point in the search space. We should be wary of this, as if the initial population does not contain any members that closely resemble the target, the duration of the search can again be extensive. Iteration of the algorithm is generally continued until a certain amount of time is exceeded or a certain level of accuracy is achieved. Because a GA is stochastic, there is

no guarantee that a globally optimum solution will be identified. A GA can function as a type of memory when the population is trained to represent multiple targets. Unlike conventional memory, there is always the risk that information pertaining to previously encountered targets is going to be overwritten. This is less likely when targets are distinct and so the fitness of a trained solution is low in comparison to other solutions within the population. Where targets are similar there is a high likelihood that information that should be retained is destroyed. Furthermore, the stochastic nature of the algorithm always presents the chance that any portion of memory may be rewritten.

The reader should hopefully be convinced that despite the wealth of ideas within the AI framework, no plausible method to truly solve the problem has been developed with sufficient rigour to even begin to emulate the capabilities of a biological brain. The extensive research presented since the inception of AI in 1956 [139] has been fraught with exciting developments and failings. Many ideas emerging from AI research have been implemented into the technologies now commonplace in the modern world. Regardless of this, the cumulative achievements to date remain far short of the expectations of the field's pioneers [30]. With this in mind, the need for new ideas is indisputable. We cannot claim that the model we shall subsequently present offers a solution to all of the shortcomings of alternative AI's, however, viewed within the wider context of the field, its development may inspire future developments and potentially a new approach to stimulate advances.

1.3 ‘Memory Foam’ Approach to Unsupervised Learning’

The shortcomings of current methods to replicate the efficacy of a neural system provide an opportunity to re-evaluate the approach to the problem. It is apparent that even with the evolution of silicon chips, we are still a significant distance from implementing neural-like circuits with the same level of connectivity and parallelization as the human brain. Even if such a circuit were currently available, we still do not know how such a system should be composed and what laws governing interactions should be applied to produce an autonomous learning system. These issues may perturb some considerations of neural network and alternative AI approaches until adequate technologies can be developed. As discussed, replicating biology is not the only way to artificially realize learning. It may not even be the optimum way. A new paradigm is presented in [74] and developed in [99]. It is a consideration of this new paradigm that we shall concern ourselves with, developing supplementary ideas and exploring further the presented results.

The basic ideology of the ‘Memory Foam’ model is that it is not how the brain works, but what it achieves that should be the primary concern. As such, the realities of biology are neglected in favour of a more flexible approach that does not involve the restraints of neuron architecture. The simplest means to introduce this new model is to use an analogy to orthopaedic memory-foam:

Memory foam takes the shape of any object that is imprinted on its surface. This imprint remains even once the object has been removed. To aid this description, it is beneficial to consult the graphic shown in figure 1.5. Consider the impacts of a series of falling objects. At the places where they land, indentations are made. These objects are comparable to a stimulus, and the foam is ‘memory’. If objects are continuously dropped at different locations on the foam then multiple indentations are formed. The foam’s profile will continuously change to reflect the distribution

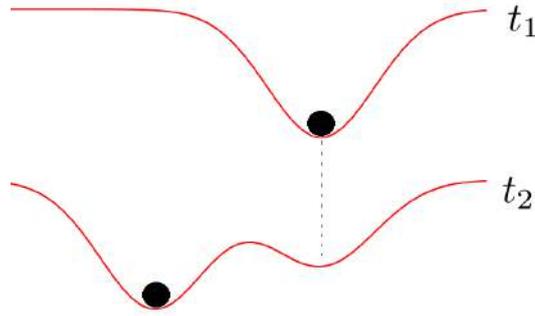


Figure 1.5: As objects fall onto the profile the shape continuously changes. As a depression caused by an impact remains beyond the removal of the stimuli we may associate deeper depressions with more prevalent stimuli. This concept can be seen as analogous to the shaping of orthopaedic memory foam. Information about the distribution of stimuli is retained by the shape of the profile. We may consider the foam to have learnt and retained information about its experiences.

of objects that have fallen. As impressions endure with time, the profile of the foam reflects the distribution of stimuli that have been applied.

The memory foam is analogous to a time-dependent evolving potential that is continuously shaped by a stimuli. Unlike in other learning systems, the stimuli are not subject to external control or manipulation. We should also add that there is no restriction on the number of stimuli that can be utilized. The formation of this potential is essentially a new method to visualize learning. Prevalent stimuli impose local minima on the potential, these minima act as attractors of the system and reflect classes of stored knowledge. Further complications may also be included to shape the potential such as the notion of forgetting. This may be described by the gradual disappearance of a local minimum.

In a typical learning model such as a neural network or a decision tree learner, learning is halted whilst the system is utilized to perform some operation. This may be to solve some optimization problem or assign a new, unseen stimulus to a class. This assignment will generally have no implication on the information that has already been learnt by the system and will not add to the system's knowledge. In contrast, this new model allows learning and recognition to occur simultaneously, a mechanism referred to as on-line learning. Recognition is deemed the assignment of an unfamiliar stimulus to a class. This is realized in [74] by taking the gradient of the potential. Learning and recognition are not separated into definable phases in the systems operation. The potential is allowed to continuously adjust to its environment, learning remaining completely unsupervised. Concurrently, the system utilizes the potential via a gradient approach to recognize new stimuli. Although the system has a maximal information storage capacity, the time dependent nature of the model ensures that any information that is lost can be re-learnt. The system will therefore train itself to any new environment. The flexibility and breadth of application of the system is therefore much greater than that of models with distinct learning phases. Despite this flexibility, the system should retain accuracy and may be utilized to represent precise information in appropriate contexts.

Before we begin to explore this model, we should first discuss the intended direction of development and highlight the tools that may achieve such ends. The behaviour of this new model is non-linear, nonetheless, the dynamics of the prototype system are relatively simple, dependent on basic gradient descent. Adjustment of the system to permit more complex, higher order attractors

should improve the viability of the model to real world AI tasks. It is therefore important to consider some of the primary elements of non-linear dynamics. We shall introduce these via the van der Pol oscillator, a paradigmatic non-linear system that has been extensively studied, and previously utilized to characterize the oscillations in neural systems. We shall look to utilize this system to develop the model postulated by Janson and Marsden.

1.4 Elements of Non-Linear Dynamics

Non-linear dynamical systems are described by time evolution equations that pertain to the state of the system. In contrast to linear systems, the variables are non-linear; this means that the output is not directly proportional to the input. Relevant dynamical variables include position, velocity, acceleration etc [61]. Dynamics is now a highly inter-disciplinary branch of science which has made important contributions across multiple subject areas (finance [29], cosmology [9], engineering [149], climate physics [35], biology [14], population modelling [13]). To date, the accomplishments of the field are extensive. Even with this in mind, its applications appear far from exhausted. Developments within the younger branch of physics: non-linear dynamics, promise to deliver yet more exciting and innovative contributions from this venerable field.

The dynamics of Janson and Marsden's gradient technique are relatively simple; the local minimum of an energy potential, surrounded by its basin of attraction, acts as a point attractor. Although the gradient approach meets the basic requirements of the learning system, utilizing the concepts of non-linearity presents an opportunity to introduce recognition in terms of higher-order oscillatory behaviour. Such behaviours are indicative of real neural systems, where limit-cycle oscillations, quasi-periodic motions or even chaotic trajectories have been observed [80]. Biology aside, considering various synchronization and bifurcation regimes introduces valuable complexity to our system. Such complexity may be utilized to reflect a hierarchical learning structure such as the formation of classes and subclasses of retained knowledge.

1.4.1 Differential Equations and Attractors in Phase Space

The time evolution of a dynamical system can be expressed in the language of differential equations (DE's). These constructs take the general form $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} = F(\mathbf{x}, t)$, where \mathbf{x} is a vector variable, t refers to time and $F()$ may be referred to as the velocity vector field. If the function $F()$ does not explicitly depend on time, then the system is referred to as autonomous. Where a DE is of order higher than first, the system can be transformed into a set of first order equations by introducing a suitable change of variables. This is advantageous as the solution of sets of first order DEs can be utilized to study the system in phase space (the dimensions of the phase space are equal to the number of dynamic variables required to define a solution to the DE). In this framework important features of the system can be observed and analysed. As an example, let us consider a DE describing simple harmonic motion: $\ddot{x} + \omega^2 x = 0$. Here x is displacement and ω is a constant referred to as the angular frequency. Introducing the change of variable $y = \dot{x}$ the system can be described by a pair of equations $\dot{x} = y$, $\dot{y} = -\omega^2 x$. Trajectories of the system can then be observed in a two-dimensional phase space with axis $x, \dot{x} = y$. Non-autonomous systems can be treated in a similar way, consider the forced and damped harmonic oscillator: $m\ddot{x} + b\dot{x} + kx = F \cos(t)$. Here m is mass, b is a damping parameter and k is a spring constant. An external oscillating force $F \cos(t)$, with amplitude F and period t perturbs the natural dynamics of the system. Applying

the change of variables $\dot{x} = y$ and $t = z$. $\dot{x} = y$, $\dot{y} = -\frac{1}{m}(by + kx - F\cos z)$ and $\dot{z} = 1$. This trick ensures that any higher order DE can be written as sets of first order equations.

The dimensionality of the phase space is important when considering the behaviour of the system. Uniqueness requires that within a one-dimensional phase space trajectories must either remain stationary, converge to some fixed point, or move away to infinity. The Poincaré - Bendixson theorem [152] dictates that phase space trajectories may never cross, as such quasi-periodicity and chaotic trajectories are not possible on the plane, only emerging in a space with three dimensions or greater. Oscillatory dynamics, known as limit cycles are however possible.

The methods and tools utilized within the field of dynamical systems allow complex behaviours to be analysed and predicted. We shall introduce many of these tools whilst considering the behaviour of the now paradigmatic system, the van der Pol (VDP) oscillator. In Chapter 4 we shall suggest methods of incorporating limit cycles with our learning system.

1.4.2 The van der Pol Oscillator

Oscillatory dynamics are ubiquitous within living systems. The regular beat of the heart [156], gaseous exchange in the lungs [39], sleep cycles [151], hormone level fluctuations [136], neural excitations [51] and the vast array of circadian rhythms occurring within our bodies, all reflect some kind of oscillatory behaviour. The requirements of homoeostasis prescribe that these oscillations must occur within a finite bandwidth. Such oscillations are therefore insensitive to small perturbations and return to a stable dynamic trajectory at the cessation of any external stimuli. Such a trajectory may be visualized in phase space as a stable limit cycle.

Oscillators capable of demonstrating limit cycle dynamics all share some critical similarities. These oscillators are often referred to as self-sustained, this name is apt as the repetitive motions do not damp. Furthermore as $t \rightarrow \infty$ the shape, amplitude and frequency of oscillation are typically insensitive to any change in initial conditions. This type of behaviour is only possible when there exists an interplay between energy dissipation and power input to the system, with one or both of these terms being non-linear. There exists an expansive array of models that meet the requirements for self-sustained oscillation, particularly eminent is the VDP oscillator.

The DE describing the VDP oscillator normally takes the form:

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} + \omega^2 x = 0 \quad (1.19)$$

It may however be referred to in the alternative form:

$$\ddot{x} - (\varepsilon - x^2)\dot{x} + \omega^2 x = 0 \quad (1.20)$$

It can be shown that equation (1.20) is essentially equivalent to equation (1.19). This is achieved by introducing the change of variable $y = \sqrt{\varepsilon}x$. The derivatives of y are then:

$$\begin{aligned} x &= \frac{1}{\sqrt{\varepsilon}}y & \dot{x} &= \frac{1}{\sqrt{\varepsilon}}\dot{y} \\ \ddot{x} &= \frac{1}{\sqrt{\varepsilon}}\ddot{y} & x^2 &= \frac{1}{\varepsilon}y^2 \end{aligned}$$

$$\begin{aligned}
\ddot{x} - \varepsilon(1 - x^2)\dot{x} + \omega^2 x &= \frac{1}{\sqrt{\varepsilon}}\ddot{y} - \varepsilon\left(1 - \frac{1}{\varepsilon}y^2\right)\frac{1}{\sqrt{\varepsilon}}\dot{y} + \frac{\omega^2}{\sqrt{\varepsilon}}y \\
&= \frac{1}{\sqrt{\varepsilon}}\ddot{y} - (\varepsilon - y^2)\frac{1}{\sqrt{\varepsilon}}\dot{y} + \frac{\omega^2}{\sqrt{\varepsilon}}y
\end{aligned} \tag{1.21}$$

Multiplying through by the constant $\sqrt{\varepsilon}$ we show that the form of equation (1.19) is equivalent to equation (1.20). The topologies of the two dynamical systems will therefore be equivalent, however certain bifurcations may occur for differing parameter values. The system may also be described by a Liénard type formulation [152]. Applying the Liénard transform $y = x - \frac{x^3}{3} - \frac{\dot{x}}{\varepsilon}$ the van der Pol oscillator can be written in the linearised form:

$$\begin{aligned}
\frac{dx}{dt} &= \varepsilon\left(x - \frac{x^3}{3} - y\right) \\
\frac{dy}{dt} &= \frac{x}{\varepsilon}
\end{aligned} \tag{1.22}$$

Each formulation is essentially equivalent, however, certain forms may be preferred in certain circumstances. For the sake of clarity, all references to the VDP oscillator henceforth refer to the system as written in equation (1.19). As this system will prove intrinsic to many of the ideas presented in later areas of this dissertation, and also serves as a suitable example on which to apply some of the basic techniques of non-linear analysis, we shall subsequently present an overview. We shall locate the fixed points of the system, study the stability of the system's attractors, and identify system bifurcations as parameter values are adjusted. Inferences and analytical results shall be compared to numerical simulations to help illustrate the systems various regimes.

Fixed Point Analysis

The VDP equation (1.19) is a second order, non-linear, differential equation. It can be expressed as a pair of first order equations by defining the variable change $y = \dot{x}$. This yields:

$$\begin{aligned}
\dot{x} &= f_1(x, y) = y \\
\dot{y} &= f_2(x, y) = \varepsilon(1 - x^2)y - \omega^2 x
\end{aligned} \tag{1.23}$$

The fixed points of this system are then identified by finding solutions where $\dot{x} = 0$, $\dot{y} = 0$. These points may also be referred to as equilibrium points. If the system's initial state is set to one of these points, then it will remain stationary for all time. It is easy to see that $y = 0$, $x = 0$ is the only fixed point. We shall denote a fixed point as x_*, y_* . The next important question pertains to the stability of the fixed point. For a point x_* to be stable $x_* + \epsilon \rightarrow x_*$ as $t \rightarrow \infty$, where ϵ is a small displacement. This infers that for a fixed point to be stable the derivatives at this point must be negative.

Greater information can be obtained by linearising about the fixed point utilizing a Taylor expansion. As the system is non-linear, it is important to remember that this approximation is only valid within a limited proximity of the fixed point. Nevertheless, considering the linearized points provides an initial basis from which to consider the system's dynamics. There are some constraints on this approach such as the requirement for the functions to be smooth and continuous. Let us expand about the fixed point of equation (1.23) with a Taylor expansion:

$$\begin{aligned}\dot{x} &= f_1(x_*, y_*) + \frac{\partial f_1(x, y)}{\partial x}(x - x_*) + \frac{\partial f_1(x, y)}{\partial y}(y - y_*) + H.O.T. \\ \dot{y} &= f_2(x_*, y_*) + \frac{\partial f_2(x, y)}{\partial x}(x - x_*) + \frac{\partial f_2(x, y)}{\partial y}(y - y_*) + H.O.T.\end{aligned}\tag{1.24}$$

H.O.T. refers to the higher order terms of the expansion which are generally assumed to be negligible. Since the expansion is taken about the fixed point $x = x_*$, $y = y_*$, it follows that $f_1(x_*, y_*) = f_2(x_*, y_*) = 0$. We are interested in the behaviour in close proximity to the fixed point, as such we should define $\hat{x} = x - x_*$ and $\hat{y} = y - y_*$ and restate the equations in the form:

$$\begin{aligned}\hat{x} &= \frac{\partial f_1}{\partial x} \hat{x} + \frac{\partial f_1}{\partial y} \hat{y} \\ \hat{y} &= \frac{\partial f_2}{\partial x} \hat{x} + \frac{\partial f_2}{\partial y} \hat{y}\end{aligned}\tag{1.25}$$

This may be written in vector form:

$$\dot{\mathbf{X}} = \mathbf{A}\mathbf{X}\tag{1.26}$$

Here $\mathbf{X} = (\hat{x}, \hat{y})^T$ and \mathbf{A} is known as the Jacobian matrix of the system, defined as:

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}\tag{1.27}$$

The behaviour of the non-linear system in close proximity to a fixed point is approximately the same as for a linear system, as such the behaviour can be determined by computing the eigenvalues λ , found by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Based on these eigenvalues, the nature of the fixed point can be predicted. An ensemble of possible types of fixed point are summarized in figure 1.6. Note that we are considering a two-dimensional system and therefore each fixed point is described by a pair of eigenvalues $\lambda_{1,2}$. These may or may not contain both real and imaginary parts $\lambda = a + ib$.

Let us now return to equation (1.23). Taking the derivatives about the fixed point, our linearized system is described by:

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega^2 & \varepsilon \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}\tag{1.28}$$

Eigenvalues are then found by taking the determinant:

$$\begin{bmatrix} 0 - \lambda & 1 \\ -\omega^2 & \varepsilon - \lambda \end{bmatrix} = \lambda^2 - \varepsilon\lambda + \omega^2\tag{1.29}$$

Equation (1.29) is referred to as the characteristic equation. Eigenvalues λ_{+-} are therefore

$$\lambda_{+-} = \frac{\varepsilon \pm \sqrt{\varepsilon^2 - 4\omega^2}}{2}\tag{1.30}$$

Considering the types of fixed point shown in figure 1.6 we can build a picture of the behaviour of the point $(0, 0)$ as a function of the parameters ω and ε . Stability is inferred by the sign of the eigenvalue, whilst a complex eigenvalue infers rotation about the point. A diagram containing information about the transition between different fixed point behaviours is known as a bifurcation

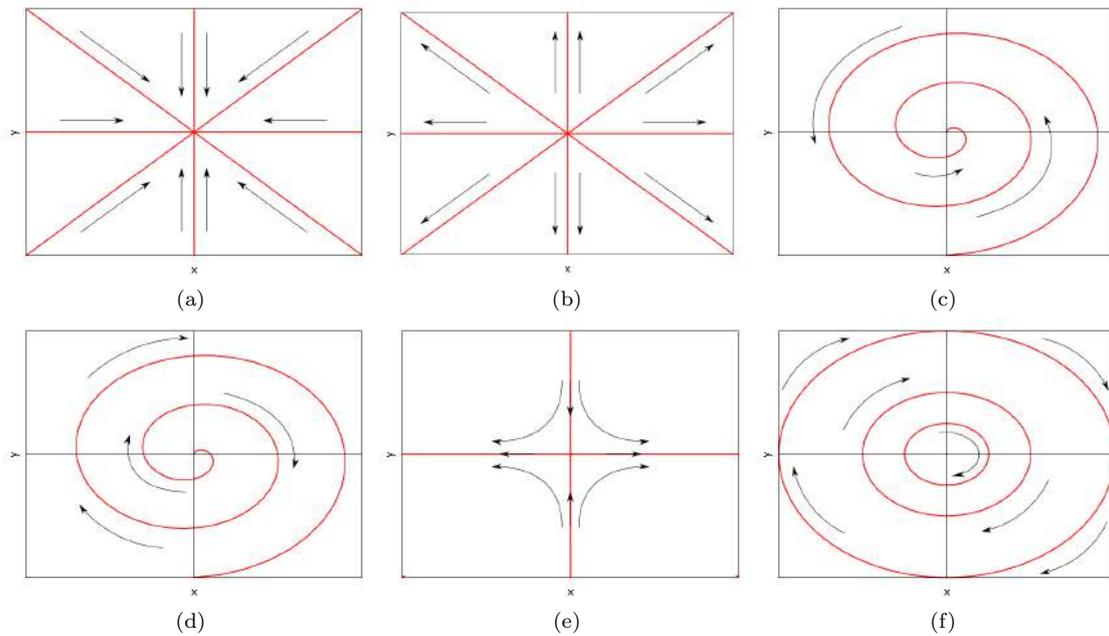


Figure 1.6: The different types of fixed point that may exist on the plane. Red lines indicate possible trajectories whilst black arrows highlight the direction of the vector field. Each type of fixed point can be associated with certain eigenvalue conditions as summarized:

- **(a)**: Stable Node: $Re\lambda_{1,2} < 0, Im\lambda_{1,2} = 0$.
- **(b)**: Unstable Node: $Re\lambda_{1,2} > 0, Im\lambda_{1,2} = 0$
- **(c)**: Stable Focus: $Re\lambda_{1,2} < 0, Im\lambda_{1,2} \neq 0$
- **(d)**: Unstable Focus: $Re\lambda_{1,2} > 0, Im\lambda_{1,2} \neq 0$
- **(e)**: Saddle Point: $Re\lambda_1 > 0, Re\lambda_2 < 0$
- **(f)**: Centre: $Re\lambda_{1,2} = 0$

plot. A bifurcation plot for the parameters ω and ϵ of the VDP oscillator is shown in figure 1.7.

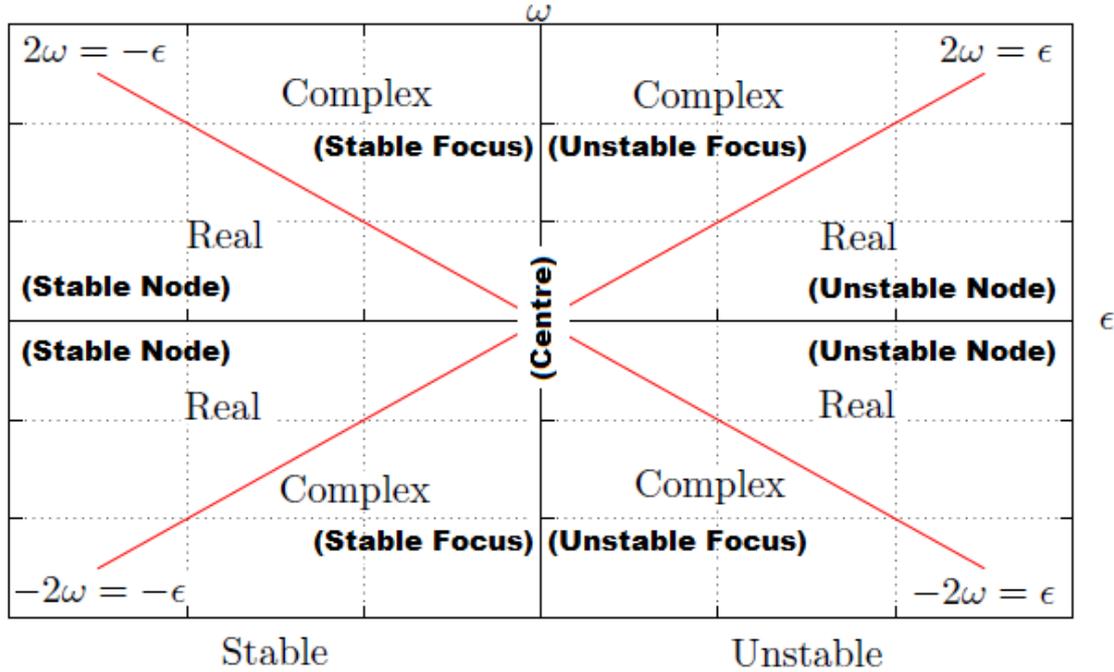


Figure 1.7: Classification of the van der Pol oscillator fixed point at $(0,0)$ in the parameter space ω, ϵ . The red line $|2\omega| = |\epsilon|$ is shown to partition the space into real and complex regions whilst the line $\epsilon = 0$ separates attraction and repulsion.

We can validate our inferences about the fixed point by numerically simulating the VDP oscillator with parameters corresponding to each of the regions of the bifurcation diagram. As is evident from figure 1.8, numerical results support our analysis of the fixed point. The trajectories shown in figures 1.8a, 1.8b, 1.8c and 1.8d also highlight the nature of the system when the fixed point is unstable. In this circumstance, the system variables do not diverge away to ∞ as would be the only possibility for a linear system, rather, they are influenced by an attractor known as a limit cycle, which can only be realized by a non-linear system.

Limit Cycles

The limit cycle that emerges in the VDP system is predicted by Liénard theorem, which applies to any differential equation of the form $\ddot{x} + f(x)\dot{x} + g(x) = 0$, which meets the following conditions [149]:

- $g(x)$ and $f(x)$ are continuously differentiable.
- $g(x)$ is an odd function and $f(x)$ is an even function.
- $F(x) = \int_0^x f(u) \cdot du$ has 1 positive zero at $x = a$, is negative for $0 < x < a$, is positive and non-decreasing for $x > a$ and $F(x) \rightarrow \infty$ as $x \rightarrow \infty$.

These conditions are such that $g(x)$ acts as a restoring force, whilst $f(x)$ has a damping behaviour. This damping behaviour is non-linear; small amplitude oscillations are amplified, but large amplitude oscillations are damped. The total energy (E) of the system is the sum of the kinetic energy ($K = \frac{1}{2}\dot{x}^2$) and the potential energy ($V(x) = \int g(x) dx$).

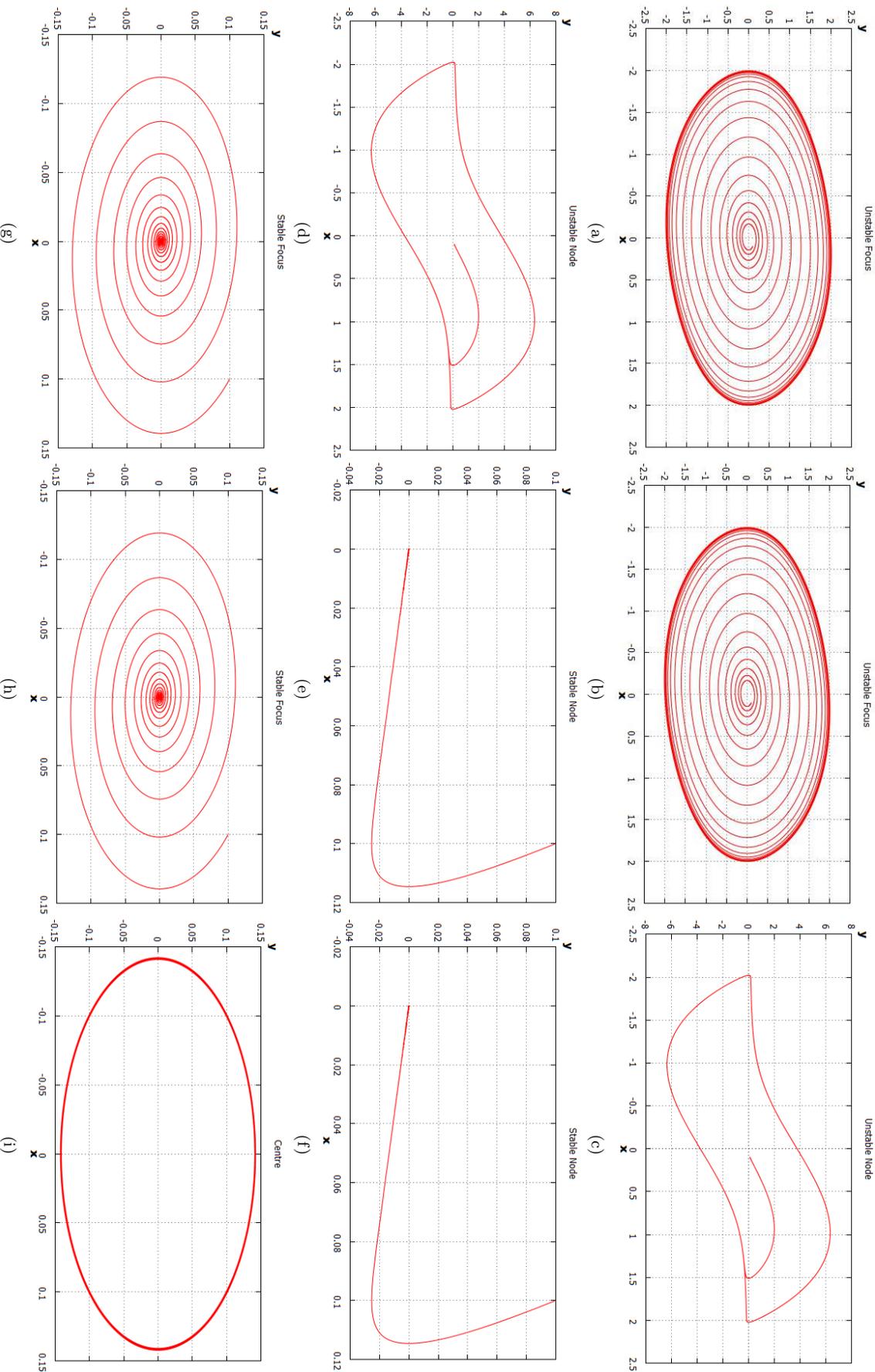


Figure 1.8: Trajectories of the van der Pol oscillator (equation (1.23)) for various parameter values. Values chosen correspond to each of the regions shown in figure 1.7. In each case the initial condition $x, y = 0.1, 0.1$ is set. (a):-**An Unstable Focus**: Parameters take the values ($\epsilon = 0.1, \omega = -1$). (b):- **An Unstable Focus**: Parameters take the values ($\epsilon = 0.1, \omega = 1$). (c):- **An Unstable Node**: Parameters take the values ($\epsilon = 4, \omega = -1$). (d):- **An Unstable Node**: Parameters take the values ($\epsilon = 4, \omega = 1$). (e):- **A Stable Node**: Parameters take the values ($\epsilon = -4, \omega = -1$). (f):- **A Stable Node**: Parameters take the values ($\epsilon = -4, \omega = 1$). (g):- **A Stable Focus**: Parameters take the values ($\epsilon = -0.1, \omega = -1$). (h):- **A Stable Focus**: Parameters take the values ($\epsilon = -0.1, \omega = 1$). (i):- **A Centre**: Parameters take the values ($\epsilon = 0, \omega = -1$).

The limit cycle in the VDP system is the result of a Hopf bifurcation (Poincare-Andronov-Hopf bifurcation) [149]. A Hopf bifurcation occurs when a complex conjugate pair of eigenvalues of the linearised flow, at a fixed point, become purely imaginary. For the VDP system this takes place when $\varepsilon = 0$. Prior to the bifurcation, the system parameters are such that the fixed point is a stable focus. When the eigenvalues cross zero, the point becomes unstable and spirals outwards. This behaviour only persists locally, remote areas of phase space may be unaffected. The interplay between the local and global dynamics culminates to generate a cyclic behaviour relating to the trade-off between non-linear damping and excitation. Initial conditions in close proximity to the fixed point move away along an unstable manifold, whilst conditions further from the fixed point are attracted. These regions must be partitioned by a separatrix. In this circumstance, the separatrix is considered stable and is known as a limit cycle. This environment is portrayed in figure 1.9.

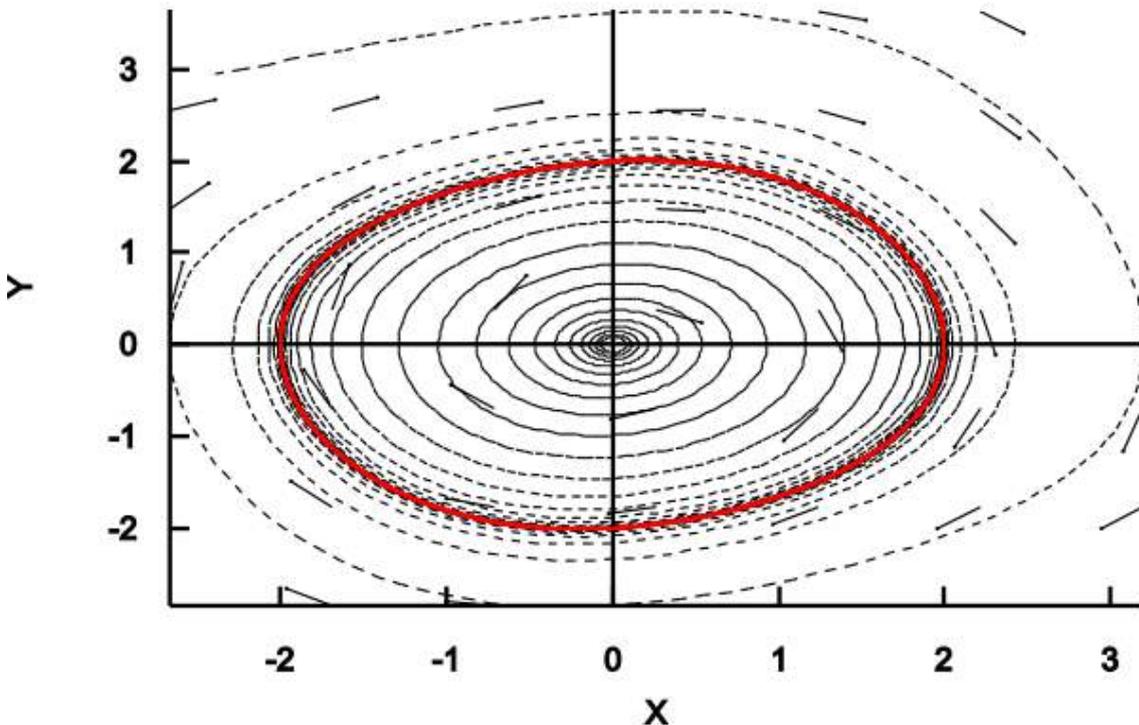


Figure 1.9: The limit cycle (red line) is attracting, all trajectories approach its surface as $t \rightarrow \infty$. In the vicinity of the fixed point, the system behaves in the manner implied by the linearized system, whereas the global behaviour is not affected. This plot is produced utilizing XPPAUT numerical continuation software [44].

A Hopf bifurcation is generically described by a normal form equation [61]:

$$\begin{aligned}\dot{x}_1 &= -x_2 + x_1 (\mu - (x_1^2 + x_2^2)) \\ \dot{x}_2 &= x_1 + x_2 (\mu - (x_1^2 + x_2^2))\end{aligned}\tag{1.31}$$

The behaviour of this system becomes more apparent if we consider the equation in the guise of polar coordinates. Introducing the change in variables $r = \sqrt{(x_1^2 + x_2^2)}$ and $\tan\theta = \frac{x_2}{x_1}$, equation (1.31) becomes:

$$\begin{aligned}\dot{r} &= r (\mu - r^2) = f(r) \\ \dot{\theta} &= 1\end{aligned}\tag{1.32}$$

For $\mu < 0$ there exists just one fixed point for r , this is at $r = 0$. The derivative at this point is μ . As such μ determines the stability of the system. Where $\mu < 0$, the fixed point will be stable. It is clear that $\theta(t) = \theta_0 + t$ is a solution for θ , as such the trajectory will spiral around the origin with increasing time. Combining these inferences we can assume that the fixed point for $\mu < 0$ will be a stable focus. For $\mu > 0$, the fixed point at $r = 0$ will become unstable. There is now a second fixed point $r = \sqrt{\mu}$ which is stable. As θ continuously spirals about the origin and $r \neq 0$ a circle must be traced. This cycle is stable and attracting, it is referred to as a limit cycle.

The limit cycle is a phase space representation of a self-sustained oscillation. Since the VDP oscillator portrays this behaviour, it has become a common starting point for deriving models of sustained oscillation beyond van der Pol's initial intentions. Applications appear within neurology, biology, seismology, chemistry, physics, electronics and economics, as well as in many other relevant areas. In what follows, we investigate how the self-sustained behaviour of the VDP oscillator can be imparted onto our learning system. We shall look at how this can be applied to a simple one-dimensional potential, before considering the more complex case of the N-dimensional learning model. In order to achieve this we shall be highly dependent on numerical methods and computer simulation. A numerical integration scheme is required to integrate the differential equations that are developed. We shall also consider the applications of noise to the system, hence appropriate means of generating uncorrelated signals are required.

1.5 Numerical Approaches

The simulation of appropriate stimuli and the integration of differential equations requires the use of several numerical methods. Such methods allow for the computation of solutions for systems which, to model any other way, would require dramatic simplification, and therefore significantly diminish the corresponding model's usefulness. Computer implemented numerical integration schemes are capable of handling large sets of non-linear equations. When utilized appropriately they rapidly provide accurate solutions. A further advantage of computer simulation is the insight offered into the effect of changing parameter values. Parameters can be manipulated and the simulation repeated in a fraction of the time it would take to implement analytical solutions.

1.5.1 Solving Differential Equations Numerically

There are many ways to numerically solve differential equations. Factors to consider when selecting a routine include the computational expense of the method, the level of accuracy, and the restrictions of the system under investigation. Because our model may incorporate a stochastic aspect, additional considerations must be made. Stochastic systems may be approached from the viewpoint of probability theory by exploring the Fokker-Planck equation [134] that describes the system. Alternatively, such problems can be postulated in the language of Stochastic differential equations (SDE)[54]. Under this guise, the problem may be posed as a Langevin equation [89] and solved via the application of either the Stratonovich [147] or Itô Integral [70]. A Langevin equation takes the form:

$$\dot{x} = a(x, t) + b(x, t) \xi(t) \tag{1.33}$$

The term $a(x, t)$ is commonly referred to as the drift coefficient and $b(x, t)$ as the diffusion coefficient. The requirements on the noise function $\xi(t)$ are discussed at length in [157], as are the methods to solve such equations. In a numerical context we can simplify the problem by selecting a 4th order Runge-Kutta scheme (see Appendix A.1) [150]. This scheme is proven to perform well in comparison to simpler integration schemes such as the Euler method [153], and does not require the computation of derivatives; an advantage over Taylor schemes.

1.5.2 Gaussian Noise

In order to simulate appropriate stimuli we are required to generate Gaussian white noise. White noise is created by a random number generator and transformed to obtain a Gaussian distribution. Computers are deterministic machines. Generating genuine random variables is therefore not possible. Every output is determined by a sequence of events; the rules that govern the behaviour of the program are not subject to random fluctuation. We must therefore make a distinction between true randomness and the pseudo-randomness that we are able to generate. An appropriately generated pseudo-random number portrays all the statistical properties of its counterpart, and as such, is a suitable tool used in stochastic models. Henceforth we will not distinguish between pseudo-random and random numbers. Testing of various routines suggests that a derivative of the Park and Miller algorithm [123] is a viable method of generating randomly distributed uniform deviates. It is clear that the choice of RNG should be carefully considered [129]. Due to the linear congruential execution of many of these algorithms, the period of repetition may render many approaches invalid for all but the simplest of models. The `rand()` function defined by the ANSI C standard has a period of approximately 2^{32} , this is small if we draw a comparison to other implementations such as the Mersenne twister ($2^{19937-1}$) [101]. Our choice of generator is motivated by evidence presented in [123], which cites the Lehmer RNG [88] as the minimal standard based on the statistical analysis presented in [69, 91, 94, 125].

Methods to numerically create GWN include the box-Muller transform [16] and the Ziggurat algorithm [98]. The box-Muller algorithm has been applied effectively across a vast range of modelling problems. In figure 1.10 we demonstrate the algorithms output in conjunction with the RNG (see Appendix A.2), this should convince readers that our numerical choices for the stochastic aspects of the model are appropriate.

1.6 Summary and Conclusions

We have outlined the general state of AI research, focussing on the deficiencies of existing techniques whilst commenting on the direction of ongoing research. The dependence on supervised and reinforcement training routines and the requirement for high levels of redundancy are two failings that the majority of methodologies fail to address. The NN and alternative AI that we have considered possess fundamental flaws that may lead to pessimism about their future development. Unsupervised learning methods are typically complex and merely filter and cluster data-sets rather than being applicable to evolving, real world, real time data processing. In contrast, the method of Janson and Marsden [74] presents a simple, plausible approach to model unsupervised learning.

The relationship between intelligent systems research and biological studies is emphasised; at the current level of biological knowledge it is felt that approaches based around brain inspired architecture may serve as misdirection. Without a proper understanding of the processes occurring

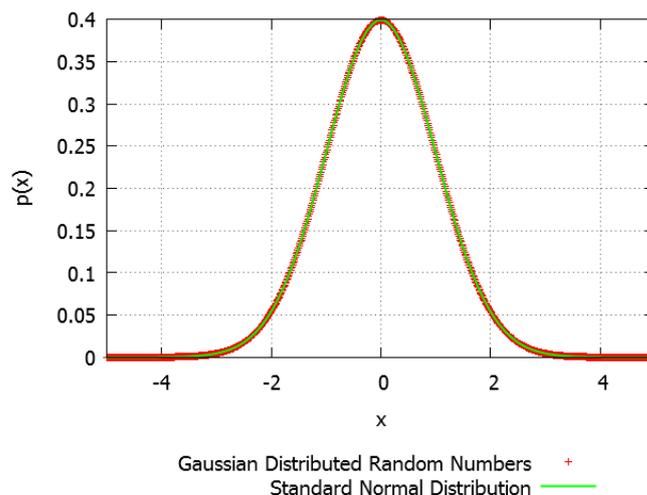


Figure 1.10: Numerically generated Gaussian random numbers compared to the standard normal distribution.

within the brain, we cannot hope to exploit partial information to generate an intelligent system. Further to this are the limitations of engineering. Even if we were to understand the functions computed by the biological system, we have no means of engineering a system with such a highly parallelized structure. The last sixty years of research is yet to meet the expectations of the field's pioneers, despite new ideologies such as back-propagation and Hopfield nets that have revitalized the field. With this in mind, there appears to be cause to peruse alternative strategies. This shows again that Janson's model should be cited as a potential solution. The futility of biological plausibility is rejected in favour of a new rational approach that does not rely on connectionism. Where most models have separate learning and recognition phases, Janson overcomes this oversight, creating a system that may be considered a better replication of how humans interact with their environment.

With the need for this new system made clear, it is the aim of future Chapters to demonstrate this systems efficacy and develop refinements that may improve the efficiency and scope. In this regard, we have already discussed the intended areas of development. The tools of non-linear dynamics should allow us to instigate higher order dynamic attractors for this model, inspired by an analogy to the van der Pol oscillator. Creating attractors more complex than simple fixed point should have implication on how this system may be used to recognize stimuli. Oscillations are ubiquitous in neural systems. Although we do not wish to replicate biological structure, the oscillatory dynamics of our system may be approached and understood with the same tools as used for signal analysis, hence creating a bridge between electronics, biology and this new model.

The study of non-linear dynamical systems is not generally possible from a solely analytical perspective. It is clear that numerical methods will play an integral role in our investigation. Towards this end we have considered numerical methods for solving differential equations and explored the approaches commonly taken to generate Gaussian distributed random numbers.

Chapter 2

Classification and Recognition by a Dynamical System with Self-adapting Vector Field

As is apparent from the overview presented in Chapter 1, the research directions pursued within the fields of AI and machine learning are exceptionally varied. It is also evident that many of the directions taken have led to complications that have not yet been fully resolved. With many of the more classical approaches continuing to present challenges, it is interesting to consider a promising new model. This recently postulated model [74] accurately elicits unsupervised learning without the complications typically observed in other unsupervised approaches [87]. It offers an opportunity to concurrently learn from, and recognize, whilst eradicating the necessity of a training phase. This ensures that the system remains relevant even when the environment changes significantly over time.

In this chapter we will evaluate the memory foam model. By testing its performance for a multitude of stimuli, and considering the implications of parameter choices, we shall verify that this method warrants further attention. It is our hope that this examination will convince readers that the model is capable of combating some of the greater issues faced by AI research. We begin this investigation by defining the differential equations that govern the behaviour of this system. Later we shall consider the implications of the systems parameters, before finally exploring ways in which the model may be developed.

2.1 “Memory Foam’ Approach to Unsupervised Learning’

We first described this model via an analogy to orthopaedic memory foam in section 1.3, here we shall consider the corresponding mathematics. The dynamics of this system are governed by the non-autonomous, non-linear differential equations presented in equations (2.1)-(2.3). Equation (2.1) prescribes the evolution of a time-dependent potential $V(t, \mathbf{x})$ that is continuously shaped by stimuli $\eta \in \mathbf{x}$. The application of a stimuli η is associated with a Gaussian function centred about η . This Gaussian function is described by equation (2.2). Due to the stacking properties of Gaussians, the profile $V(t, \mathbf{x})$ contains all the information pertaining to the classification of stimuli.

We shall later show that this profile converges towards the probability density distribution (PDD) of the ensemble of stimuli. Recognition is simultaneously inferred by the autonomous differential equation (2.3). By computing the gradient of the potential, any initial condition on \mathbf{x} will locate a local minimum of the potential $V(t, \mathbf{x})$. As the local minimum is characterized by the derivative $\frac{dV}{dx} = 0$, the corresponding \mathbf{x} is a fixed point of the systems vector field (defined by equation (2.3)). The gradient approach ensures that all fixed points are attracting. The set of all \mathbf{x} that converge to this same point may be considered to fall within the basin of attraction of the same fixed point. These processes are summarized in figure 2.1.

$$\frac{\partial V}{\partial t} = -\frac{1}{t}(V + g(\mathbf{x} - \eta)) - kV \quad (2.1)$$

$$g(\mathbf{z}) = \frac{1}{(\sqrt{2\pi}\sigma^2)^N} \exp\left(-\frac{\mathbf{z}^2}{\sigma^2}\right) \quad (2.2)$$

$$\frac{d\mathbf{x}}{dt} = -\gamma \frac{\partial V}{\partial \mathbf{x}} + \xi(t) \quad (2.3)$$

In equation (2.2) N refers to the number of dimensions of the system, this is equal to the number of dimensions of the stimulus. A stimulus may be represented as a multi-dimensional vector when it represents more than one aspect of the environment. As an example we may consider a stimulus describing an object. The different dimensions of the stimulus may refer to the objects length, breadth, thickness, weight, colour and a whole host of other relevant parameters. It is clear that when η is N -dimensional, \mathbf{x} must also be an N -dimensional vector.

We previously stated that the shape of $V(t, \mathbf{x})$ converges to the PDD of the stimuli. This feature will be explained in section 2.1.1. It should become clear that the term $\frac{1}{t}$ is important as it ensures that a stationary solution exists. We must also define the topology of $V(t, \mathbf{x})$ when $t = 0$ as without assigning a suitable definition we will have a discontinuity.

The concept of forgetting is represented in this model via the term $-kV$ of equation (2.1). The parameter k controls the rate at which the profile $V(t, \mathbf{x})$ approaches a flat surface. Biological systems tend to forget information that is not regularly re-enforced. By manipulating the constant k we can control the rate at which local minima disappear. For the purpose of this investigation we shall consider that the system does not forget, hence, in all future formulations we shall consider the value of k to be zero.

Another important parameter that may influence the shape of the potential is σ , this is introduced as part of the Gaussian function described by equation (2.2). σ^2 is referred to as the variance of the Gaussian. It controls the width of the curves that represent the application of a stimulus. As we shall later show, σ plays a critical role in defining the number of local minima that our potential may contain. We must also comment on the parameter γ of equation (2.3). This controls the rate at which \mathbf{x} approaches the local minima of the system. As recognition is not achieved instantly, the potential will continue to shape and adjust to new stimuli whilst a local minimum is identified. The rate at which a recognition trajectory approaches a minimum may therefore have implications on what is actually recognized. A new minimum may form and existing minimum can disappear. This may happen rapidly, the rate of gradient descent will therefore have a significant effect on what is actually recognized. We may consider the parameter γ to relate to decision making time. In general, we require our model to recognize stimuli quickly, however, it may be advantageous in some instances to permit the system greater time to locate a minimum. This may result in a

more accurate realization. The function $\xi(t)$ is noise, which can be incorporated to demonstrate the resilience of the system to external perturbation. We should be wary however, as its inclusion may have undesirable consequences on the long term behaviour of the system.

The memory foam model automatically forms clusters of stimuli that may be regarded as distinct classes. A class represents a body of knowledge derived from the experiences of the system. The ability to automatically form classes is an achievement that should not be under-appreciated, as the realisation of unsupervised learning in neural networks or the wider AI field is not easily accomplished. Where such claims are made by alternative models, the approach is typically computationally complex and resource intensive. There may even remain some human intervention, suggesting that the approach is semi-supervised rather than unsupervised. Further, AI models typically operate in two distinct phases. First is a period of training, then unfamiliar stimuli are encountered and recognized. Janson and Marsden’s approach presents a method to recognize and classify stimuli simultaneously, without the requirement for a training phase or significant data-manipulation. The ‘curse of dimensionality’ pertains to a severe shortcoming of the neural network methodology. Due to the distributed architecture and finite number of activation paths, outputs can be portrayed that are not correlated to any specified input. These are known as spurious attractors. As the number of retained stimuli increase, so too must the size of the system if erroneous responses are to be avoided. If too many patterns are learnt, the result can be a cascade of unit activations that render the network useless [60]. Due to the deviation from a rigid unit architecture, the memory foam model should not experience this issue. The properties of the Gaussian function (equation (2.2)) ensure that the profile remains smooth and minima only appear at desirable locations. Because the system continues to learn throughout its operation, it is not restricted to learning about a finite number of stimuli. Over time it may forget certain classes, however, this information can be re-learnt if the system experiences similar stimuli again. The amount of information that the model may represent is therefore unlimited. We should note that, at any finite time moment, there will be a maximal number of classes that can be represented by the potential, these classes can however be redistributed given enough time. Here we may draw some parallels to a human learner. Although we have an exceptionally large capacity for information storage, it is certainly true that we forget things. Presented with some information that we have previously learnt, we are not guaranteed to remember. This is not a problem, as, like this new model we continuously learn from our environment.

2.1.1 Constraints on the Stimuli and Convergence of the Potential

Let us consider how we may approach finding an analytic solution to equation (2.1). Note that we shall only consider the case where $k = 0$. Hence the differential equation becomes:

$$\begin{aligned} \frac{\partial V}{\partial t} &= -\frac{1}{t}(V + g(\mathbf{x} - \eta)) && \text{this may be rewritten in the form:} \\ \frac{\partial V}{\partial t} + \frac{1}{t}V &= -\frac{1}{t}g(\mathbf{x} - \eta) \end{aligned} \tag{2.4}$$

From which we can define an integrating factor $e^{\int \frac{1}{t} dt} = e^{\ln|t|} = t$, allowing equation (2.4) to be written:

$$V = -\frac{1}{t} \int g(\mathbf{x} - \eta) dt \tag{2.5}$$

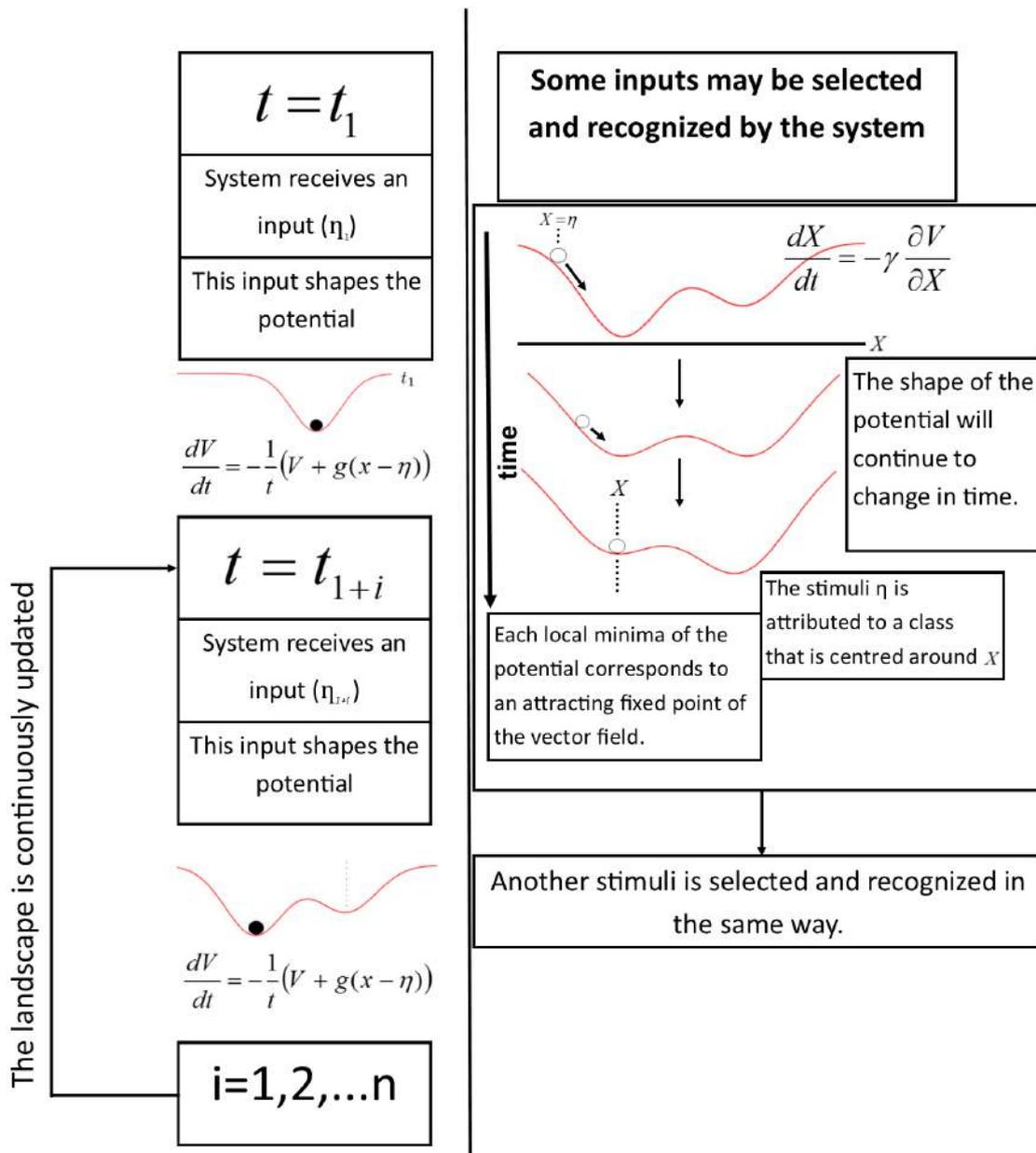


Figure 2.1: Graphical representation of the operation of the autonomous learning model. The shaping of the system's potential (unsupervised learning) occurs simultaneously with the recognition of a subset of stimuli.

This is the negative time average of the function $g(\mathbf{x} - \eta)$. It is apparent that the term $\frac{1}{t}$ is required to ensure that the system is able to reach a stationary state. Where the stimuli is ergodic we may replace a time average with a statistical average. We must therefore make some assumptions about the behaviour of η .

We assume that a stimuli is a result of a stationary, ergodic random process $H(t)$, with a probability density distribution $P(\eta_1, \eta_2, \dots, \eta_N)$. Stationarity implies that the stochastic process has a joint probability distribution that is not time dependent. This is the case if the stimuli $\eta = (\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_k})$ has a distribution such that:

$$P(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_k}) = P(\eta_{t_1+\tau}, \eta_{t_2+\tau}, \dots, \eta_{t_k+\tau})$$

for all k and all τ . Here τ is a shift in the time parameter t and η is a k dimensional vector representing the input. An ergodic function retains all statistical properties in a single sufficiently long realization. This means that all statistical measures remain unchanged regardless of whether η is a time ensemble or a single realization.

We should also consider the behaviour of the function $g()$. This is a Gaussian function as defined by equation (2.2). If we consider the case where $\sigma \rightarrow 0$ this Gaussian function can be replaced with a delta function. In this limit we can therefore define:

$$V = -\frac{1}{t} \int g(\mathbf{x} - \eta) dt = -\frac{1}{t} \int \delta(\mathbf{x} - \eta) dt \quad (2.6)$$

Then, considering the behaviour of $V(t, \mathbf{x})$ as $t \rightarrow \infty$ we may replace the time average with the statistical average:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int g(\mathbf{x} - \eta) dt = \int P(\eta) \delta(\mathbf{x} - \eta) d\eta \quad (2.7)$$

Using the properties of delta functions [18] $\int P(\eta) \delta(\mathbf{x} - \eta) d\eta = P(x)$. Hence, in the limit $t \rightarrow \infty$ it is apparent that $V(t, \mathbf{x}) \rightarrow -P(x)$, which is the negative probability density distribution of the inputs.

We have made several assumptions to reach this solutions. Most significantly we have defined the statistical nature of the stimuli and replaced the Gaussian function $g()$ with a dirac-delta function. Generally, we will require the use of numerical methods to find solutions of equation (2.1). This requirement is a result of the way the system interacts with stimuli. Stimuli are interpreted in a iterative manner, a complete knowledge of a stimuli's distribution is unknown, but is developed as $t \rightarrow \infty$. This property is intended to replicate the way a biological learning system would receive inputs.

2.2 Initial Simulations

The first demonstration of the versatility of this new model was presented in [74]. The model was shown to learn and recognize musical stimuli supplied by a flute. Musical notes were characterized by frequency, with the highest spectral peak of a fourier transform [72] identifying the main frequency of the note played. This then served as the input to the learning system. Further to exploring the classification of notes within a one-dimensional framework, Janson and Marsden extended the model to four-dimensions and considered temporal patterns of four notes from the

children’s song “Mary had a little lamb”. Examples of the foams convergence for various stimuli with different probability density distributions were shown in [99]. Methods to create correlated and uncorrelated inputs were also compared. Marsden showed that the energy potential $V(t, x)$ converges to the negative probability density distribution of the input faster when inputs remain uncorrelated. The main contribution of [99] is to demonstrate a method to generate oscillatory and chaotic behaviour utilizing delay-differential equations. It is suggested that such methods may be applied to the problem of global optimization by utilizing a new procedure referred to as ‘Delay Annealing’. Our investigation will deviate from this time-delay approach, instead focussing on the paradigm of self-sustained, autonomous oscillation. It is our intention to represent the system with dynamics that may infer a hierarchy of classes. Each class of knowledge should be represented in the phase space by an attractor that has qualities specific to the class.

Having introduced the model and explained the various aspects of its composition, we should now progress to test and verify the assertions that have been made. The equations of the model may be integrated using a number of different numerical schemes [153]. Due to its accuracy and computational efficiency we shall utilize a fourth order Runge-Kutta method, as outlined in appendix A.1. The potential of the system should evolve to replicate the probability density distribution of the ensemble of stimuli. We should be aware that this stationary behaviour may only be realized as $t \rightarrow \infty$. We should consider the time scale on which an approximate convergence is reached and may also investigate the effect of parameter choices on convergence times. Thus far, the model has only been considered for a small set of differently distributed stimuli. We will now consider a method to generate stimuli in accordance with a prescribed distribution. This will allow us to compare the evolution of the potential to a known stimuli distribution. We can then contrast the shaping of the system for different stimuli. To effectively test the model it is imperative that we are able to define a wealth of different inputs with varying degrees of complexity. This is important as we need to investigate the system’s efficacy in a number of different artificial environments.

2.2.1 Simulating Stimuli

Two simple methods of simulating a stimuli with a prescribed probability density distribution are provided in [99]. The key difference between these approaches is the correlation between subsequent values. Marsden shows that the model converges faster to a given probability density distribution in the event that inputs are uncorrelated. Later we shall consider a real world stimuli observed via a web-camera: in this circumstance stimuli may be considered uncorrelated. At present we shall concentrate on producing simple correlated inputs. Correlated stimuli may be produced by integrating a stochastic differential equation of the form:

$$\dot{\eta} = -\frac{dF(\eta)}{d\eta} + D\xi(t) \quad (2.8)$$

This formulation is equivalent to the the behaviour of a Brownian particle within an arbitrary potential $F(\eta)$ [95]. In this context, η is the stimuli, ξ is Gaussian white noise of intensity D and $F(\eta)$ is a function prescribing the desired distribution of stimuli. We shall utilize the stochastic differential equation (2.8) to generate a series of differently distributed stimuli. These may then be used to investigate the scope of the model’s application. It should be noted that a non-correlated stimuli may be generated by applying a non-linear transform to a Gaussian white noise (GWN). Alternatively, we may convert a correlated stimuli to a uncorrelated equivalent by utilizing a

numerical random sampling method.

The equations (2.9(a-f)) indicate a number of different functions that may be utilized to describe $\frac{dF(\eta)}{d\eta}$. These functions correspond to the illustrations of $F(\eta)$ shown in figure 2.2.

$$\frac{dF}{d\eta} = 0.06\eta^3 - 0.9\eta - 0.3 \quad (2.9a)$$

$$\frac{dF}{d\eta} = \int \frac{1}{1500} (\eta - 7) (\eta - 5) (\eta - 3) (\eta + 2) (\eta + 4) (\eta + 7) .d\eta \quad (2.9b)$$

$$\begin{aligned} \frac{dF}{d\eta} = & 0.08\eta^3 + \frac{40(\eta + 4)}{\left((\eta + 4)^2 + 0.4\right)^2} + \frac{50(\eta + 2)}{\left((\eta + 2)^2 + 0.5\right)^2} + \frac{50(\eta - 0.5)}{\left((\eta - 0.5)^2 + 1\right)^2} \\ & + \frac{300(\eta - 6)}{\left((\eta - 6)^2 + 2\right)^2} + \frac{300(\eta + 7)}{\left((\eta + 7)^2 + 2\right)^2} \end{aligned} \quad (2.9c)$$

$$\begin{aligned} \frac{dF}{d\eta} = & 0.2\eta + 6(\eta - 4)e^{-0.5(\eta-4)^2} + 3(2\eta + 10)e^{-(\eta+5)^2} \\ & + 8\eta e^{-\eta^2} + 5(8\eta - 56)e^{-4(\eta-7)^2} \end{aligned} \quad (2.9d)$$

$$\begin{aligned} \frac{dF}{d\eta} = & 0.2\eta + 5(0.4\eta + 0.4)e^{-0.2(\eta+1)^2} + 7(6\eta + 42)e^{-3(\eta+7)^2} \\ & + 6(\eta - 6)e^{-0.5(\eta-6)^2} \end{aligned} \quad (2.9e)$$

$$\begin{aligned} \frac{dF}{d\eta} = & 0.2\eta + 5(0.2\eta + 0.4)e^{-0.1(\eta+2)^2} + 10(\eta + 7)e^{-0.5(\eta+7)^2} \\ & + 6(\eta - 6)e^{-0.5(\eta-6)^2} + 6(6\eta - 12)e^{-3(\eta-2)^2} \end{aligned} \quad (2.9f)$$

We have chosen various types of function to simulate stimuli. This is intended to demonstrate the scope of the approach and allows us to investigate the shaping of the memory foam model for various stimuli compositions. Equation (2.9a) may be considered the simplest of these distributions. The function $F(\eta)$, as shown in figure 2.2a, is a double well polynomial. A similar distribution is considered in [74], which suggests that this function is a good place to start in order to validate the previously published results. Equation (2.9b) describes the derivative of a different polynomial distribution. In this instance $F(\eta)$ is a sixth order polynomial as shown in figure 2.2b. It contains three local minima, each with a different curvature. This profile may be seen as the logical progression from the first example. Equation (2.9c) and the corresponding graphic of $F(\eta)$ shown in figure 2.2c utilizes a series of Lorentzian functions and a quartic term to describe the stimuli. This potential is the most complex that we shall consider at this stage. $F(\eta)$ contains five local minima. The quartic term ensures that η remains within a bound range in the vicinity of these Lorentzian curves. The final three functions $\frac{dF}{d\eta}$, described by equations (2.9d,e,f), correspond to figures 2.2d,e,f. These consist of ensembles of Gaussian functions. In each case a quadratic term is included to ensure values remain within the necessary bound region. In each example the distribution and number of minima are chosen so as to create a profile with unique features. Figures 2.2d,f both contain four minima, whilst 2.2e has only three. The Gaussian approach is relatively simple to manipulate as we can define the variance and mean of each Gaussian curve independently. This method is particularly simple if we draw a comparison to the polynomial descriptions. By applying different normalization coefficients to each Gaussian function we can also control the depth of each minimum. As shown, there is a large scope for creating interesting distributions with this function.

2.2.2 Application to the Memory Foam Model

Applying the different stimuli η , as described in section 2.2.1, to the differential equation (2.1), we can observe the evolution of the potential $V(t, x)$ which characterizes the classification of stimuli. Figure 2.3 illustrates the time evolution of the potential for the various stimuli generated by equation (2.8). The stimuli distributions, described by the functions $F(\eta)$, correspond to equations (2.9a-f). The labels of figure 2.3 correspond to the labels assigned to each of the functions described in equation (2.9). In each case $V(t, x)$, given sufficient time, converges towards the negative probability density distribution of the stimuli (shown by black line). We can show this probability density distribution by normalizing a histogram composed of all the stimuli experienced by the model. Comparing figures 2.3 and 2.2 it is clear that the stimuli distribution is related to the functions $F(\eta)$. It is important to note that in each of the examples given in figure 2.3 the Gaussian width parameter $\sigma = \sqrt{0.05}$. In general, this is sufficiently narrow to model each of these prescribed stimuli (the small discrepancy in figure 2.3c, between the distribution of stimuli and potential, may be attributed to an excessively wide Gaussian parameter σ). If the stimuli distribution contains finer details then this width parameter may need to be reduced. An alternative to reducing σ would be to linearly transform the stimuli onto a wider range of x . Selection of σ may have a significant effect on the number of classes that emerge within the system. It may also influence the time it takes for the system to effectively reflect the distribution of stimuli. We shall consider the implications of σ later, for now we shall continue to build a solid understanding of the prototype model.

Recognition is illustrated in each plot of figure 2.3 by a set of initial conditions that follow the contours of the potential (coloured lines at the base of each plot). These trajectories locate the local minima of the system. This is achieved via a gradient descent function described by equation (2.3). The local minima of the potential correspond to fixed point attractors of the systems vector field. The system may consider the recognition of any number of stimuli simultaneously. We choose to demonstrate the recognition of a cross section of from across the range of x in order to approximate the basins of attraction of each fixed point. Note that the same initial condition $\eta = x$ may be recognized differently depending on when a stimulus is encountered by the system. This is because the system's potential $V(t, x)$ is continuously evolving. By this reasoning, the recognition rate parameter γ may influence which class a stimulus is recognized as belonging to. In the examples shown in figure 2.3 $\gamma = 1$. Parallels may be drawn between our choice of this parameter γ and thinking time. Taking longer to make a decision (allowing the gradient descent point to converge more slowly) generally results in a better decision (means the point is attributed to a class that better resembles the stimulus).

2.3 The Influence of Gaussian Width on Class Formation

To understand the effect that the Gaussian width parameter σ has on the learning system we can study the evolution of the model for various values of this parameter. In each case we require an appropriate stimuli, that presents a high enough degree of complexity to illustrate the possible problems. A stimuli generated by the application of equation (2.8), where $F(\eta)$ is defined by equation (2.10) appears to meet this demand.

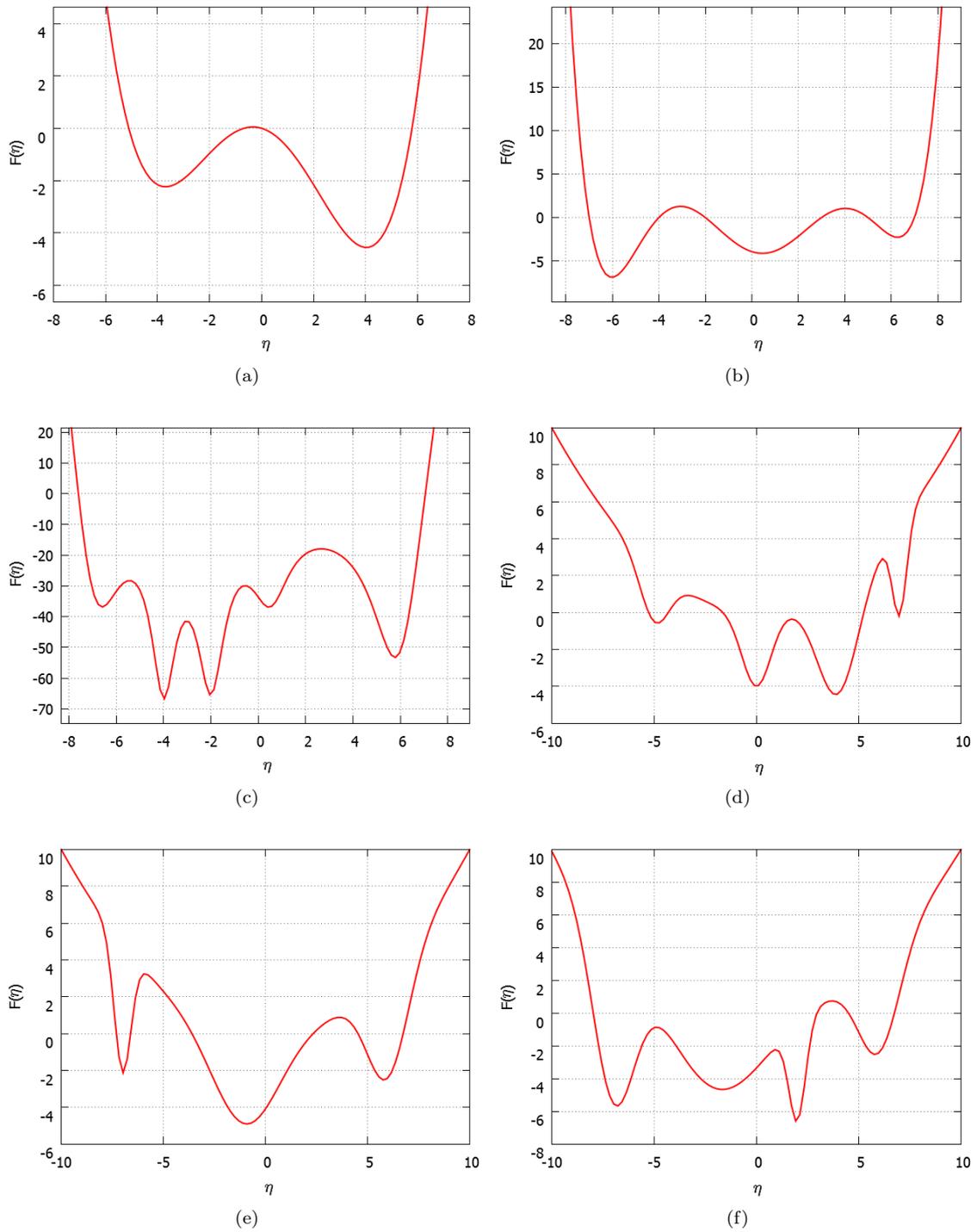


Figure 2.2: Graphical illustrations of the various functions $F(\eta)$ that are employed to generate stimuli with specific distributions. The corresponding derivatives $\frac{dF}{d\eta}$ are described by equations (2.9). (a):- $F(\eta)$ is a fourth-order polynomial. (b):- $F(\eta)$ is a sixth-order polynomial. (c):- $F(\eta)$ is a series of Lorentzian functions combined with a fourth order polynomial. (d),(e),(f):- $F(\eta)$ is a series of Gaussian functions combined with a parabolic function.

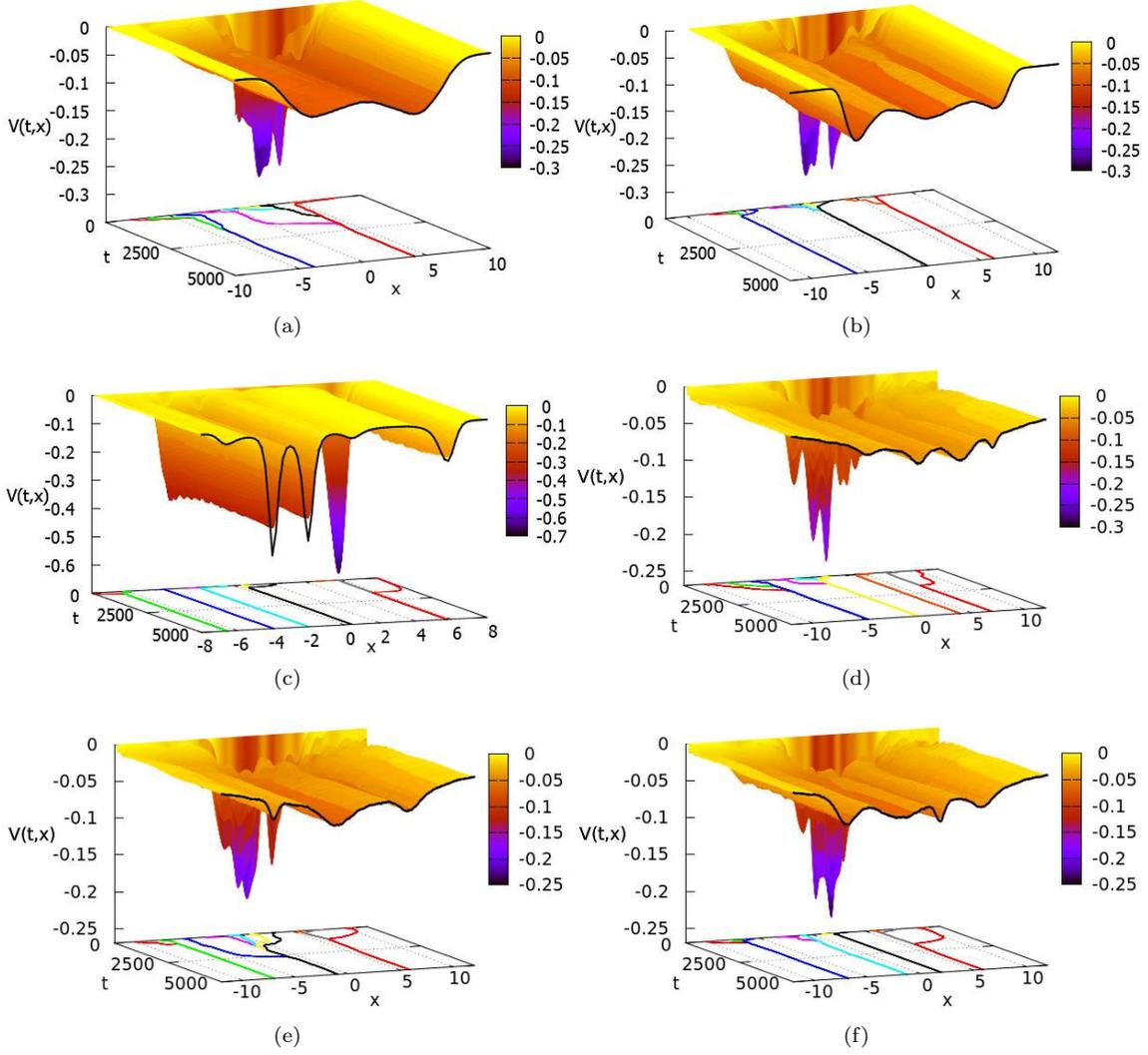


Figure 2.3: Six examples of the evolution of $V(t, x)$ under the influence of different stimuli. These stimuli are described by equation (2.8). The corresponding $\frac{dF(\eta)}{d\eta}$ are provided in equations (2.9). In each figure the probability density distributions of stimuli is represented by a solid black line at $t = 5000$. This distribution is determined by normalizing a histogram containing all the stimuli that the system has experienced. Each figure also shows a cross-section of recognition trajectories on the plane (x, t) , these identify the local minima of the potential. The local minima correspond to the attractors of the system due to the gradient approach outlined by equation (2.3). Unknown stimuli with arbitrary initial conditions are assigned to a class, the phase space is separated into basins of attraction surrounding each fixed point. Note that in each example $\sigma = \sqrt{0.05}$, $\gamma = 1$ and $D = 5$.

$$\begin{aligned}
F(\eta) = & 0.005\eta^4 - 0.4\eta^2 - 10e^{-0.5\frac{(\eta-2)^2}{0.08}} - 10e^{-0.5\frac{(\eta-3)^2}{0.2}} - 10e^{-0.5\frac{(\eta-4)^2}{0.08}} - 10e^{-0.5\frac{(\eta+4)^2}{0.08}} \\
& - 10e^{-0.5\frac{(\eta+3)^2}{0.15}} - 5e^{-0.5\frac{\eta^2}{0.2}}
\end{aligned} \tag{2.10}$$

For each simulation we ensure that the PDD of the stimuli remains constant. we show that the choice of σ has a significant influence on the formation of classes. In figure 2.4 we compare the evolution of the potential for the same stimuli, but with various values of σ . Where smaller values of σ are selected, the potential $V(t, x)$ bares a closer resemblance to the probability density distribution of stimuli on the respective time scale. In contrast, a large σ (see figure 2.4a) prohibits the formation of some local minima altogether. Figure 2.4a demonstrates the case where $\sigma = \sqrt{0.5}$, in this situation the profile only contains four local minima, however, the PDD of stimuli remains the same eight minima shape as in the other examples. Reducing σ increases the number of local minima. Where $\sigma = \sqrt{0.1}$, the potential contains five local minima and where $\sigma = \sqrt{0.05}$, seven local minima exist. We observe all eight local minima for $\sigma = \sqrt{0.01}$. There appears to be very little difference between the stimuli distribution and the potential $V(t, x)$ if we reduce σ further. A comparison of the potential (at $t = 5000$) for various values of σ with the probability density distribution of inputs is provided in figure 2.4f.

Selecting an appropriate value for σ is clearly very important if the potential is to converge to the probability density distribution of the stimuli. It is evident that some problems may be avoided by simply selecting a small value of σ . For the example given in figure 2.4 this may mean $\sigma < \sqrt{0.01}$, however, how we define small is clearly dependent on the distribution of stimuli. Although a small σ is beneficial in the sense that classes form as required, selecting a small value has an adverse effect on the systems ability to generalize when stimuli form distinct bands. Due to the properties of the Gaussian function, any location on the potential should be associated with a basin of attraction of a fixed point. However, when there is a large separation between stimuli there may be large regions of the phase space that are characterized by a potential that is approximately flat. If we wish to use this model as a type of auto-classifier this is clearly objectionable. We require that any point in the phase space associates with a class within a reasonable time period. Utilizing a Gaussian with a larger variance will help remove these flat regions of potential and ensure faster gradient descent. We are therefore presented with the problem of balancing between generalization and replicating the distribution of stimuli. At this stage we will present two suggestions to overcome this issue. The first involves making σ a function of time, the other involves scaling σ with a function of $V(t, x)$.

σ as a function of time: Let us consider how a human might observe and learn from the world around them. Consider a child learning colours for instance. A child would first learn the basics ‘red’, ‘green’, ‘blue’ etc. Given different shades or tones they are likely to still refer to only a few basic, broad classes. As time passes they begin to distinguish between different types of colour that they previously referenced as the same. i.e. They may now distinguish between scarlet and crimson. This concept may be captured in the learning model by describing σ as a function of time. Initially we require very broad classes, this means that σ must be large, as time passes σ should be reduced so that it can reflect the difference between two similar classes such as scarlet and crimson, that may not have been distinguishable with a larger σ . Clearly this function cannot be linear as we require that $\sigma > 0$ and bound within some range. We propose the non-linear

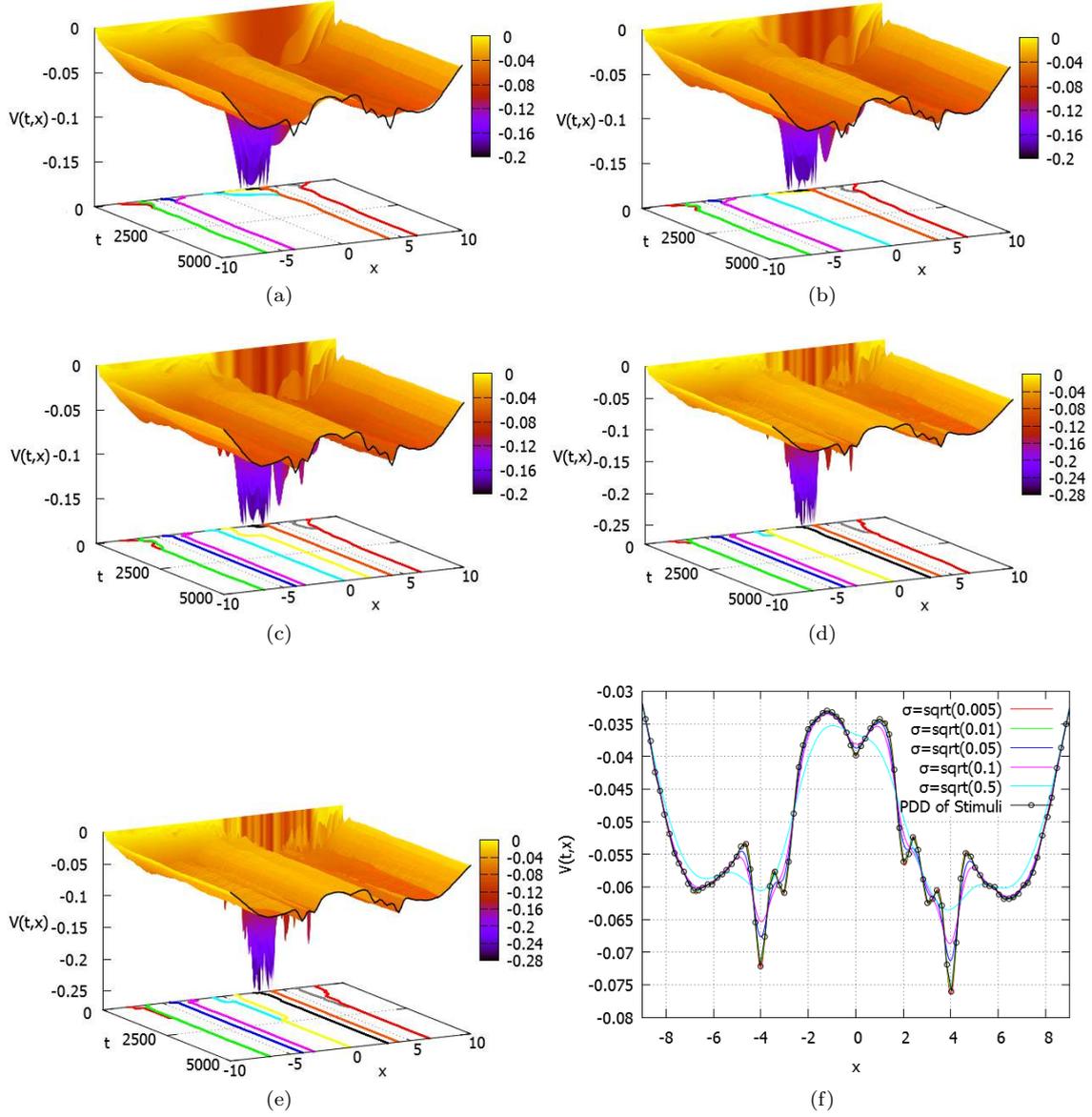


Figure 2.4: **(a)-(e)**: Evolution of $V(t,x)$ under the influence of a stimuli described by equation (2.8) where $\frac{dF}{dn}$ is given by equation (2.10). In each case $D = 5$ and $\gamma = 1$. The evolution of initial conditions demonstrating recognition (coloured lines) are shown on the plane (t,x) of each plot, these points are evenly interspersed along the range of x . These trajectories approach the fixed points of the system, which act as attractors due to the gradient method applied. The parameter σ is shown to effect the shape of $V(t,x)$. In **(a)**: $\sigma = \sqrt{0.5}$, **(b)**: $\sigma = \sqrt{0.1}$, **(c)**: $\sigma = \sqrt{0.05}$, **(d)**: $\sigma = \sqrt{0.01}$ and in **(e)**: $\sigma = \sqrt{0.005}$. It is apparent that the distribution of stimuli (which is the same in each plot and has been obtained by building a histogram of the inputs) as indicated by the black line at $t = 5000$ in each plot does not strictly coincide with the shape of the potential at $t = 5000$ unless σ is suitably small. In **(f)** the shape of the potential at $t = 5000$ for **(a)-(e)** is plotted along with the PDD of stimuli.

function described by equation (2.11) and shown in figure 2.5a as an appropriate method to scale σ . This function is continuous and smooth, ensuring that σ^2 tends to a small positive constant as $t \rightarrow \infty$. This is necessary as we do not want our Gaussian function to become equivalent to a delta function. In some circumstances an excessively narrow Gaussian function would impair the smoothness of the potential and prevent trajectories associated with the gradient of the potential converging to a fixed point of the system within a reasonable time scale. The parameters A , B and C of equation (2.11) influence the slope of the curve, the intersect at $x = 0$, and the asymptotically approached constant.

$$\sigma = \sqrt{A (\log (1 + e^{(-Bt-C)}) + C)} \quad (2.11)$$

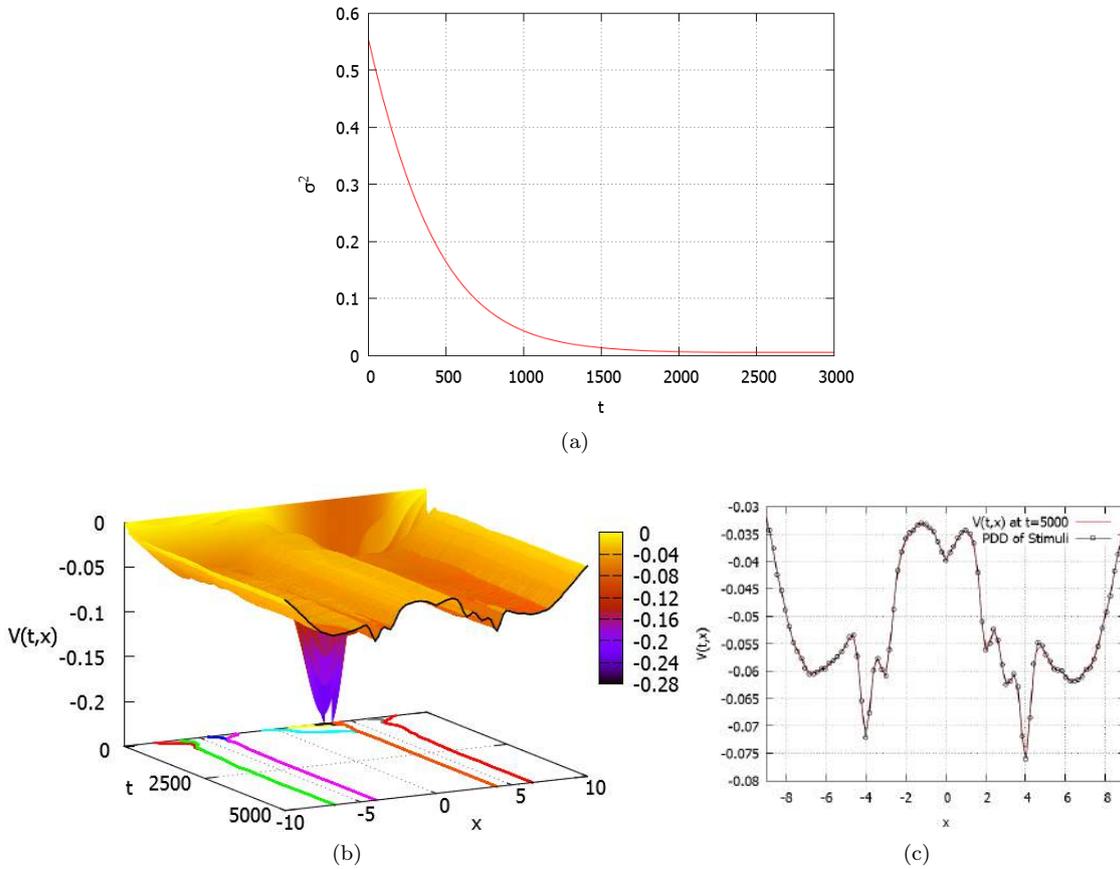


Figure 2.5: **(a)**: The value of σ^2 may be presented as a smooth function of time by utilizing equation (2.11). Here the parameters of the function are $A = 0.8$, $B = 0.003$ and $C = 0.006$. σ^2 tends to a small positive constant as $t \rightarrow \infty$. As the system spends more time learning it should be able to appreciate final details of the stimuli. **(b)**: The shaping of the potential $V(t, x)$ when sigma is described by equation (2.11) with the above parameters. $V(t, x)$ converges to the shape of the PDD of stimuli, shown by the black line at the $t = 5000$. On the plane t, x a cross-section of recognition trajectories are included (coloured lines). Note that these have the same initial conditions as those shown in figure 2.4 and are controlled by the rate parameter $\gamma = 1$. **(c)**: Comparison of the PDD of stimuli and the potential at $t = 5000$ for the potential shown in (b).

It is evident from figures 2.5b and 2.5c that describing σ via a function of time does not impair the shaping of the potential. A comparison between the PDD of stimuli and the potential highlights that the system still shapes to the stimuli as required. The illustration shown in figure

2.5b is clearly comparable to figure 2.4e (where $\sigma = \sqrt{0.005}$). This method should ensure that regions of the potential do not remain approximately flat even if stimuli are well separated across x . The dependence of the system on the Gaussian width function is somewhat negated by allowing the function to adjust over a range during the runtime of the model. Despite this, there still remains some difficulty. An appropriate choice of range must be made, as well as defining an appropriate time scale over which this range should be covered. These factors are described by parameters A, B, C in equation (2.11). Figure 2.5 demonstrates the case where $A = 0.8$, $B = 0.003$ and $C = 0.006$. Although effective in this circumstance, the same parameters may not be favourable for alternative stimuli. Using a range of σ values may be seen as a more flexible and general approach than selecting a single value. In addition to improving the versatility of the model, this adjustment may also be considered to replicate an additional feature of the way a human learns. The model starts by making some vague generalizations, before developing the ability to discern finer-details with time.

σ as a function of $V(t, x)$: An alternative to representing σ as a function of time is to make it dependent on $V(t, x)$ at the point $x = \eta$. Similar stimuli that have been seen many times will cause a deep well to form in the potential. It is logical to propose that if we receive similar stimuli many times, then we should be able to recognize finer differences between them. To appreciate finer details in the structure of the stimuli we require a narrower Gaussian. By making σ a function of $V(t, x)$, we ensure that if the stimulus is close to a local minimum, the Gaussian is narrow and there is therefore a chance that a new class will emerge. Stimuli that are a long way from local minima will be represented by a Gaussian with a much higher variance. This approach can be seen as a different way to capture the same human based learning behaviour as making σ time dependent. In this case the variance is dependent on what has already been learnt, rather than simply on time spent learning. This may be considered a more realistic proposition. A function of the same type as equation (2.11) can again be utilized. Equation (2.12) describes this alternative approach to obtain a scaled value of σ .

$$\sigma = \sqrt{A (\log(1 + e^{(-BV(t,x)-C)}) + C)} \text{ taken at } x = \eta \quad (2.12)$$

A demonstration of the application of this function is shown in figure 2.6. Here we have set $A = 0.8$, $B = 110$ and $C = 0.0006$. A visualization of the function defining σ is given in figure 2.6a, whilst the time dependent shaping of the potential, and a comparison between the potential at $t = 5000$ and the PDD of stimuli are given in figures 2.6b and 2.6c respectively. It is evident from figure 2.6c that there is good agreement between the PDD of inputs and the potential. What is also interesting is the number of classes recognized in this example. Although the same initial conditions are selected (as shown by the coloured lines at the base of figure 2.6b), more classes are recognized in this illustration than in figures 2.5b or 2.4e.

We may consider that assigning a range of values to σ is a more general approach than assigning a single value to the parameter. Despite this, it remains unclear over what range this function should operate and how steep the slope of the transform should be. As the value of σ required to model the distribution of the stimuli is dependent on the stimuli itself, no simple method to ascertain an appropriate range is apparent. A further conceptual problem with this idea is that several closely related stimuli may not necessarily be defined by a deep well. A deep well may arise due to stimuli occupying a small region of the phase space, whilst other areas remain approximately

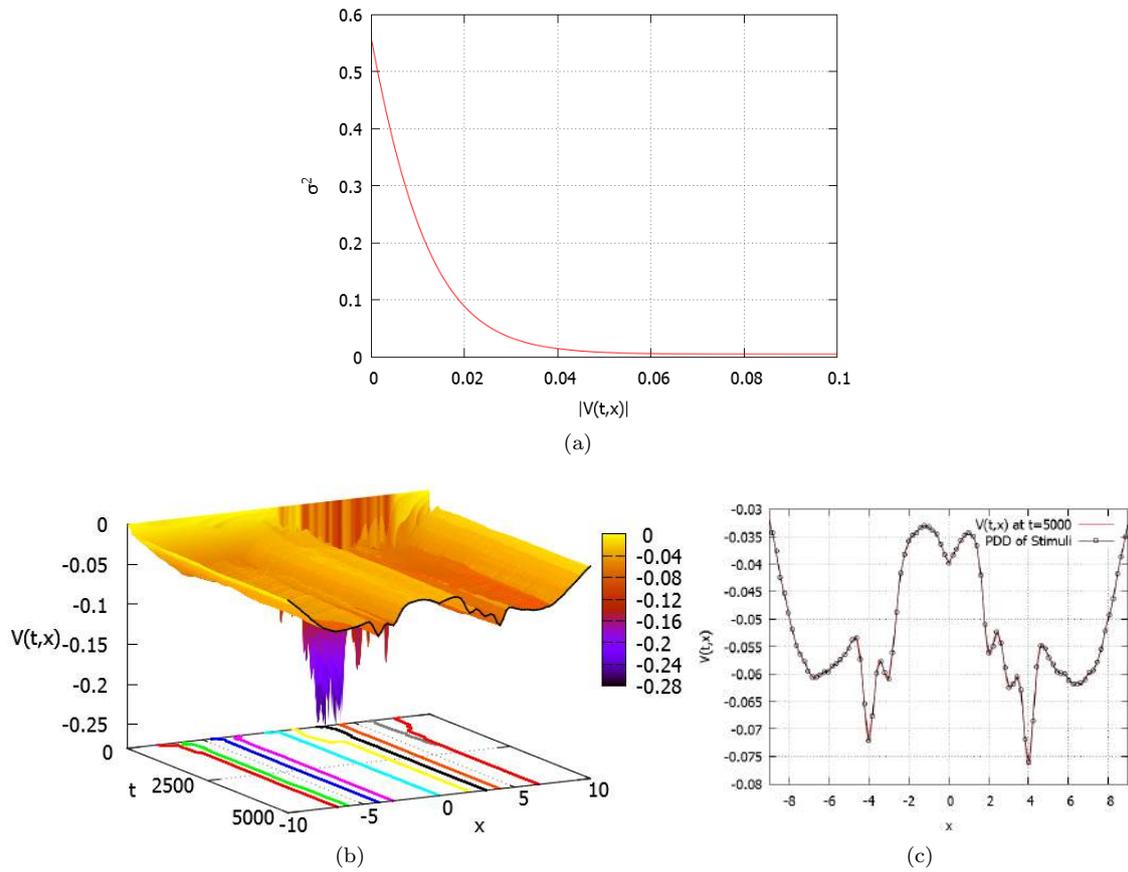


Figure 2.6: **(a)**: The function described by equation (2.12) for the parameter values $A = 0.8$, $B = 110$, $C = 0.006$. This relates σ^2 to the depth of the potential $V(t, x)$. Smaller values of $V(t, x)$ are associated with more prevalent stimuli, the system should be able to better differentiate between things it has seen more often, this is achieved by relating a deep potential well to a low value of σ^2 . **(b)**: The shaping of the potential when σ is described by equation (2.12). The PDD of stimuli is shown by the black line at $t = 5000$. A set of recognition trajectories is shown on the plane t, x , these again have the same initial conditions as the trajectories shown in figure 2.4 and are influenced by the rate parameter $\gamma = 1$. **(c)**: Comparison of the PDD of stimuli and the potential at $t = 5000$ where σ is described by a function of $V(t, x)$.

flat. Stimuli that are similar, may, in some circumstances be better inferred from a broad potential well. What is defined as broad is again hard to quantify. If the basin of attraction is broad, then σ should take a lower value. However, following this concept we will typically want a small σ across the majority of the potential. If the class is wide, we wish to reduce σ to see if a new class emerges. If the class is already narrow, it seems illogical to artificially make it broader. A further approach to adjust σ may be to smoothly oscillate its value across a range so as to emulate a circadian rhythm. Alternatively we may select values at random, possibly with a prescribed probability distribution. This should permit narrow sub-classes to form, whilst ensuring broader classes exist which guarantee effective gradient descent. Although we have focussed on a Gaussian function (equation 2.2) to represent a stimulus, we may consider how other functions may perturb the potential $V(t, x)$.

2.4 Extending the Model to N-Dimensions

In previous sections the basics of the system were demonstrated for a one-dimensional stimuli. The system is not limited to this restricted case, rather, stimuli may be represented by any N-dimensional input. As such, the potential energy function $V(t, \mathbf{x})$ may be considered as an N-dimensional profile with trajectories evolving in accordance with the N-dimensional vector field of the system. In [74] Janson considers a four-dimensional system to model the learning of temporal patterns. In addition to pattern learning, each of the N dimensions may be utilized to represent a different parameter of a learning goal. Consider a system that performs facial recognition using fuzzy data. Each dimension may encode information about a distinct parameter, such as: ‘eye-colour’, ‘skin tone’, ‘hair colour’, ‘hair length’, ‘facial dimensions’, ‘piercings’, ‘glasses’ etc. In this way the model builds an internal representation of the subjects under study. New stimuli are prescribed to classes, or if unknown, a new class develops. Following on from the one-dimensional demonstrations, two-dimensional (i.e. the potential is a function of time and $\mathbf{x} = (x_1, x_2)$) models can also be portrayed. Combinations of the stimuli prescribed in equation (2.9) are utilized in figures 2.7, 2.8 and 2.9 to characterize learning. We are again able to compare the profile to the PDD functions of stimuli, and also consider the evolution of initial conditions under the influence of the systems vector field. In each of the figures we have included a contour plot at the base of the potential energy to help illustrate the local minima of the system. Because of the dimensionality of the system we have only represented the state of the potential at finite time intervals. Readers should be aware that in reality the profile evolves smoothly over time. This smoothness is easier to observe from the included recognition plots. In figures 2.7g, 2.8g and 2.9g we demonstrate recognition via the evolution of a cross-section of initial conditions as they locate the local minima (fixed points) of the system (black circles). If we change the parameter γ we may increase/decrease the rate at which local minima are located. In figures 2.7f, 2.8f and 2.9f we show the normalized probability density distribution of each stimuli. Comparison with each potential demonstrates that the same concepts apply regardless of the dimensionality of the system. The potential converges to the shape of the negative PDD of stimuli. Figure 2.7 demonstrates the case where both η_1 and η_2 are described by double well polynomials, in this case the stimuli is generated by the application of equation (2.8) where both $\frac{dF_{1,2}}{d\eta}(x)$ are given by equation (2.9a). The second demonstration (Figure 2.8) illustrates the case where η_1 has a five well distribution and η_2 has a three well distribution. This is accomplished by the same stochastic differential equation, however, in this circumstance $\frac{dF_1(\eta)}{d\eta}$ is described by equation (2.9c) and $\frac{dF_2(\eta)}{d\eta}$ by equation (2.9b). The final demonstration

(Figure 2.9) is again based on the same principal, however, $\frac{dF_1(\eta)}{d\eta}$ is given by equation (2.9d) and $\frac{dF_2(\eta)}{d\eta}$ by equation (2.9e).

Although the model may be considered in higher dimensional spaces it becomes more difficult to effectively illustrate. In Chapter 3 we shall consider how this model may be utilized to learn colour as defined by the RGB colour palette. In this circumstance the dimensionality of the system becomes an obstruction to intuitive graphical representation. We can however present enough evidence about the state of the system to navigate this problem. The extension of the model to higher dimensions should be simple to execute and offers an effective solution to learning and storing complex stimuli, despite the issues with visualization.

2.5 Summary and Conclusion

Having introduced Janson and Marsden’s model of unsupervised learning we have numerically simulated the model’s behaviour for a number of different stimuli. Stimuli may be generated with prescribed probability density distributions as shown in [99]. Here we have generated six different stimuli based on a range of functions; each with very different shapes. First we considered a basic double well distribution, before increasing the complexity to consider a stimuli with five local minima. In addition to a one-dimensional stimuli we have also considered the case where η is a two-dimensional vector. There appears to be no obstacle, beyond computational power, that should prohibit the system functioning effectively for any N-dimensional stimuli. This feature is attractive as it should allow the model to represent highly complex classes of information in an efficient way.

If we contrast this concept with other unsupervised learning routines then the advantages should be apparent. Rather than requiring special functions/algorithms to derive relations between received stimuli, clusters/classes of retained information are formed automatically. Two stimuli that are closely related will share a fixed point with a basin of attraction related to the potential of the model. The local minima of the potential (center of the class) will be adjusted with the application of each new stimulus. This is computationally simple in comparison to typical unsupervised clustering models [87]. Generally such models require the computation of distance metrics between stimuli to locate nearest neighbours before clustering is considered. The memory foam model has the advantage that the system can immediately access new information to recognize any new stimulus. The system does not function in two distinct phases, unlike most learning models everything occurs autonomously, in what may be called an ‘on-line’ operation.

One area for concern is that the Gaussian width parameter (σ) may play an important role in determining the number of classes/sub classes that can be formed by the model. Although a narrow Gaussian width ensures that a large number of classes may be permitted, when stimuli are well separated the smoothness of the system may be impaired. This is an issue as there is then the possibility that each new stimulus may form its own class rather than clustering with other stimuli. Although this is desirable in some circumstances, we require the system to have some ability to generalize. As such the value of σ should be carefully considered. As a means to negate this problem we have explored various functions that could be used to determine σ . This includes making σ a function of time, or utilizing retained knowledge as described by $V(t, \mathbf{x})$.

In its current guise the memory foam model relies on gradient descent to associate a stimulus with a class. We have been dealing with a simulated stimuli, generated with a prescribed probability density distribution. In this format it seems irrelevant to select individual stimuli to recognize,

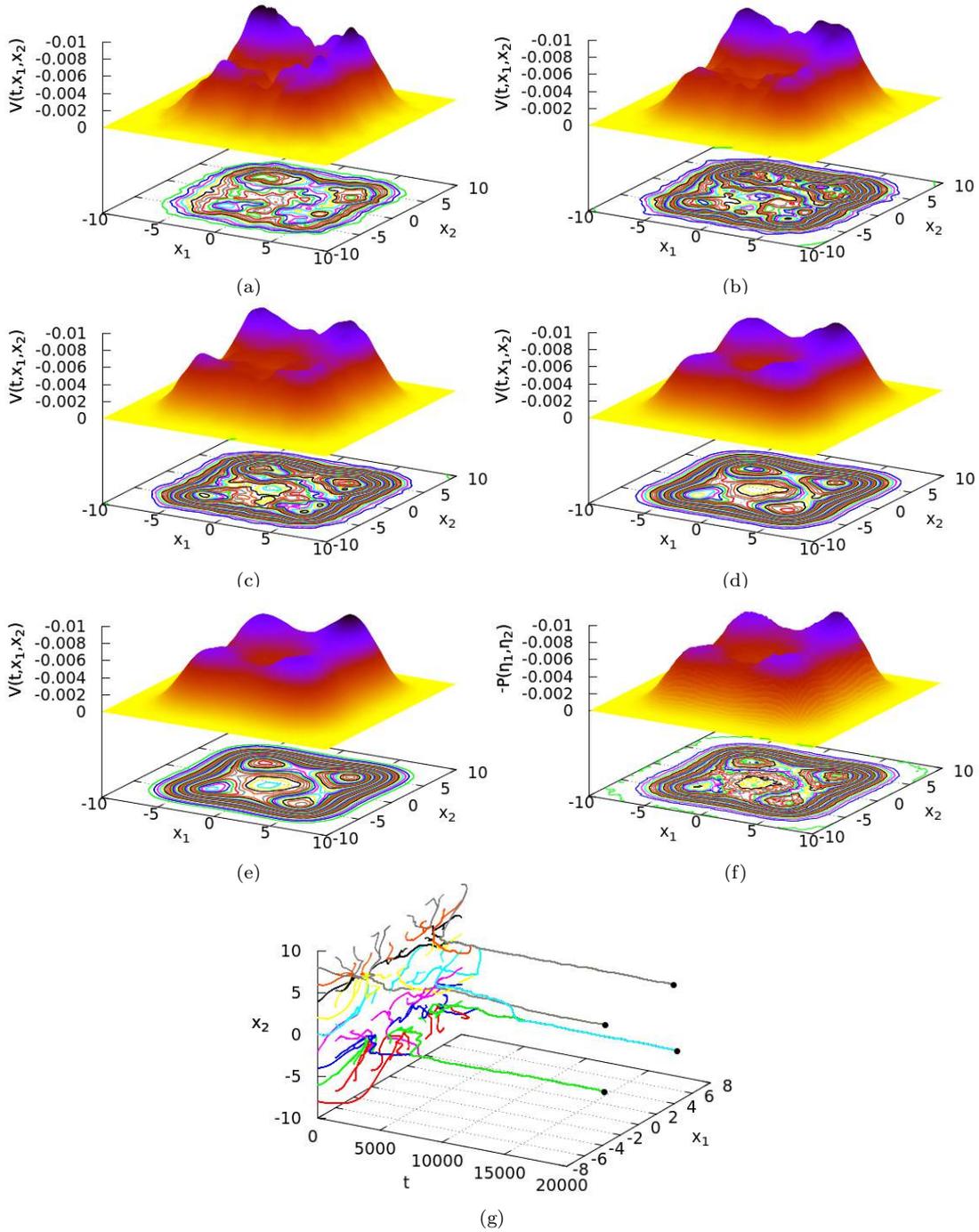


Figure 2.7: Evolution of $V(t, \mathbf{x})$ under the influence of a stimuli $\eta = \eta_1, \eta_2$. The stimuli are created using equation (2.8). In this instance the derivatives $\frac{dF}{d\eta}$ that prescribe the distribution of stimuli are given by equation (2.9a). Both η_1 and η_2 are associated with a double well polynomial distribution. Figures (a)-(e) demonstrate the energy potential at the time instances $t = 400, 800, 1600, 8000, 20000$. At the base of each profile is a gradient mapping highlighting the contours of the potential, from this it is possible to appreciate the locations of the local minima. Figure (f) illustrates the probability density distribution of the inputs. Again a gradient mapping is included on the plane x_1, x_2 . The potential $V(t, \mathbf{x})$ is seen to converge towards the shape of the PDD. Figure (g) demonstrates the concept of recognition by illustrating the evolution of 81 initial conditions interspersed across the plane x_1, x_2 . These trajectories move towards the local minima of the energy potential. Note that $\sigma = \sqrt{0.1}$, $\gamma = 1$ and $D = 5$.

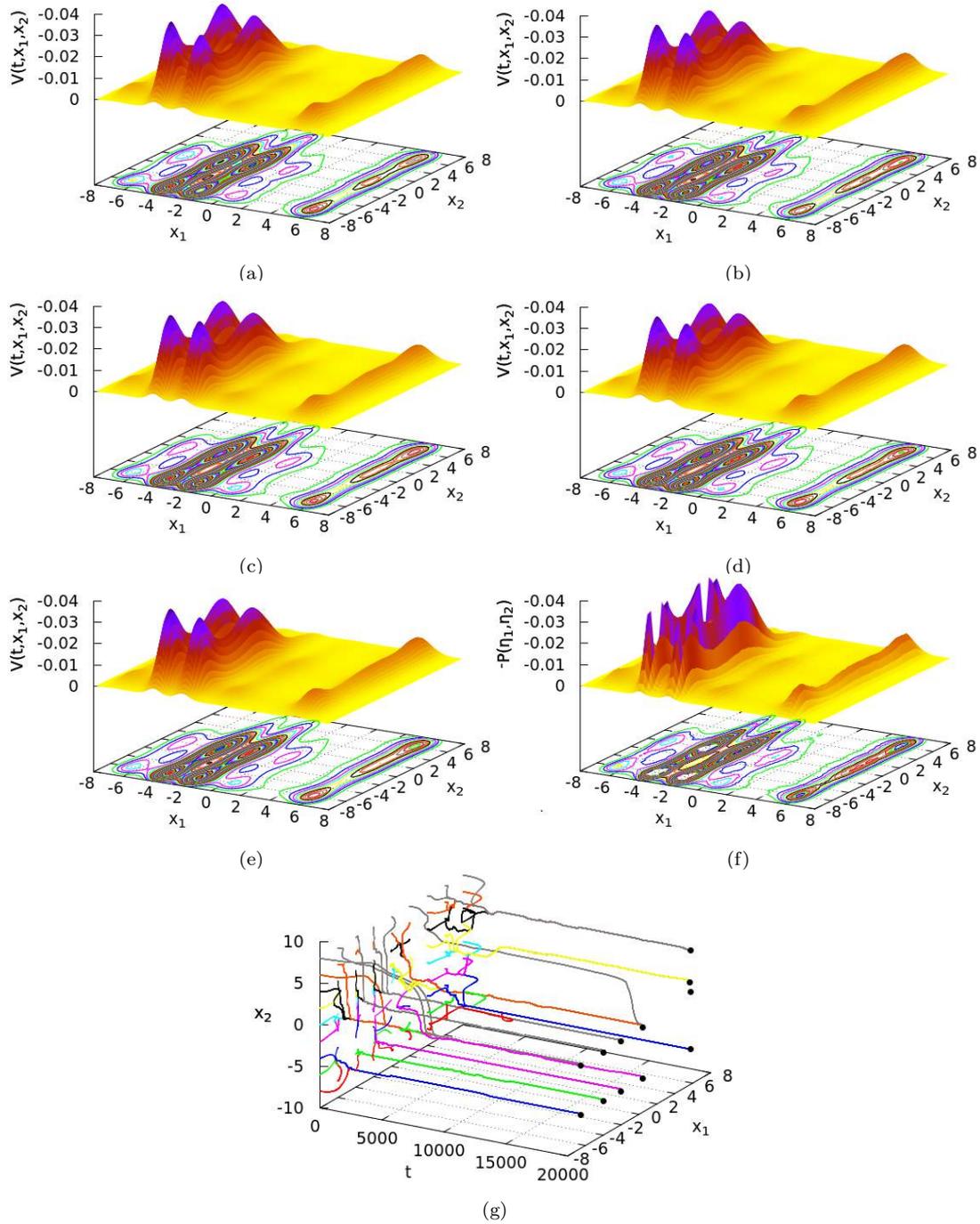


Figure 2.8: Evolution of $V(t, \mathbf{x})$ under the influence of stimuli $\eta = \eta_1, \eta_2$. The stimuli η are created by the application of the differential equation (2.8). The distribution of η_1 is described by equation (2.9c), whilst the distribution of η_2 is related to equation (2.9b). Figures (a)-(e) demonstrate the energy potential at the time instances $t = 2000, 4000, 8000, 12000, 20000$. At the base of each profile is a gradient mapping highlighting the contours of the potential, from this it is possible to appreciate the locations of the local minima. Figure (f) illustrates the probability density distribution of the inputs. Again a gradient mapping is included on the plane x_1, x_2 . The potential $V(t, \mathbf{x})$ is seen to converge towards the shape of the PDD. Figure (g) demonstrates the concept of recognition by illustrating the evolution of 81 initial conditions interspersed across the plane x_1, x_2 . These trajectories move towards the local minima of the energy potential. Note that $\sigma = \sqrt{0.1}$, $\gamma = 1$ and $D = 5$.

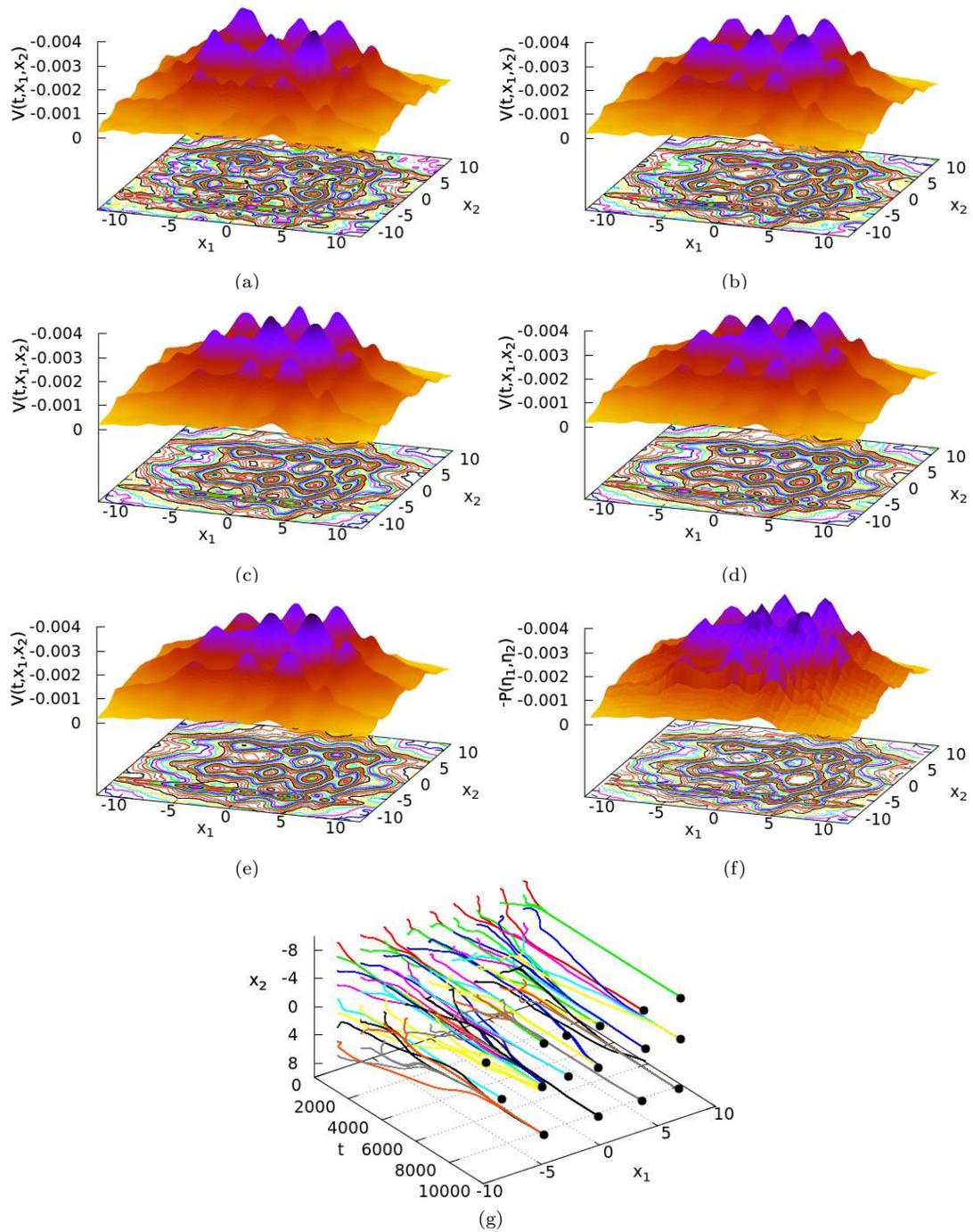


Figure 2.9: Evolution of $V(t, \mathbf{x})$ under the influence of stimuli $\eta = \eta_1, \eta_2$. The stimuli η are created by the application of the differential equation (2.8). The distribution of η_1 is described by equation (2.9d), whilst the distribution of η_2 is related to equation (2.9e). Figures (a)-(e) demonstrate the energy potential at the time instances $t = 2000, 4000, 8000, 12000, 20000$. At the base of each profile is a gradient mapping highlighting the contours of the potential, from this it is possible to appreciate the locations of the local minima. Figure (f) illustrates the probability density distribution of the inputs. Again a gradient mapping is included on the plane x_1, x_2 . The potential $V(t, \mathbf{x})$ is seen to converge towards the shape of the PDD. Figure (g) demonstrates the concept of recognition by illustrating the evolution of 81 initial conditions interspersed across the plane x_1, x_2 . These trajectories move towards the local minima of the energy potential. Note that $\sigma = \sqrt{0.1}$, $\gamma = 2$ and $D = 5$.

rather, we have demonstrated the concept by introducing a cross-section of initial conditions and observing their evolution in conjunction with the vector field of the model. In the next chapter we will deal with stimuli selected from a real world environment. In such a circumstance recognition can be demonstrated in a more organic manner, for instance, the model may observe a colour it has never before encountered and may recognize it as a shade of red.

Thus far we have only considered the functionality of this model for simulated stimuli. Clearly, if this model is to be adopted as a paradigm of learning its relevance to real world environments should be investigated. Janson and Marsden previously demonstrated one such example with the classification of musical notes and chords [74]. Alternative learning models have long been utilized across a broad spectrum of real-world tasks. These tasks include image and speech recognition [75, 127, 77], clustering of large data sets to identify associations [112], and even playing mind board-games such as checkers [96]. To emphasise the scope of this model we shall subsequently discuss the learning of colours, a typical task associated with computer vision.

Chapter 3

Non-Linear Dynamics of Learning and its Application to Colour Recognition

This new dynamical system that learns via the shaping of its plastic vector field is applicable to a number of real-world learning tasks. Janson and Marsden have already described one such investigation, related to the learning and recognition of musical notes and phrases [74]. This study presented an interesting initial test, however, a more complex stimuli should now be studied. At this juncture we consider the recognition of colours in a changing unsupervised environment. Starting from a flat vector field, we postulate that the system will develop classes that are related to the colours observed by the system. Given sufficient time and inputs the system will be able to recognize unfamiliar colours in relation to what it has previously learnt. To test this hypothesis we construct a simple experiment composed of a web-camera and computer. The web-camera observes the environment and the colours perceived are relayed to the model in the form of RGB triplets. As the RGB scheme is composed of three components, the potential of the system is a function of three spatial co-ordinates plus time. Utilizing this arrangement we are able to place the web-camera to observe any backdrop. By situating the camera in the proximity of an LCD computer screen, the system is able to learn from highly varied stimuli. In this Chapter we first consider a stimuli composed of known colours, this allows us to compare the classes of the system to a known input. After this, we compare the performance of the system for two cartoons, each composed of very different colour distributions. It is important that we investigate this new model in a variety of different situations so as to understand its range of applicability. The use of colour stimuli may be considered a more complex, real-world, relevant example, than the previous demonstrations. The challenge of learning colours autonomously presents a good opportunity to compare the performance of this model to human learning. It should be clear that a human can only learn from what it experiences or is taught. Likewise, the memory foam model should only form classes that correspond to the environment it has observed. If the environment only contains a limited colour palette, our model should only contain a restricted number of classes. Any RGB colour should be related to a corresponding attractor in accordance with the systems vector field. The learning of colours in a supervised systems would appear to be an exceptionally time consuming task. Consider providing training examples to teach an artificial system, each

colour would have to be presented in turn and labelled accordingly. In contrast, the automatic learning of colours should proceed rapidly. Related colours should form classes without the need for external interference. The memory foam model allows for continuous learning so the system can continuously adjust to any changing circumstances.

3.1 Colour Perception

If we are going to compare the performance of our artificial learning system to a human it is important to quantify the performance of the receptors involved. Within the animal kingdom the ability to perceive colour is incredibly species dependent, it can even vary significantly intra-species. The eye is a highly complex organ responsible for colour identification as well as all other faculties of sight. Light captured by the eye causes excitatory and inhibitory action potentials, these pass down the optic nerve to the brain. Within the brain's visual cortex action potentials are interpreted, this forms the basis of sight. Colour perception is largely affected by the type and number of photo-receptive cells found within the fovea, part of the retina, within the eye. These cells are classified as either rods or cones. Colour vision is the primary concern of cone receptors, different types being responsive to different wavelengths of light. It is the distribution, number, and types of these receptors that can vary substantially, this potentially leads to differences in colour perception. We should note that an exact understanding of the mechanism by which we interpret colour remains elusive [57]. Differences in perception may also be attributed to processes occurring within the brain's visual cortex.

The human eye contains as many as six million cone cells of 3 principal types; red (long), green (medium) and blue (short); their names indicate the regions of the spectrum to which they are sensitive. Inputs to these receptors are combined to identify colours. The typical trichromatic human is able to observe approximately 1 million colours, however, deficiencies exist which prevent some individuals from distinguishing between certain colour groups. An inability to distinguish between two colour groups is referred to as dichromatism and the afflicted person is said to be colour-blind. Approximately 8% of US males with Northern European ancestry had red-green colour blindness as of April 2011 [116]. Dichromacy significantly limits the range of perceivable colours, however, as many as 10,000 colours may still be discernible. If a person has only one functioning type of cone cell as few as 100 colours may be perceived. On the other end of the spectrum exist people with a 4th type of cone cell, sensitive to wavelengths between red and green. These individuals are referred to as tetrachromatic, they possess a potential colour palette of 100 million colours [76]. In total the average human eye has around 130 million photoreceptors, yet the optic nerve is composed of only 1 million fibres, this suggests that there must be some compression of signals before information is transferred to the brain [45].

The study of the eye and its interaction with the brain is an ongoing concern within biology, so too is colour perception [103]. Because of the subtle differences between individuals it would seem unlikely that we all observe colour in the same way. One suggestion is that different people may perceive colours differently but assign the same 'label' or 'name' to them [161]. Determining if this is the case presents a considerable problem as we cannot simply swap input device and compare results.

3.1.1 Machine Vision

Machine vision [71] can be seen as the artificial parallel of biological vision. As expected, it typically involves the acquisition of an image followed by the application of various image filtering techniques. Critical aspects of the image are identified and then utilized to drive some autonomous process. Images may be obtained by standard digital cameras or by any number of alternative imaging procedures; these may include more obtuse methods such as x-ray, electron microscopy or infra-red imaging techniques. The ability to distinguish between colours is sometimes limited by these methods, firstly by the hardware used and then by the algorithms applied. A machine vision model is not restricted to the realms of human vision and may incorporate stimuli from the entire range of the electromagnetic spectrum. Consider what advantages there could be for a learning system capable of understanding stimuli emitted within the infra-red range or ultra-violet regions. The world we see with our eyes is limited by the receptors we possess, increasing the range of sensitivity would completely transform the world we perceive. Equally, a reduction in sensitivity would have a similar effect. This should be considered when comparing the performance between natural and artificial systems. Machine vision currently has widespread applications within the realm of expert systems [106, 113, 34, 56]. It is utilized in industry for tasks such as inspecting machined parts for defects or making precision measurements and responding autonomously. Applications may also be found in situations where human action may be dangerous or impractical. A pertinent example of a system utilizing machine vision is a spacecraft, here numerous on-board functionalities are automated [106], this may even be extended as far as the landing procedure [113]. Machine vision is an expanding field that is expected to become intrinsic to the development of future technologies as we continue to apply automation to more and more applications. Modern day production has become synonymous with automated robots forming production lines, in future such production may become completely devoid of human intervention. We may also consider other fields such as farming or transport where automation may remove the necessity for the human component [163]. Developing autonomous vehicles is a current popular trend [34], this research area is being pursued by a number of developers including ‘Google’ [56].

Machine vision in the context of this work is presented in a relatively simple way. We demonstrate that the self-shaping learning system is able to learn about different colours. The system will recognize new colours as part of existing classes or develop new classes to represent unfamiliar stimuli. The input device used to observe the environment is a standard computer web-camera. From this RGB triplets are extracted and utilized to autonomously build an internal representation of colour.

3.2 Colour Recognition and Classification by a Dynamical System with Self-adapting Vector Field

Stimuli to the learning system are three-dimensional RGB triplets that define the colour of the central pixel of a web camera. These RGB triplets are identified using code published under a GNU general public license. Each component takes a value between 0-256, inferring a finite range of colours rather than a continuous spectrum. Thankfully, the discretized 16,581,375 distinct colours provide a more than adequate approximation of a continuous spectrum. Notice that the number of colours realized by this scheme far exceeds the level of distinction the average trichromatic human is able to perceive. For convenience we scale each component of the stimulus $\eta = (\eta_R, \eta_G, \eta_B)$

onto the range $0 \rightarrow 10$. Here $\eta_R = R$, $\eta_G = G$ and $\eta_B = B$ where R,G and B are identified by the camera. As the stimuli consist of three components, the system's potential must be a function of the vector $\mathbf{x} = (R, G, B)$ and time t . Note that R,G,B are also scaled onto the same range as the stimuli. This scaling allows us to continue to utilize a familiar range of the Gaussian width parameter σ . A larger range of \mathbf{x} would require us to up-scale σ accordingly. The self-shaping potential may therefore be described by the differential equation:

$$\frac{\partial V(t, R, G, B)}{\partial t} = -\frac{1}{t} \left(V(t, R, G, B) + \frac{1}{(\sqrt{2\pi}\sigma)^3} \exp\left(-\frac{(R-\eta_R)^2 + (G-\eta_G)^2 + (B-\eta_B)^2}{\sigma^2}\right) \right) \quad (3.1)$$

Note that we again omit any forgetting term, this is equivalent to setting $k = 0$ in equation (2.1). The potential $V(t, R, G, B)$ cannot be entirely visualized due to its dimensionality, we must therefore make an attempt to describe the systems composition in an alternative way. This is achieved by taking projections of the profile at discrete time intervals, considering the location of local minima and observing many recognition trajectories to approximate the size and shape of each fixed point's basin of attraction.

Ultimately we should be able to use the memory foam model in any real life environment. The problem with this for demonstration purposes is that only a few related colours are typically found in an isolated situation. The limitations of our hardware impose further constraints in regard to stimuli collection. To overcome these issues we position our web-camera to observe a standard LCD monitor. This is shown in figure 3.1. We can apply either a known controlled stimuli, such as a colour with a known RGB value (section 3.3), or allow the camera to observe a randomly selected video and learn about its colour composition.

3.2.1 Assessing the accuracy of RGB value identification

For our model to emulate human recognition of colours the identification of RGB triplets must be accurate. As discussed previously, colour perception is not uniform across populations and species, this fact should be considered when appraising this model. We should not require the colours recognized to match what we observe, rather, they should be compared to the colours and distributions that the model receives. This is dependent upon the camera. We may consider the camera as a person with an inferior level of visual acuity, or as a different 'animal' altogether. We may still consider the identification of RGB triplets and compare these to our own perceptions, this proviso is simply intended to reconcile any disagreement between man and learning model.

If we consider the LCD monitor to be correctly calibrated then we can display colours with prescribed RGB values. We may position the web-camera to observe the screen and use dedicated web-camera software to observe what the camera views. This setup is demonstrated in figure 3.2. In this instance the screen is set to shown a pure red (255,0,0). It is apparent that the colour in the left hand box, which is obtained using the web-camera software 'Camorama', is not exactly the same. In fact, this colour is close to the value identified by our code, a colour defined by the approximate triplet (254,65,0).

Any colour can be compared to both the RGB value identified by the code and also to a photograph of the display obtained by the web-camera. Figure 3.3 contrasts RGB values for a range



Figure 3.1: The web-camera is positioned in front of a LCD monitor. This allows us to select stimuli such as television shows that have a wide colour contrast. This static set-up allows us to optimize the position of the camera to obtain the strongest possible relation between a known colour and the colour identified by the camera (see subsection 3.2.1). This is achieved by manipulating lighting conditions and adjusting the distance between the camera and the screen.

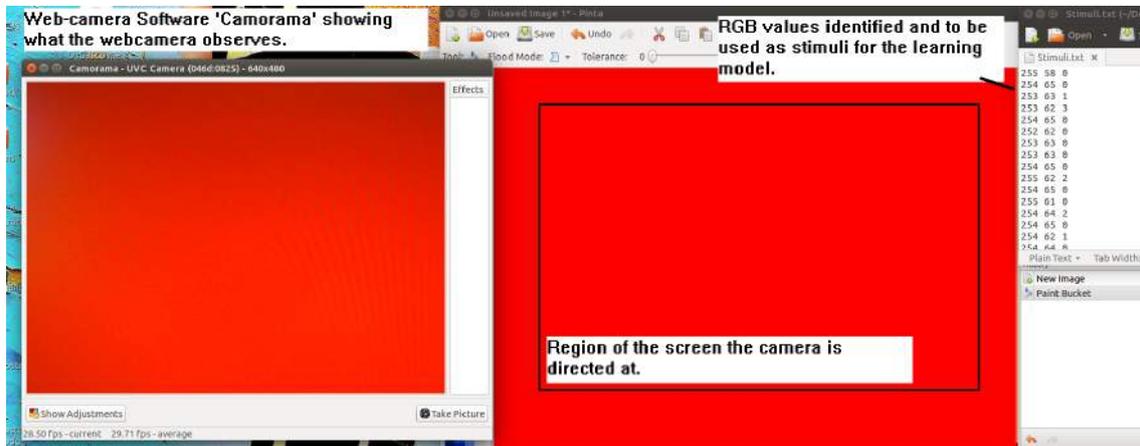


Figure 3.2: The camera is placed to observe the screen. On the screen is a colour with a prescribed RGB triplet. We can compare the colour of the screen, the colour seen by the camera, and the output of the code in a single screen-shot. A colour with a known RGB value, observed by the camera, is shown centre screen. The colour observed through web-camera software is shown to the left of the screen. The RGB triplet identified by the code is shown to the right of screen. It is apparent that the three colours are not identical. In this instance the RGB colour prescribed is 255,0,0. The camera appears to observe a colour closer to 254,65,0 as indicated by the code and validated by the web-camera software screen-shot.

of different colours. Note that light intensity and camera-screen distance have been manipulated to try to obtain an optimal relationship. It is evident from figure 3.3c that the RGB value identified by the code generally agrees with the camera observation. Unfortunately, figures 3.3a and 3.3b demonstrate that the colour identified by the hardware (observed on screen through web-camera software) does not always agree with the known screen colour. This difference is observed more severely at the edges of the range. Where a component is close to 255, the hardware typically underestimates this value. Components that should be close to 0 are typically over-estimated. As such, the colours recognized by the learning model will not exactly match those perceived by humans. This should not be seen as detrimental or inhibitory to the function of the model at this stage. In circumstances where such a result is undesirable greater emphasis should be placed on the selection and calibration of hardware, utilizing superior equipment is likely to incur significant expensive, but should remedy this problem.

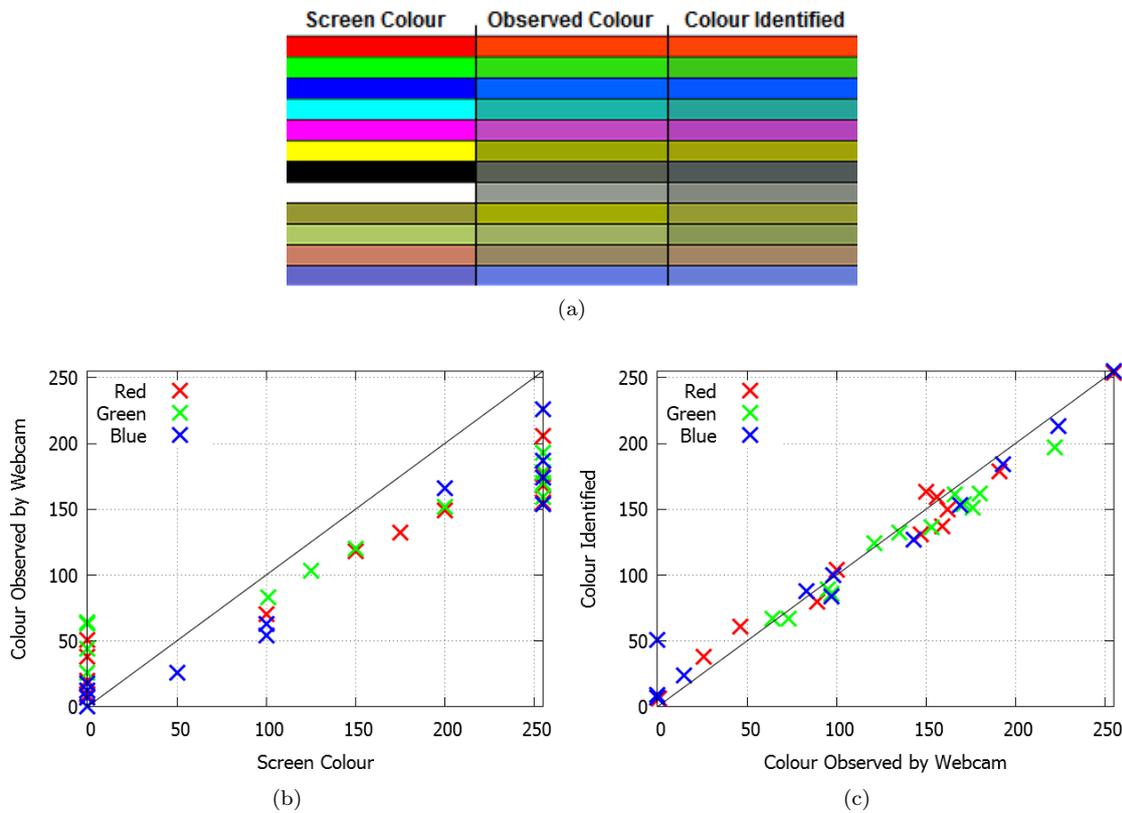


Figure 3.3: **(a)**: Comparison between the screen colour, the observed colour (picture taken by web-cam software) and output RGB colour (determined by code). **(b)**: A graphical comparison of the screen colour and the observed colour. The optimal case is described by the black line. **(c)**: Observed Colour against identified colour. Note that the colour of the marker matches the corresponding component of the RGB value. The term ‘colour identified’ refers to this output colour (this is the RGB triplet identified by our code).

3.3 Applying a Simple RGB Stimuli

A continuously looping video of pre-defined coloured slides can be observed by the web camera to provide a stimuli to the memory foam learning system. Although such an approach is rather contrived, it allows us to ensure that classes form as expected and colours are recognized accordingly.

We have considered how the camera and code interact to provide stimuli, we must now consider how this information is perceived by the learning model. This is an important initial test that is intended to validate the approach before we test the system for more complex stimuli. This method is particularly suited to provide an initial assessment as we should be able to infer which colours correspond to which local minimum of the potential, hence, we may compare recognized colours to known inputs.

The camera is positioned in front of the screen as shown in figure 3.1. Table 3.1 provides a list of the colours that the system experiences as part of the observed slide show. As explained in section 3.2, each component of the stimuli is scaled to take a value between 0 and 10. With sufficient application of the input, the classification landscape $V(t, R, G, B)$ should possess features that reflect the observations of the system. Initially the potential is completely flat ($V(t, R, G, B) = 0$), any local minimum of the potential that develops is therefore a result of the stimuli. Because our video is composed of eight distinct colours, we may expect eight well-defined local minima to form. We must stress that illustrating these local minima and the topology of the potential in their vicinity is a somewhat challenging problem due to the dimensionality of the system. As the colours selected to form the stimuli lie at the extremes of the RGB space, the minima of the potential may be expected to form at the extreme edges of $V(t, R, G, B)$. It is therefore advantageous to select a large value for σ , here we apply $\sigma = \sqrt{0.5}$. This ensures that the basins of attraction of the systems fixed points, corresponding to the potential's local minima, are quite wide. It is undesirable for regions of the potential to remain approximately flat, this would inhibit the rate of gradient descent. The time for which each colour is observed by the system is approximately equal, whilst the model undergoes a total of 500000 iterations. Here we assume that each iteration of the model takes approximately the same time. This is certainly not guaranteed and may explain why some regions of the potential contain deeper local minima.

Colour	(R,G,B)
Red	255,0,0
Green	0,255,0
Blue	0,0,255
Cyan	0,255,255
Yellow	255,255,0
Magenta	255,0,255
Black	0,0,0
White	255,255,255

Table 3.1: Colours used to create a contrived slide-show to be utilized as a stimuli source for the camera.

Projections of the system's energy potential at finite time moments are shown in figure 3.4. Following the discussion of the two-dimensional case in Chapter 2, the extension to three-dimensions (plus time) has the additional complication that even at finite time moments the entire potential cannot be observed. To overcome this problem we illustrate projections along the principal axis (R, G) , (R, B) and (G, B) . At the base of each projection we include a gradient mapping of the potential in order to highlight the location of each minimum. Although we show only three subsequent time moments, $t = 100$, $t = 300$ and $t = 500$, it should be stressed that this potential is a continuous, smooth function of position and time and therefore continuously evolves with the addition of new stimuli. We may also note that given a sufficient number of stimuli this profile may become approximately stationary. Comparing the state of the potential at each of the demon-

strated time intervals it may appear that the potential has indeed reach a semi-stable stable. We may consider this profile to act as a memory containing information about all the colours that form the stimuli. Other illustrations may also prove useful when attempting to understand the state of the system. Not least are the positions of the local minima of the potential within the three-dimensional RGB space. These are shown in figure 3.5. This depicts the different locations of the minima for twelve different time moments. In this instance the colour of each point corresponds to the location of the minimum in the RGB space and the size is related to the depth of the potential. Larger points are associated with deeper minimum. Based on this description it would appear that the main features of the system have emerged by $t = 50$, there is of course the possibility that the potential continues to shape, but the locations of the minima remain relatively static. Utilizing figures 3.4 and 3.5 in unison would appear to suggest that this is not the case.

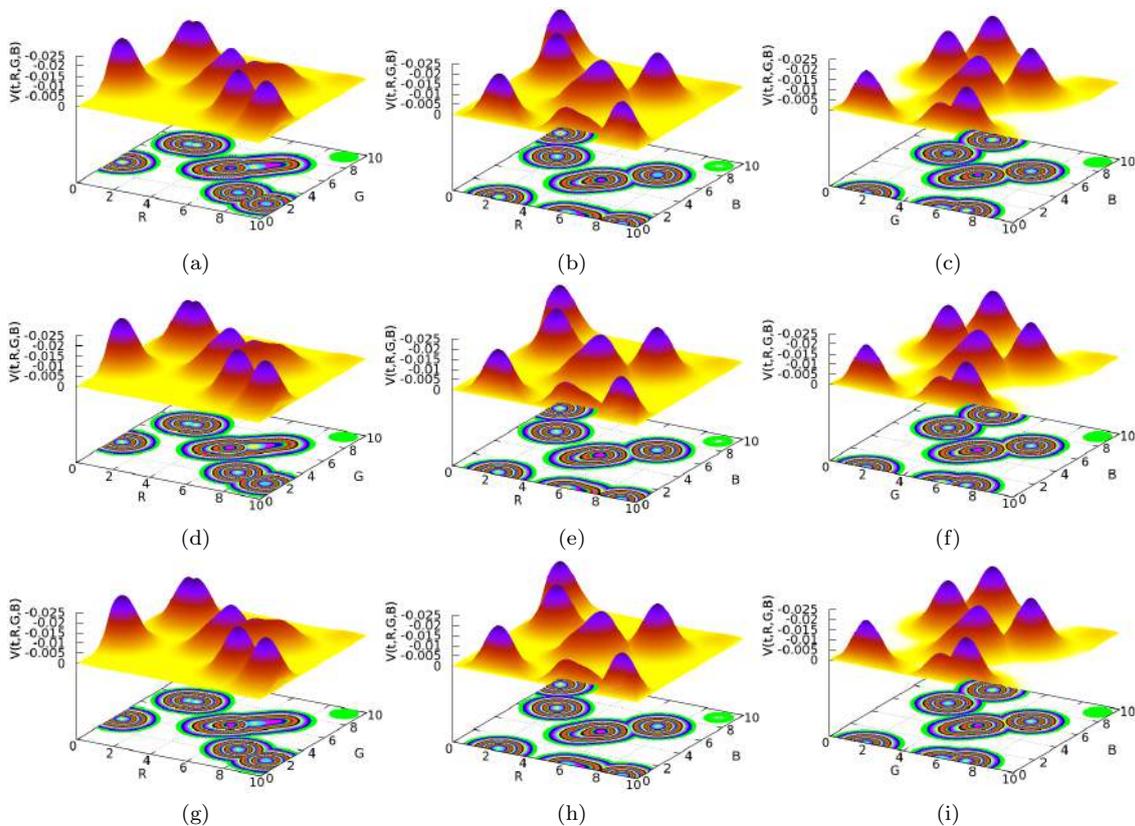


Figure 3.4: Subsequent applications of a three-dimensional stimuli cause the shaping of the memory foam systems potential $V(t, R, G, B)$. Figures (a),(b),(c) show projections of the three-dimensional potential onto the prescribed two-dimensional space at $t = 100$. Figures (d),(e),(f) and (g),(h),(i) show the same projections at the subsequent time moments $t = 300$ and $t = 500$. Each figure also includes a gradient mapping at its base in order to illustrate the locations of the local minima and the local topology of the potential on the plane. The parameter $\sigma = \sqrt{0.5}$, this ensures that the basins surrounding each minimum remain relatively large.

As we may surmise from figures 3.4 and 3.5 the potential contains information relevant to all stimuli by $t = 500$ (in fact this appears to occur much earlier). We may therefore focus on describing the system's state at this particular time moment. Perhaps the simplest way to understand the state of the system is to consider the attractors illustrating recognition. The gradient description

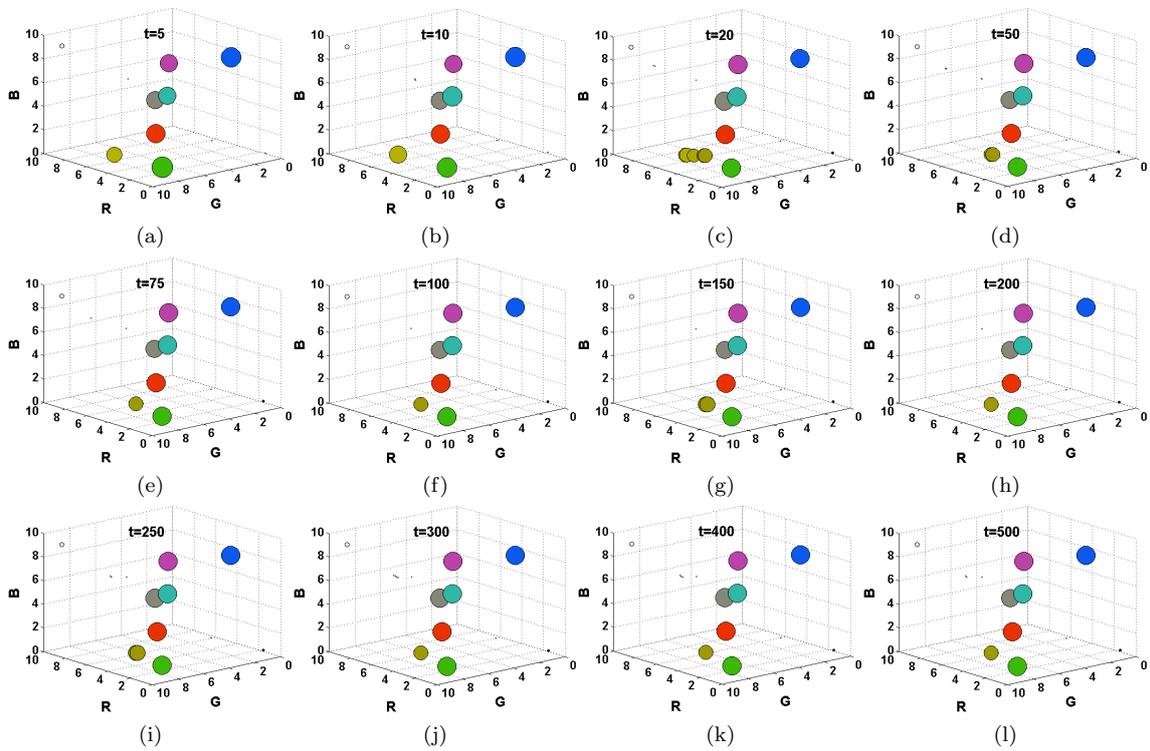


Figure 3.5: Positions of the local minima of the energy potential function $V(t, R, G, B)$ at static time intervals as denoted on each figure. The colour of each point corresponds to the position in the RGB space. The size of each point is scaled to the value of $V(t, R, G, B)$, larger points correspond to deeper minimum of the potential.

of recognition causes each local minimum of the potential to act as an attracting fixed point. By observing the evolution of a mesh of initial conditions from across the RGB space we not only demonstrate the recognition of a cross-section of stimuli, but also gain a vague appreciation of the different sizes of each fixed points basin of attraction. The sizes of these basins are related to the shape of the potential. We may also consider the local maximas of the potential which may be considered to act as separatrices of the system.

The recognition of a cross-section of different stimuli, each with different initial conditions within the RGB space is shown in figures 3.6a and 3.6b from two different perspectives. Here we only consider the evolution of trajectories in conjunction with the stationary potential as it appears at $t = 500$. This recognition of stimuli may also be visualized as a colour spectrum comparison. Figure 3.6c shows the same stimuli as contained in figures 3.6a and 3.6b. The top row corresponds to the colour of the stimulus whilst the bottom row demonstrates the classes recognized.

Notice that the system does not exactly replicate the colours expected. As the stimuli only contains eight colours it is interesting that eleven colours are recognized. These eleven classes may be shown to correspond to the minima of the potential. As this potential remains relatively stationary over the interval $t = 50 \rightarrow 500$ this property cannot be attributed to receiving an insufficient number of stimuli from which to learn. This difference can be reconciled by consulting the actually RGB triplets inferred by the camera. As shown in section 3.2.1, the colour observed by the camera is not always directly related to the screen colour, it is not uncommon for some differences to arise. We should therefore consider the actual stimuli that are received by the learning model. This information is inferred from figure 3.6d, which indicates the maxima of a three-dimensional histogram built using the stimuli received by the system. Clearly such a profile is not smooth, hence the high number of points. Larger points correspond to greater local maxima of the probability density distribution landscape. Recall that we anticipate the memory foam profile to converge to the shape of the negative probability density distribution of inputs (given small enough σ and sufficient inputs), as such we should expect the positions of these local maxima to correspond to the local minima of the potential. Although the camera is only presented with 8 colours it has perceived several more. This may be due to factors such as changing lighting conditions or a lack of exposure time. It is clear that a strong relation exists between the actual colours of the stimuli (figure 3.6d), the shaping of the potential (figures 3.4 and 3.5) and the colours recognized by the system (3.6a, 3.6b and 3.6c). Each visualization infers that the system has learnt about a set of colours that are defined by the stimuli. The system has autonomously compiled these into classes and will recognize unfamiliar stimuli based on what it has learnt.

The difference between the colour the camera is shown and what it perceives is a problem that will clearly persist throughout this study. As previously indicated, we do not need to be concerned with this. We may consider this particular camera to be equivalent to a person suffering from some minor form of visual impairment. Changing to a higher end model may reduce or even negate this disagreement. Consider a person suffering from colour-blindness, they are still able to learn about the different colours that they see, they just see things differently to other people. Colour-blindness often has little effect on how people interact with their environment, as their knowledge of colours is specific to them. Where one person always perceives red, another may see green. In most situations this does not result in any negative consequence.

Advancing from this contrived demonstration, it is now logical to consider a more complex stimuli. The world of cartoon is characterized by a wealth of vibrant colours, making cartoons

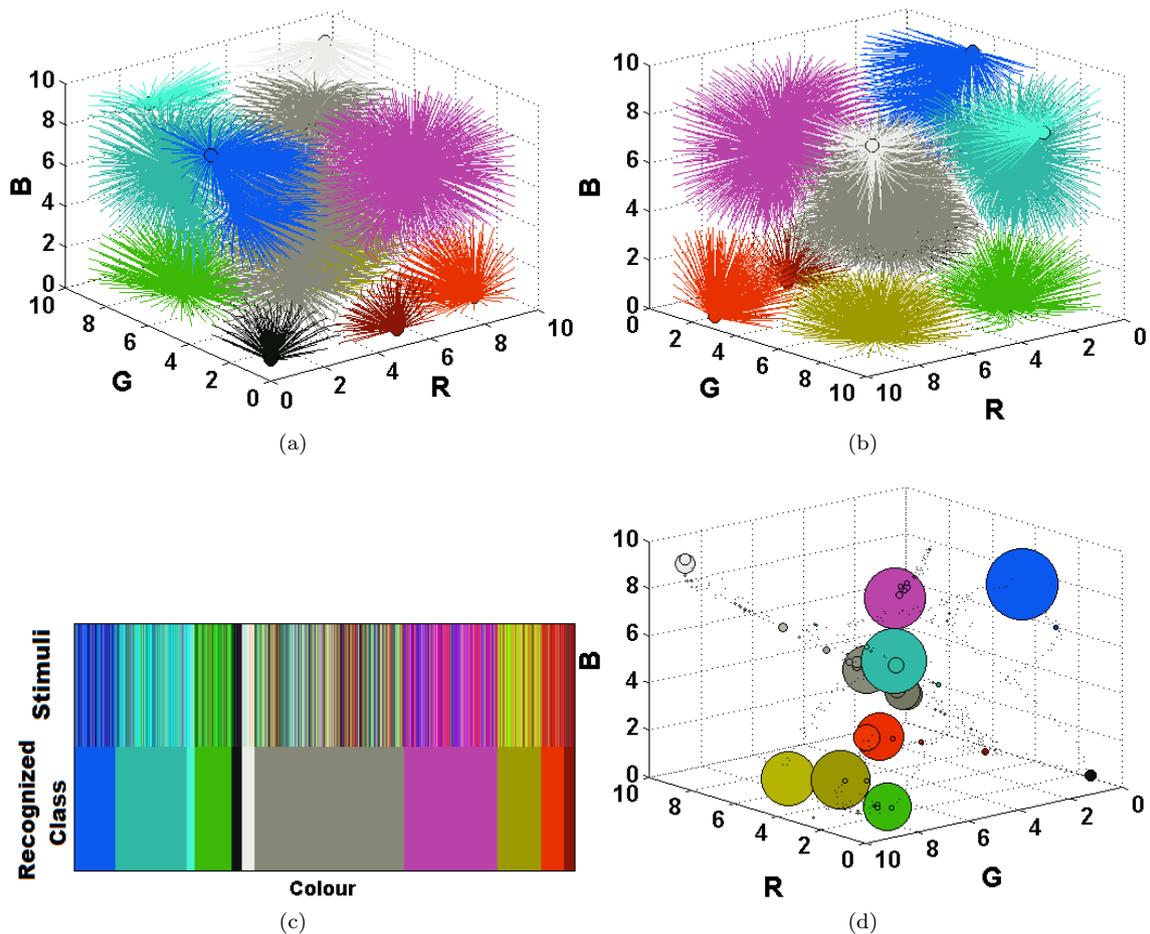


Figure 3.6: Illustrations of the formed classes and the recognition of stimuli by the system. (a),(b): Different perspectives of a mesh of evenly interspersed initial conditions as they evolve towards the attracting fixed points of the stationary profile. These dynamics are in accordance with gradient descent. Line colour infers the final class that the trajectory reaches. This indicates the recognition of a stimuli corresponding to each initial condition. The size of each class may also be approximately inferred from the size of each coloured region. (c): An alternative illustration of recognition. The initial stimuli are demonstrated on the top row whilst the class to which they are assigned is shown on the bottom. (d): The maxima of a three-dimensional histogram describing the distribution of stimuli. Point size is dependent on the probability density at each location. Larger points therefore correspond to more prevalent stimuli. The colour of each maxima corresponds to the location in the RGB space.

such as ‘Futurama’ and ‘Southpark’ ideal candidates for further exploration of this approach.

3.4 Learning Colours by Observing Cartoons

The previously discussed approach can be applied for any stimuli that comprises RGB information. Next we shall demonstrate the approach for inputs taken from two well known cartoons. We shall again shown how the memory foam model autonomously creates an internal representation of its experiences and draws upon this information in order to recognize new colours. Like a human learner the system may only learn from its surroundings. We should point out that learning colours is generally an unsupervised routine in humans. We observe our environment and remember aspects of what we have seen. To learn about colours via a supervised approach would appear implausible, it would certainly seem an automatic task that does not require outside interference or complex reasoning. Likewise, the memory foam model applies unsupervised learning to store information about the colours perceived by a web-camera. As with humans, the system is able to continuously learn about new colours as new stimuli are observed. In these two demonstration the system receives two million stimuli, each one defining a colour that composes a cartoon. To teach a supervised system would therefore require two million training examples, learning would take an exuberant amount of time and would also fail to replicate the mechanism that appears to occur in humans.

It is important to consider the behaviour of the system for a range of different stimuli. We may anticipate that the colour compositions of these two cartoon are significantly different. By considering two examples we shall demonstrate that the evolution of the system is highly dependent on the environment.

3.4.1 Southpark

The cartoon Southpark has been selected as our first example due to the rich vibrant colours of which it is composed. Selecting a single episode and retaining the camera and screen in constant relation to each other the video was played in a loop until 2000000 iterations of the model had been exceeded. The potential is initially flat and hence contains no pre-existing information ($V(R, G, B, t) = 0$). Two million subsequent stimuli equates to a model run time of $t = 2000$. To ensure that the systems vector field characterising recognition is smooth, the parameter σ is chosen as $\sqrt{0.5}$. As with the previous contrived example, we represent the development of the systems attractors via a variety of different illustrations.

Let us first consider the recognition of a set of stimuli. Figure 3.7a depicts the recognition of a set of periodically selected stimuli that are encountered by the system as its potential is shaped. These stimuli provide the initial conditions for trajectories that converge towards the attractors of the system. The attractors of the vector field correspond to the local minima of the potential which are marked as black circles in the RGB space. We may also illustrate this information in the form of a colour comparison as shown in figure 3.7b. The colours on the top row correspond to the observed stimuli, whilst the colours on the bottom row demonstrate the class to which the colour is assigned. Note that we have manipulated the order of stimuli from left to right so that all stimuli assigned to the same class appear together. This allows us to compare the range of colours that the system may consider to be similar. As the recognition of colour is dependent on the evolution of the potential, it is possible that the same stimulus may be recognized differently

when encountered again at a later time moment. This is because the vector field of the system continuously changes with the application of more stimuli. The same process may be inferred for a human learner as more experience is gained. Initially we may only draw a distinction between large classes of colour. For instance, a learner may initially only recognize colours as belonging to one of the classes ‘red’, ‘green’ or ‘blue’. With more experience we may learn about the existence of other classes, such as a cyan or magenta. Where a stimulus close to magenta was previously assigned to ‘red’, we now know that a superior classification is available. We may anticipate that where two identical stimuli are recognized by the memory foam model at subsequent time intervals, the recognition in the second event is likely to prove more relevant to the systems experiences than when encountered the first time.

An attractor of the system is not guaranteed to exist for any prescribed period of time as the potential evolves. Only once/if the potential reaches a stationary state may we infer that the attractors of the system will endure. This has important ramifications on the recognition of stimuli as a class may persist for a long period of time before disappearing. Where this is the case, gradient descent will cause the corresponding trajectories to locate a different attractor of the system. There is always some uncertainty regarding whether the recognition of a stimulus is indeed optimal. Only when the system has reached a stationary state before a stimulus is encounter may we be confident that the relation between stimulus and class will remain for all time.

As we would expect, some colours appear much more prevalently in this cartoon than others. It would appear from figure 3.7b that the most prevalent colour is grey. Some stimuli that may appear to be more related to the classes blue and yellow also fall into this attractor. These alternative classes may not have formed when these stimuli were first encounter, or these particular shades may actually more closely align with grey due to the inherent structure of the RGB scheme. It should be expected that the model will recognize some colours differently to a human. It appears to be a problem for different people to agree on how to define a colour in some circumstances [93]. Consider the colour turquoise (48,213,200), different people would argue that this is a colour more closely related to blue than green or vice versa. We should therefore be careful to avoid contradicting the models recognition of a colour without first consulting the stimuli that it has experienced and the configuration of the systems potential.

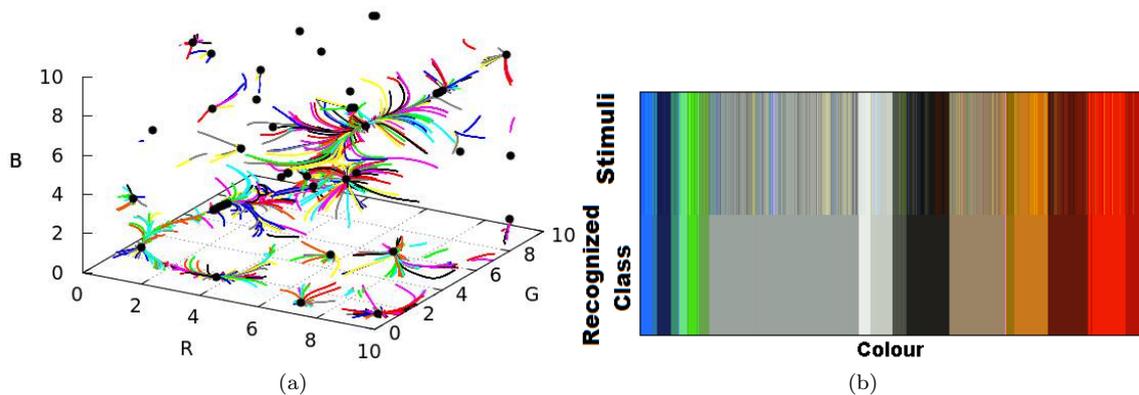


Figure 3.7: Recognition of 2000 periodically selected stimuli. **(a)**: Evolution of initial conditions towards the local minima of the energy potential (black circles). **(b)**: Comparison of the initial condition (stimuli)(top row) and the recognized colour (bottom row). The recognition rate parameter $\gamma = 5$.

The assertion that grey is the most prevalent colour within this cartoon is supported by the

projections of the potential at finite time intervals on the principal axes of the system. These projections are shown in conjunction with a demonstration of the local minima of the system in figure 3.8. Each row of this figure corresponds to a different time interval ($t = 400, 800, 1200, 1600, 2000$). It is apparent that the local minima of the system remain relatively stable and constant across the time scale investigated. This suggests that the formation of classes occurs relatively quickly in this example. We may discern that the basic structure of the system has already been composed by $t = 400$, beyond this we do observe some fluctuation in the size, number and shape of classes, however, the most eminent features remain stable. More projections of the potential at earlier time intervals are included in appendix figure A.1. These images highlight the changeable nature of the potential between $t = 5 \rightarrow 200$.

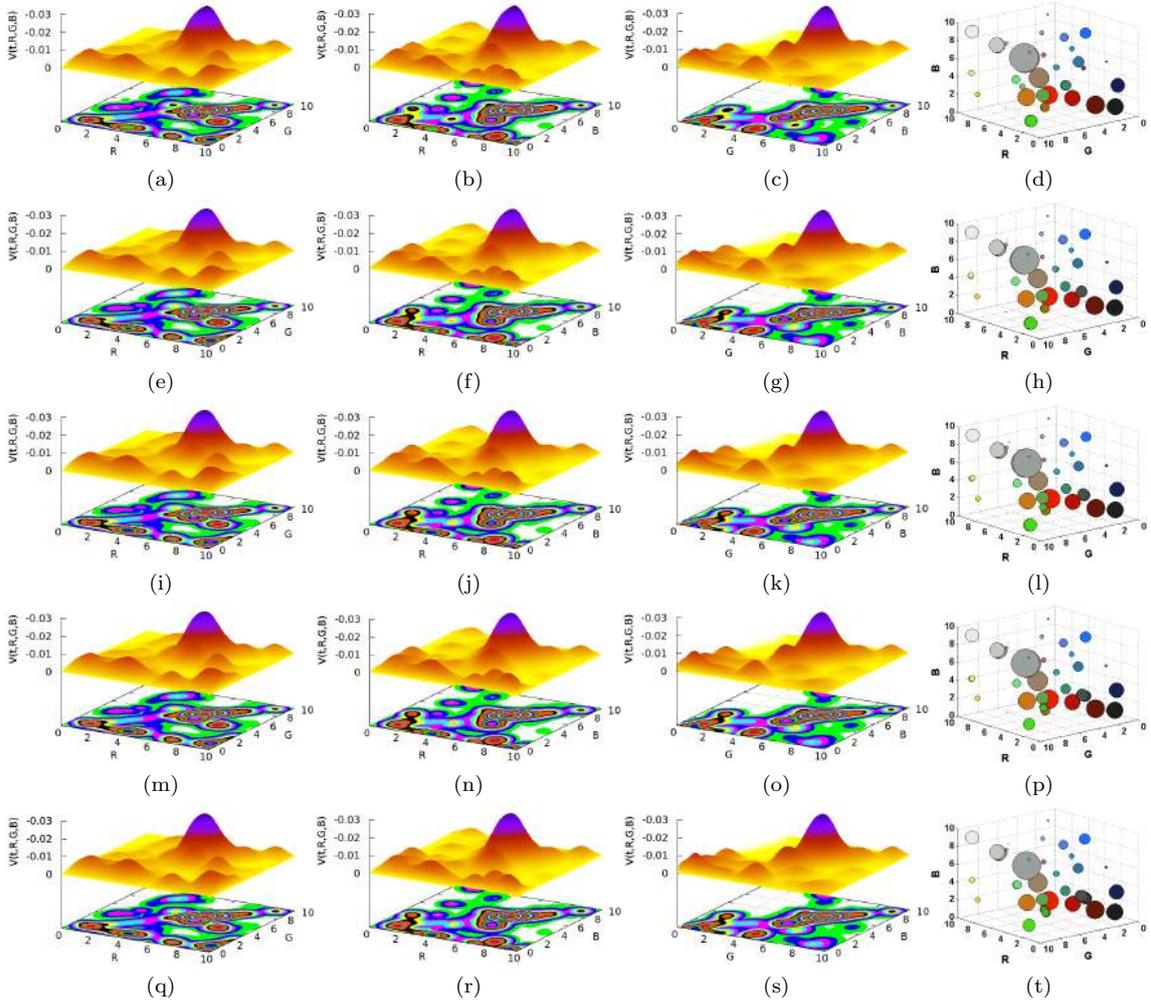


Figure 3.8: Projections of the energy potential $V(t, R, G, B)$ at time instance $t = 400, 800, 1200, 1600, 2000$. Each row relates to a different time moment. The first three columns demonstrate the projections of the potential on the axes (R, G) , (R, B) and (G, B) . At the base of each graphic is a gradient mapping of the contours of the potential. The final column illustrates the locations of the local minima of the potential in the RGB space. The colour of each minima indicates the location of the point in the space, whilst the size is scaled to the magnitude of the function $V(t, R, G, B)$. The Gaussian width parameter $\sigma = \sqrt{0.5}$ throughout.

For the demonstrations of the local minima the depth of the potential at the minima is again

related to the size of the point, whilst the location in the RGB space corresponds to the colour. Although this provides an indication of the number of classes, it does not demonstrate the size or shape of each class. This information is also hard to infer from the projections of the potential, although some characteristics can be seen with careful consideration. A deep potential at a local minimum does not necessarily infer that the fixed point will have a large basin of attraction or vice versa. To intimate the shapes and sizes of each basin we may introduce a dense mesh of initial conditions and again consider recognition. Figure 3.9 illustrates the recognition of 1000 trajectories that converge to the fixed points of the system in accordance with the static vector field of the system as it exists at $t = 2000$. Gradient descent in conjunction with a static potential may be seen to infer the approximate sizes of each fixed points basin at this time moment. The same information is provided in both figures 3.9a and 3.9b but from two different perspectives. As was shown in figure 3.7b, the same information may be presented as a comparison between initial input colour and the colour of the class recognized. This procedure is again replicated in figure 3.9c. The top row corresponds to the range of initial conditions, whilst the bottom row shows the assigned classes. From these illustrations it is evident that grey does indeed have the largest basin of attraction, but other colours such as green, brown, orange and blue also preside over large areas of the RGB space and must therefore form a significant part of the stimuli learnt by the system.

3.4.2 Futurama

The same concepts as utilized for Southpark may be applied to any stimuli. To demonstrate this we pursue the same treatment for the cartoon Futurama. It is important that we test the model for a range of stimuli, this allows us to observe any limitations the system may have. All previous experiments have demonstrated that the model is able to autonomously learn from its environment. It is useful to consider the shaping of the model for two cartoon that we may perceive to contain very different sets of colours.

Again we begin by demonstrating how periodically chosen stimuli are recognized in figure 3.10. Apparent from this demonstration is the dominance of the colours red and blue (cyan). This is in stark contrast to the large regions of grey and green observed in the previous example. One alteration that we have made for this demonstration is to increase the recognition rate parameter γ so that $\gamma = 10$. This ensures that recognition trajectories are more susceptible to the gradient of the potential and locate the fixed points of the system faster. From a human perspective the mappings between stimulus and corresponding class (shown in 3.10b) appear to be better correlated, fewer of these mappings appear disputable than was evident in the previous example. This may be the result of selecting a larger recognition rate parameter or may be an intrinsic quality of this stimuli choice.

In figure 3.11 we demonstrate the evolution of the systems potential in terms of projections and the locations of local minima in the RGB space. As expected, some of the basins and local minima grow, shrink, appear, and disappear across the time interval considered. Despite this, the enduring local minima remain relatively stable across the timespan. We may again conclude that the eminent features are defined during the interval $t = 0 \rightarrow 400$ and so we include more projections across this time period in appendix figure A.2. The locations of the local minima appear to be more widely dispersed across the RGB space than for the Southpark stimuli. Based on the relative depth of each minimum there also appears to be a number of colours that are appreciated in relatively equal amounts. This notion is supported by figure 3.12 which illustrates the approximate sizes

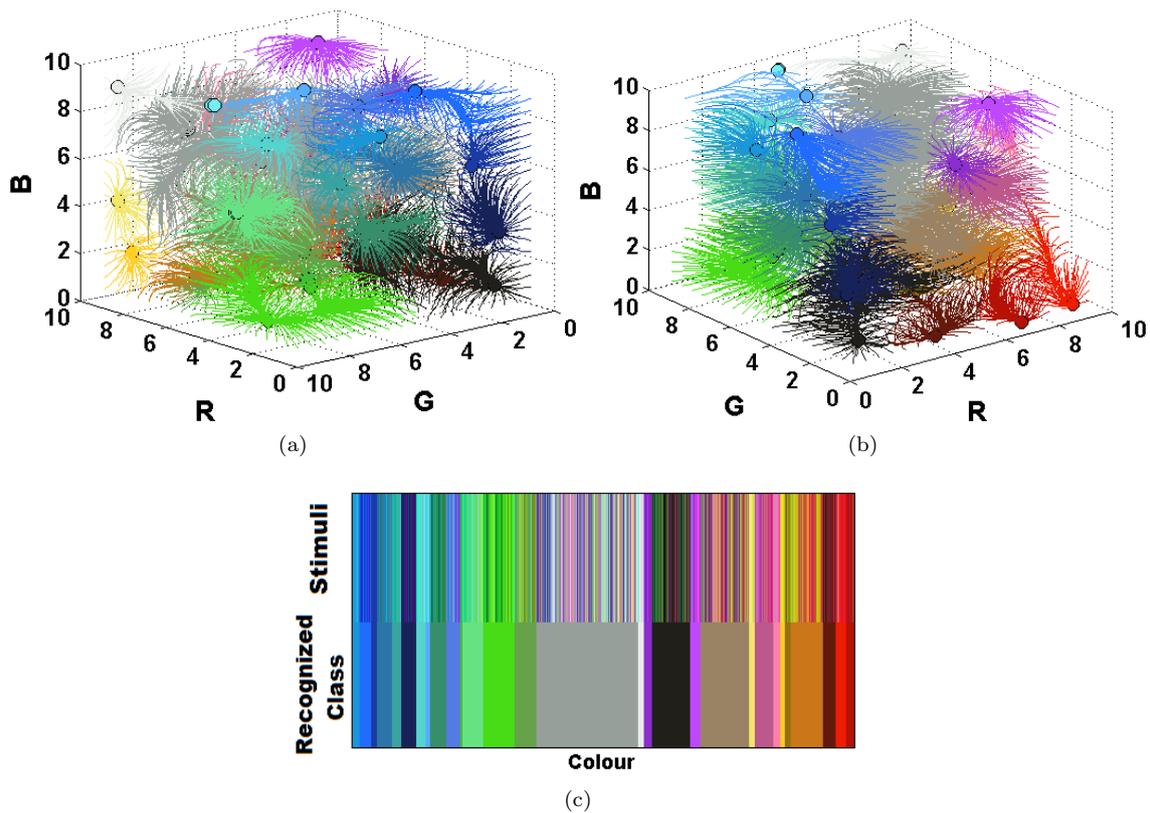


Figure 3.9: Recognition of an ensemble of 6859 stimuli with initial conditions distributed evenly across the RGB space. These recognition trajectories evolve in accordance with the stationary potential shown in figures 3.8q, 3.8r, and 3.8s. The vector field of the system utilizes gradient type dynamics. It may be considered that these trajectories provide an estimate of the size of each fixed point's basin of attraction. In (a)(b) the same trajectories are illustrated in the RGB space from two different perspectives. The colour of each line corresponds to the location of the fixed point that the trajectory identifies. This is equivalent to saying that the colour of the line indicates the class recognized. (c) demonstrates the same set of stimuli as shown in (a)(b). On the top row are the ensemble of stimuli, the bottom row shows the class that has been recognized. Here we have grouped stimuli together depending on how they are identified.

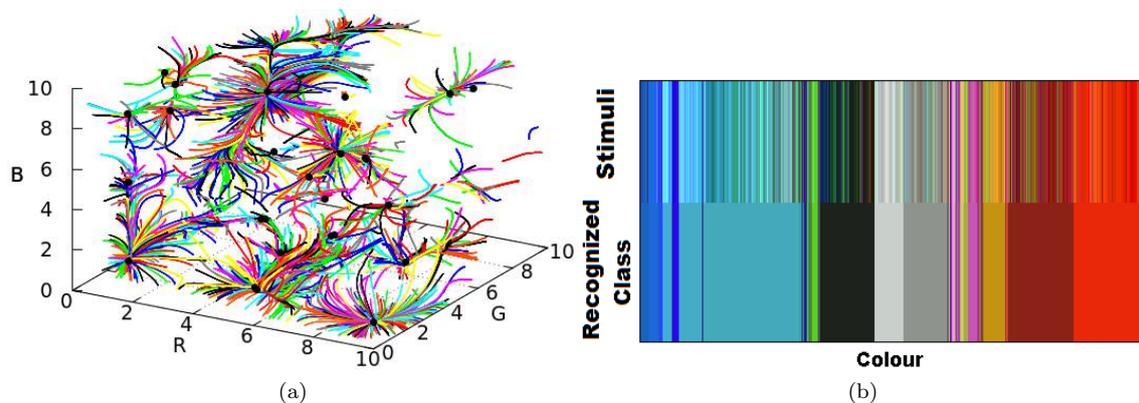


Figure 3.10: Recognition of 2000 periodically selected stimuli. (a): Evolution of initial conditions towards the local minima of the energy potential (black circles). (b): Comparison of the initial condition (stimuli) (top row) and the recognized colour (bottom row). The recognition rate parameter $\gamma = 10$.

and shapes of the basins of attraction at $t = 2000$. In figure 3.12 we again consider how a mesh of stimuli points are recognized by this system. The agreement with a human observer, in general, appears to be better than for Southpark, this indicates that we cannot expect the same levels of performance for every stimuli. Again we reiterate that the model is only able to learn from the information that it receives, if this is limited we cannot expect high levels of precision, especially when assessing never before seen colours, from distant regions of the RGB space.

3.5 A One-dimensional Approach to Colour Coding

An evolving three-dimensional profile is computationally expensive to model and cannot be easily visualized. An appropriate mapping of the three-dimensional profile onto a single dimension should alleviate these issues while preserving the key functionality of the system. This requires a mapping of R,G,B triplets onto a one-dimensional colour spectrum. Such a spectrum should present colour as a continuous, gradually changing variable, with classes and sub-classes. The problem of mapping the entire RGB space onto a single axis is complex and introduces ambiguity. Nevertheless, a number of approaches have been developed that relate the colours of the visible electromagnetic spectrum to RGB colours [107, 159, 2, 24]. Within an RGB space 16,581,375 unique colours are uniquely labelled, this far exceeds the one million colours that the average trichromatic human is able to perceive. Here we shall consider a method of mapping RGB values onto the wavelength's of a simplified light spectrum. Many RGB colours such a black and white are not realized in such a spectrum, this issue is overcome by identifying nearest neighbour $R^*G^*B^*$ triplets. Any RGB point can then be mapped onto a neighbouring point that is part of the spectrum. Such a mapping allows us to represent RGB colours by wavelength, hence reducing dimensionality.

3.5.1 Mapping RGB to Wavelength

The RGB scheme is derived from biological data which infers a relationship between colours and wavelengths [165, 166]. This data, obtained from biological investigations, forms the basis of the 1931 CIE Colour space standard [144]. Over time many refinements have been applied to this colour model to reach the current RGB definition [23]. Various approaches to approximately convert between wavelength and RGB have also been suggested [107, 159, 2, 24]. After comparison of these methods we choose to utilize a model similar to [2], which creates the spectrum shown in figure 3.13 from RGB triplets.

This model is not a 1:1 mapping between wavelength and RGB, as such a substantial part of the RGB space is not represented by the spectrum. To overcome this issue we must restrict our RGB space (R, G, B) to permitted (R^*, G^*, B^*) values. Any triplet can be transformed to a corresponding permitted triplet by finding the (R^*, G^*, B^*) that minimizes Δ in equation (3.2). This is equivalent to calculating the nearest neighbour. We may then utilize the associated wavelength value of this neighbour point (R^*, G^*, B^*) as an input to the learning system.

$$\begin{aligned}
 |R - R^*| &= r \\
 |G - G^*| &= g \\
 |B - B^*| &= b \\
 \Delta &= \sqrt{r^2 + g^2 + b^2}
 \end{aligned}
 \tag{3.2}$$

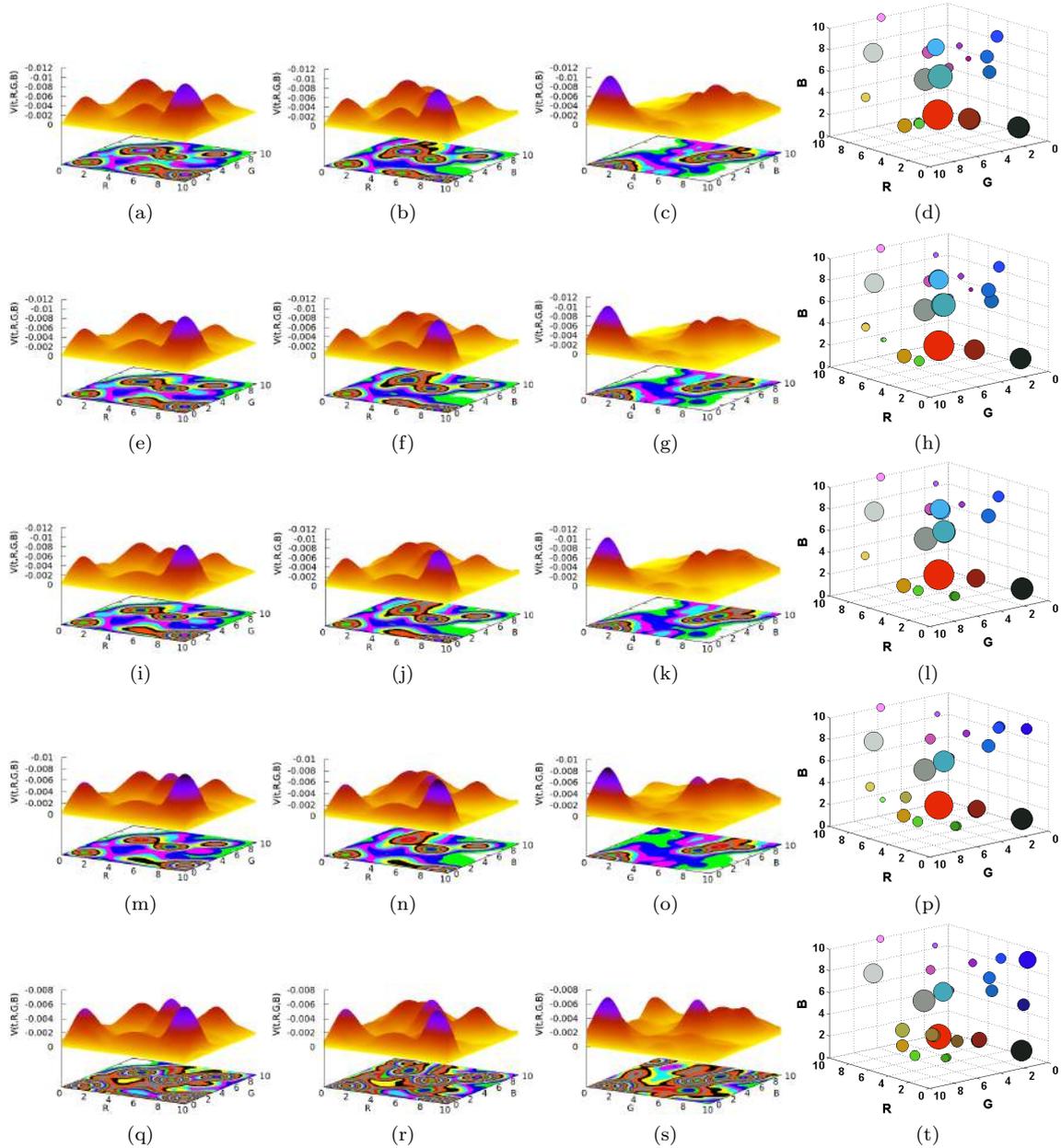


Figure 3.11: Projections of the energy potential $V(R, G, B, t)$ at time instances $t = 400, 800, 1200, 1600, 2000$. Each row contains the three projections for the corresponding time instance and the corresponding locations of the local minima in the RGB space. The colour and size of these minima points indicate the position of the minima and magnitude of the function $V(t, R, G, B)$ respectively. Each projection of the potential is illustrated with a gradient mapping of the profile at its base. This helps to identify the location of each local minimum and indicates the surrounding topology. In this example $\sigma = \sqrt{0.5}$.

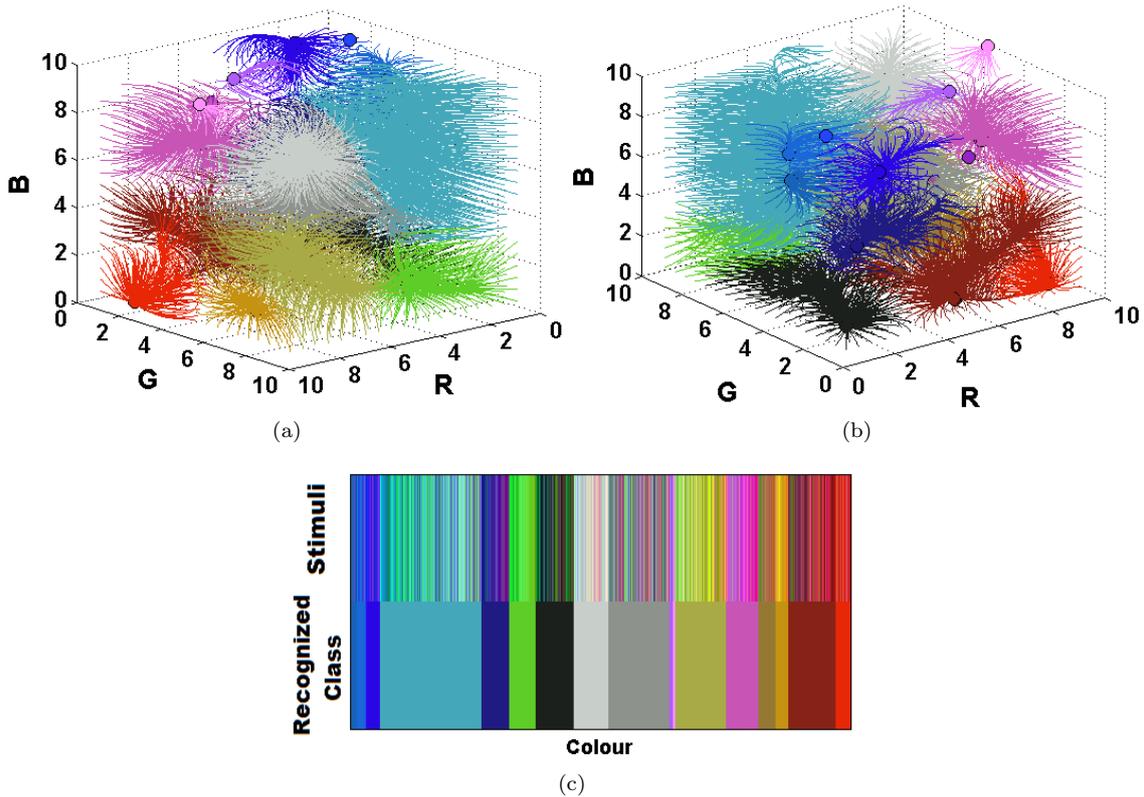


Figure 3.12: The recognition of 6859 different stimuli with initial conditions evenly interspersed across the RGB space. The potential remain stationary and takes the shape shown in figures 3.11q, 3.11r and 3.11s. Trajectories evolve in accordance with gradient descent. In (a) and (b) the colours of the lines represent the local minimum to which the initial conditions converge. The colours relate to the location of the corresponding fixed points in the RGB space. Figures (a) and (b) contain the same information but from different complementary angles. These illustrations may be used to estimate the approximate size of each fixed points basin of attraction. (c) contains a subset of the same information but as a spectrum. The top row identifies the initial condition whilst the bottom row indicates the local minimum to which the stimulus is attributed.

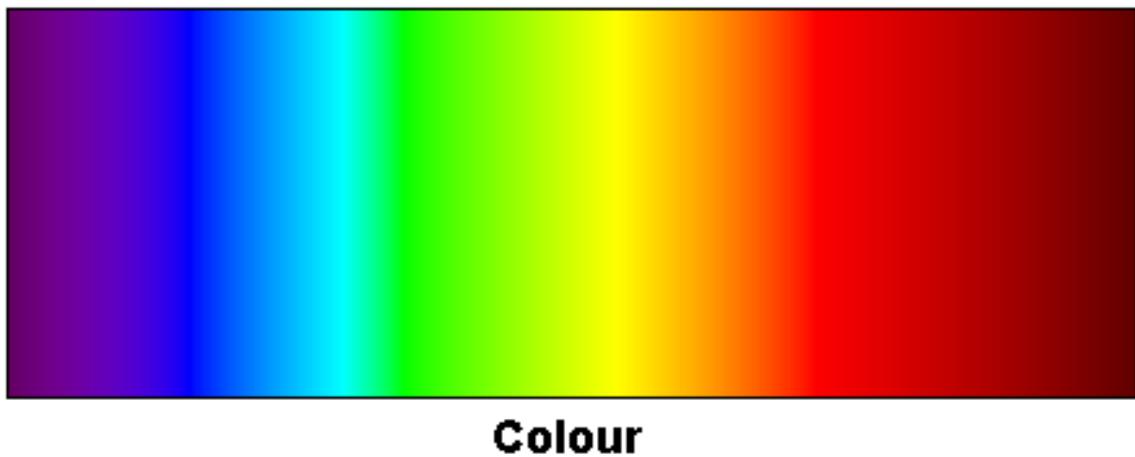


Figure 3.13: Visible Colour Spectrum built of distinct RGB colour triplets utilizing the method described in [2].

An application of this approach to a grid of initial conditions corresponding to RGB values is shown in figure 3.14a. Note that some transforms appear to be large in the context of the RGB space. We should consider the implications of such transforms and also consider the possibility of a $1 : n$ mapping in the case where a RGB value is equidistant from n different (R^*, G^*, B^*) triplets.

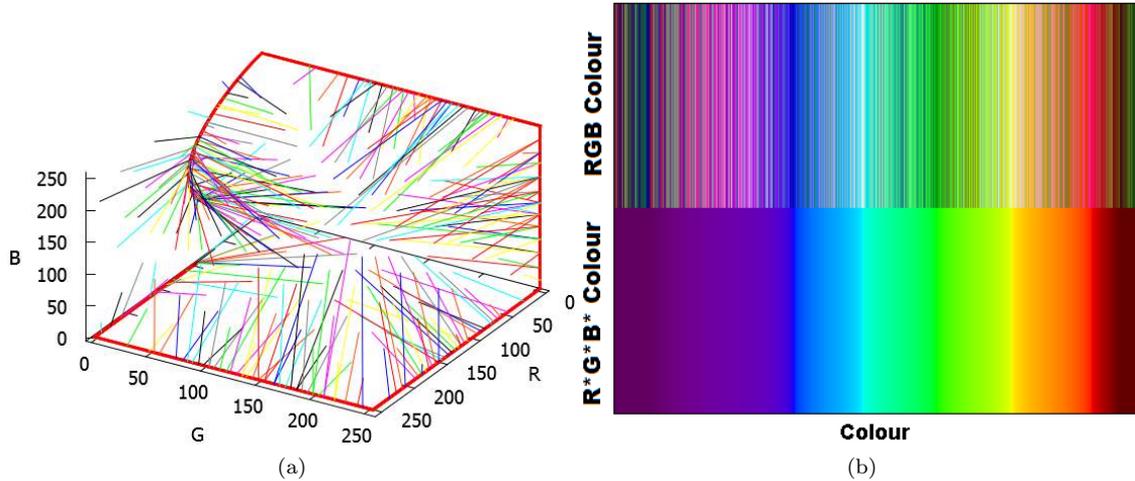


Figure 3.14: **(a)**: Mapping a mesh of initial conditions from across the RGB space onto the line $R^*G^*B^*$ shown in red. Stimuli may take any RGB value, however, our learning system is only capable of utilizing colours on the line $R^*G^*B^*$, the accuracy and applicability of this mapping is therefore very important for the majority of stimuli. **(b)**: Colours of the RGB space (top) and their corresponding $R^*G^*B^*$ transforms (bottom). These mappings are the same as those shown in (a).

Implications of this Approach

Figure 3.14b suggests that the proposed mappings are generally plausible, certainly there appears to be correlation between colours and their transforms. Of particular interest is how the model deals with black $(0,0,0)$ and white $(255,255,255)$. These colours, like many others, are not intrinsic within the visible colour spectrum. Black $(0,0,0)$ becomes $(98,0,0)$ which appears acceptable (figure 3.15a), however white presents more of a problem. $(255,255,255)$ is equidistant from two $R^*G^*B^*$ points; $(255,255,0)$ and $(0,255,255)$. These points are well separated within the spectrum and reflect radically different colours. This is also shown in figure 3.15a. Computationally we can overcome this problem by specifying that, in such a scenario, the colour is assigned to the highest/lowest wavelength. A problem that is harder to overcome is the subtle dependence on initial conditions when points are close to this equidistant relation. Figure 3.15b shows the assignment of $R^*G^*B^*$ values for the range $R=250:255, G=250:255, B=250:255$. Here an assignment has been made to highest wavelength in the case of equidistance. The subtle dependence on initial conditions is abundantly evident, the implications are visualized in figure 3.15c. Note that in figure 3.15c the top row represents the array of RGB values shown in figure 3.15b, these colours tend not to be discernible with the human eye. The corresponding mappings of these colours are therefore unrealistic, a more appropriate mapping should be applied.

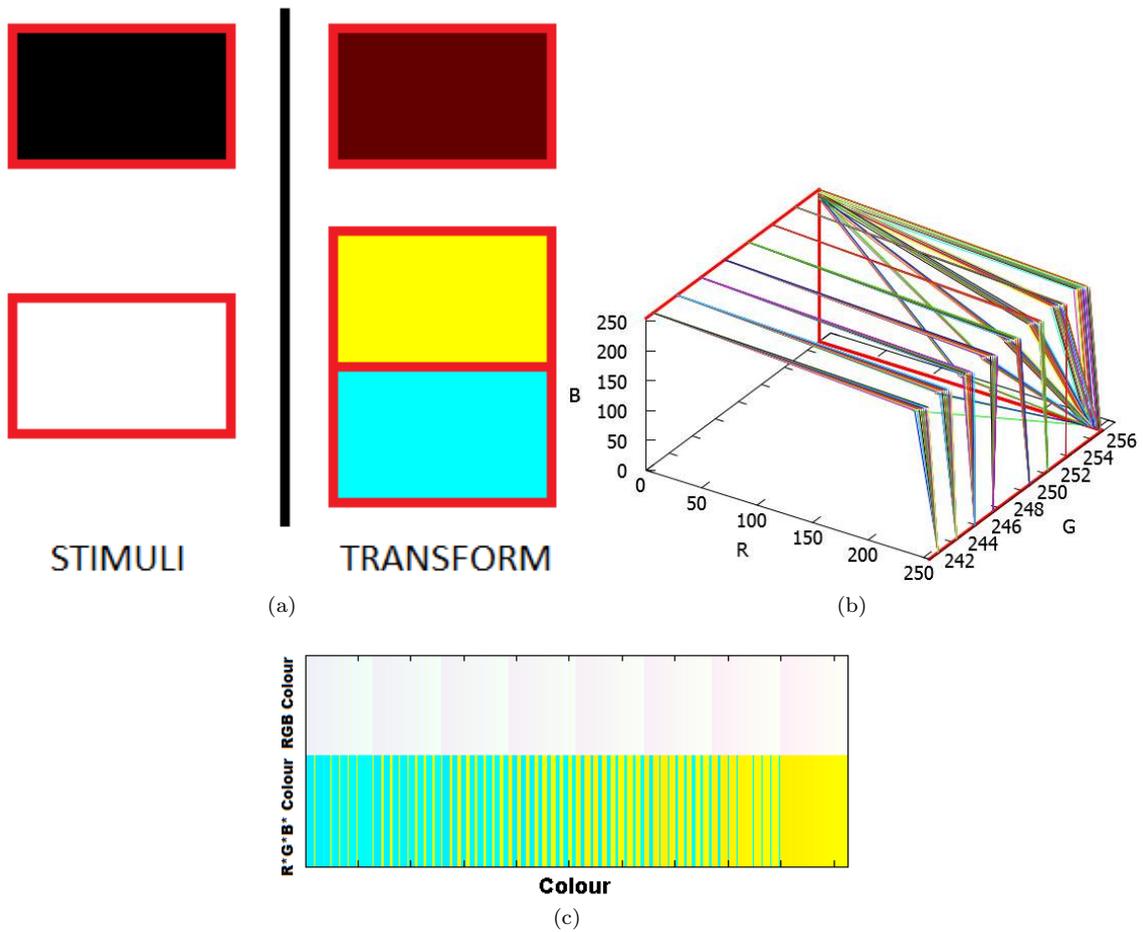


Figure 3.15: (a): The transformation of the RGB colours black and white to a corresponding $R^*G^*B^*$ colour. White lies at an equidistant point between yellow and cyan, hence there are two possible transforms. (b): RGB values within the range $R=250:255$, $G=250:255$, $B=250:255$ are all very similar colours, however, they can be transformed into very different $R^*G^*B^*$ colours (red line) by utilizing equation (3.2). This is reinforced by the visualization shown in (c). The upper box contains all colours within this range, the lower shows each related transform.

Inclusion of a Grey scale

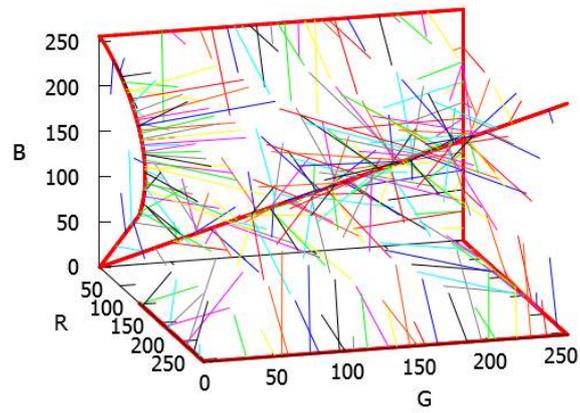
The fundamental issue with this mapping approach lies with the treatment of colours on the line $R = G = B$. A standard wavelength-colour mapping neglects colours on the black-white line, instead making large transforms to colours that, in most cases, are poor representations of the original. In such a circumstance it has also been shown that the sensitivity to initial condition is severe. Biologically speaking the eye contains two types of receptor; the colour sensitive cone cells, of which there are approximately six million, and light sensitive rod cells; of which there are around 120 million. When light intensity is low cone cell performance is poor, the light sensitive rod cells allow us to continue to perceive our environment, these cells do not respond to colour, we view the world in a grey-scale.

Extending the spectrum of colour to include a grey scale should allow us to perform a more relevant mapping and circumvent the other encountered issues. Such an extension should be continuous with the previously discussed spectrum. Under such circumstances the term wavelength loses significance, colour change remains gradual, with each distinct colour referred to by a unique pseudo-wavelength or numbered colour label. We introduce this new colour palette by introducing the line $R = G = B$ which gradually fades from white to black. This line is continuous with the existing permitted $R^*G^*B^*$ as can be seen from a comparison of figure 3.16a with figure 3.14a. A definition of the spectral colour line (red line in figure 3.16a) is given in table 3.2. This definition of the spectrum is based on the values prescribed in [2]. The attenuation term ν is utilized to account for the reduction of colour intensity at the extremes of the spectrum. As can be seen in figure 3.16a this function traces a line around the outside of the RGB space, it is therefore appropriate to represent a large range of colours. The different regions that we have previously referred to as wavelength should not really be considered as wavelength in the true sense, but rather as a set of numbered labels referring to different colours. Each section of table 3.2 can be seen to describe one of the distinct changes in the direction of the line R^*, G^*, B^* . Although we have chosen to follow this prescription of the line, it is clear that a number of different lines may also be drawn that visit large regions of the RGB space. We may postulate that several alternative lines may also be applicable as suitable colour palettes.

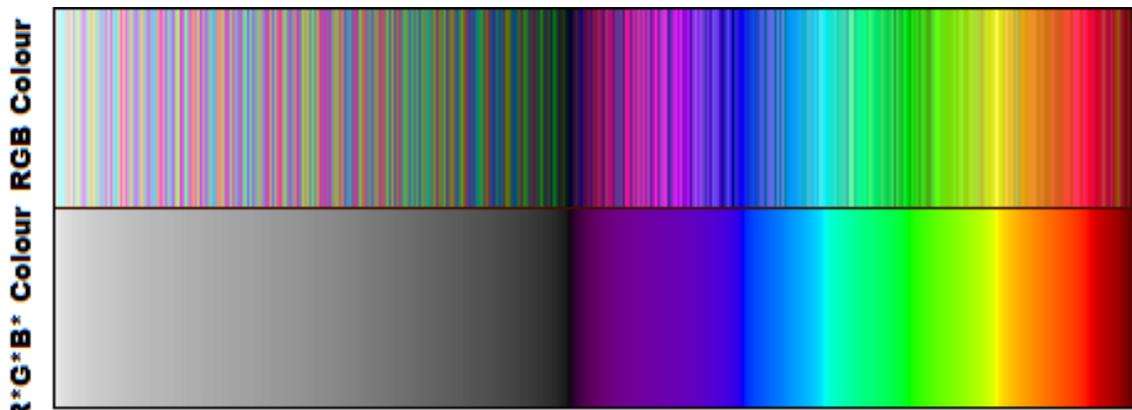
Utilizing the spectrum described by table 3.2 we can again compare a range of RGB colours and the corresponding mappings. A large portion of the RGB space is now converted to grey scale as shown in figure 3.16b. Figure 3.16a highlights the minimal distances Δ , a comparison with figure 3.14a shows that under this proposed conversion, distances are generally shorter, as such mappings are generally more appropriate. A further step to ensure that the transformed colour reflects the RGB triplet would be to choose a maximally permitted value of Δ . When $\Delta > \Gamma$, (where Γ is the threshold) the stimulus is rejected. This restriction may be compared to the finite range of wavelengths that the human eye is able to perceive. A comparison between figure 3.16b and 3.16c should convince readers that such a procedure is valid and sometimes necessary. In figure 3.16c the threshold parameter $\Gamma = 40$, the lower the parameter, the closer the mapping will be. We must find a balance between precision and generality, as selecting smaller values of Γ will also mean that a larger number of stimuli are omitted.

Wavelength	RGB
220-359	$R^* = 255 - 1.82 (\text{wavelength} - 220)$ $G^* = 255 - 1.82 (\text{wavelength} - 220)$ $B^* = 255 - 1.82 (\text{wavelength} - 220)$
360-380	$R^* = 97.97 + 4.87 (\text{wavelength} - 380)$ $G^* = 0$ $B^* = 97.97 + 4.87 (\text{wavelength} - 380)$
380-439	$A = 0.3 + 0.7 \left(\frac{\text{wavelength} - 380}{440 - 380} \right)$ $R^* = 255 \left(-\frac{\text{wavelength} - 440}{440 - 380} A \right)^\nu$ $G^* = 0$ $B^* = 255 A^\nu$
440-489	$R^* = 0$ $G^* = 255 \left(\frac{\text{wavelength} - 440}{490 - 440} \right)^\nu$ $B^* = 255$
490-510	$R^* = 0$ $G^* = 255$ $B^* = 255 \left(-\frac{\text{wavelength} - 510}{510 - 490} \right)^\nu$
510-580	$R^* = 255 \left(\frac{\text{wavelength} - 510}{580 - 510} \right)^\nu$ $G^* = 255$ $B^* = 0$
580-645	$R^* = 255$ $G^* = 255 \left(-\frac{\text{wavelength} - 645}{645 - 580} \right)^\nu$ $B^* = 0$
645-750	$A = 0.3 + 0.7 \frac{750 - \text{wavelength}}{750 - 645}$ $R^* = 255 A^\nu$ $G^* = 0$ $B^* = 0$

Table 3.2: The relation between our one-dimensional wavelength value and the line $R^*G^*B^*$. Here we have extended the initial method of [2] to also include colours on a line spanning black-white in the RGB space. The term wavelength may be misleading, by this we mean a single value that can be used to define a colour instead of the RGB triplet. The attenuation term $\nu = 0.8$ is responsible for reducing the intensity of colours at the limits of the visual range. Each component of $R^*G^*B^*$ takes a value between 0 and 255.

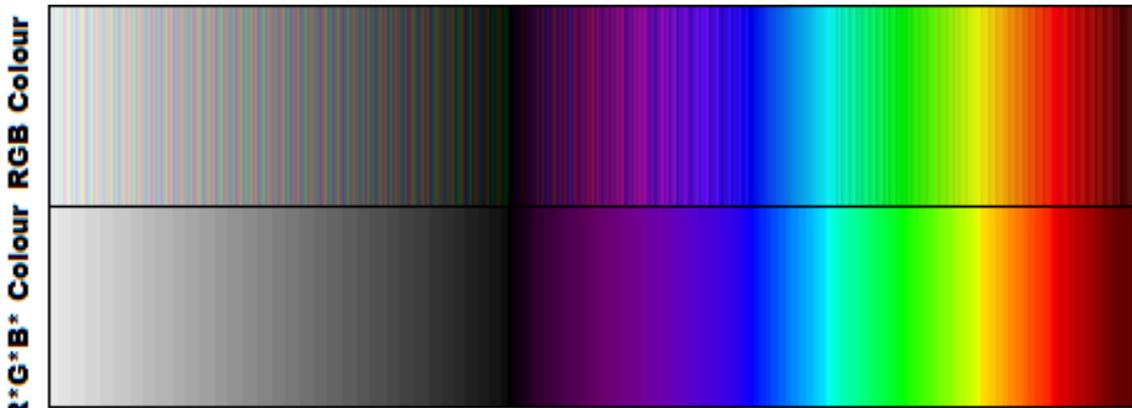


(a)



Colour

(b)



Colour

(c)

Figure 3.16: (a): The mapping of various RGB points onto the line (red) defining $R^*G^*B^*$. (b) A colour comparison between an RGB colour (top) and the $R^*G^*B^*$ colour it is mapped onto (bottom). (c) Presented with the same set of RGB colours as in (b) not every mapping onto a colour $R^*G^*B^*$ is permitted. A threshold $\Gamma = 40$ is used to compare the distance of the mapping. If Δ (as defined by equation (3.2)) is greater than Γ then the mapping will not be permitted and the system will ignore this stimulus. As shown, the restriction of RGB inputs results in a more apparent relationship between RGB and $R^*G^*B^*$ colours.

3.5.2 The One-Dimensional Learning Model

In section 3.5.1 a method to map RGB values onto a one-dimensional continuous spectrum was discussed. Such an approach limits the range of colours utilized by the system but provides the opportunity to fully visualize the time-dependent dynamics of class formation and recognition. This is possible as the dimensionality of the learning system's flexible vector field is reduced. The one-dimensional model is also computationally less expensive and so runtime is much faster. The colour spectrum that we utilize consists of 1060 distinct colours, however, we apply a linear mapping so that all stimuli are limited to the range $0 \rightarrow 10$. As with the previously presented three-dimensional models, we may take stimuli from genuine environments in the form of RGB triplets, these can then be transformed into $R^*G^*B^*$ values that can be associated with a value between $0 \rightarrow 10$ and applied to the one-dimensional learning model.

In figures 3.18, 3.19 and 3.20 the result of applying three different stimuli to the one-dimensional learning model are shown. In each of these figures we demonstrate the time dependent shaping of the potential for three different values of σ and the shaping of the potential with/without the inclusion of a threshold that restricts the range of permissible stimuli. The first row of each figure corresponds to the case where $\sigma = \sqrt{0.05}$, the second row is for $\sigma = \sqrt{0.01}$ and the third row corresponds to $\sigma = \sqrt{0.001}$. The first three columns of each figure show the evolution of the system when no restriction is applied to the stimuli, alternatively the next three columns only permit certain stimuli by setting the threshold $\Gamma = 40$. In each figure we demonstrate the shaping of the potential and recognition of stimuli for the same inputs as encountered by the three-dimensional models discussed in section 3.2. The recognition of a stimulus is described by gradient descent which locates the attracting fixed points of the system. These fixed points may be considered to represent retained knowledge, obtained via learning. The restricted approach, that only permits certain stimuli, utilizes the same initial set of input data, however, the threshold may be seen to act as a filter, reducing the total number of stimuli that are used to shape the potential. This restriction is intended to ensure mappings remain realistic. The first and fourth columns of each of the figures 3.18, 3.19 and 3.20 illustrate the shaping of the potential as a function of time and position for the unrestricted/restricted stimuli sets. On the base of each of these plots we also demonstrate the recognition of a cross-section of 2200 stimuli. This is indicated by a set of coloured lines originating at $t = 500$, that locate the local minima of the potential. The rate of gradient descent is effected by the parameter γ , here we have set $\gamma = 10$.

The shaping of the system's potential is determined by the density and distribution of stimuli received by the system. We have previously shown that $V(t, x) \rightarrow -P(\eta)$ as $t \rightarrow \infty$ is guaranteed in the limit that $g(z) \rightarrow \delta$, where δ is a delta function and $g(z)$ is a Gaussian. $g(z) \rightarrow \delta$ in the limit $\sigma \rightarrow 0$. In general this is not desirable as we require a smooth potential. Reducing σ ensures that the shape of the potential converges to the negative probability density distribution of the stimuli. Each distribution of inputs is shown alongside $V(t, x)$ at $t = 2000$ in the second and fifth columns of figures 3.18, 3.19 and 3.20. It is apparent that these lines bear the closest resemblance when σ is smaller. To demonstrate that σ is indeed the restricting factor, rather than time (number of stimuli applied), we may compare the state of the potential under different conditions. In figure 3.17 we compare the shape of the potential to the distribution of stimuli for various values of σ and for different lengths of time. Where time is increased we ensure that the distribution of stimuli remains unchanged. Notice that increasing time in this example does not improve the fit between our curves. This is only achieved by reducing σ . We should mention that over fitting

may become an issue when the Gaussian width, dependent on σ , becomes excessively narrow. We expect that any real stimuli received by the system will have a stochastic component, we require our system to be robust to such noise. Two stimuli that may be considered noisy versions of the same input may otherwise form separate classes. This problem is comparable to the over fitting of data in a deep NN during the training phase [120]. Here the solution to our problem is somewhat simpler, as increasing the parameter σ will ensure robustness. Selection of this parameter should be made based on the trade-off between generalization and colour sensitivity. Beyond representing the recognition of a colour as a simple trajectory, we are able to compare an observed RGB colour, the transformed colour (acts as input to the learning model) and a colour corresponding to the class recognized by the system. This comparison is shown in the 3rd and 6th columns of figures 3.18, 3.19 and 3.20. The top row of each plot shows a cross-section of RGB colours, in the centre are their transformed R^*, G^*, B^* colours and below this are the corresponding classes of the system. Recall that these colours coincide with the local minima of the potential on the x range. Demonstrating recognition via this medium is intended to provide an intuitive representation of the successful segmentation of the colour spectrum.

Figure 3.18 is a somewhat contrived example which allows us to make some critical inferences about this one-dimensional approach. Here the stimuli corresponds to a video created of known colours that are well separated in the RGB space. The video is looped repeatedly throughout the runtime of the model to ensure that the system experiences each colour. As the video only contains eight colours, it is interesting that as many as 25 colours are recognized as part of the learnt spectrum. The exact number of classes that develop in the systems memory is shown to depend on the value of σ and whether stimuli are excluded by the parameter Γ . When the same stimuli was discussed in section 3.3, in relation to a three-dimensional RGB space we again observed the development of more system attractors than would be expected for this input. We attributed this difference to changes in the lighting as the video progressed and to some fallibility of the input device. The disagreement between what was observed and expected was reconciled by comparing the state of the potential to the probability density distribution of inputs. As shown for both the unrestricted and restricted cases, provided in the second and fifth columns of figure 3.18, the same reconciliation may be reached here. We may note however, that although the locations of the local minimas coincide on the line x , the profiles of the potential and probability density distribution do not generally match. This disparity is alleviated by reducing σ as can be seen from figures 3.18n and 3.18q. We must consider whether the value $\sigma = \sqrt{0.001}$ is too small to ensure robustness to noise. There is also the issue of creating very narrow classes and so impairing the ability of the system to recognize unfamiliar colours. This would appear to be an issue, the potential shown in figure 3.18m appears to contain fourteen local minima whilst the recognition of colours shown in the same figure, and also indicated in figure 3.18o, demonstrates many more classes. We may surmise that these trajectories have not yet reached a fixed point of the system due to the potential remaining approximately flat across large regions.

By excluding some stimuli from our model due to their distance from the line R^*, G^*, B^* it is evident that the potential of the system typically incorporates fewer local minima. We may consider that these fewer minima are a more accurate representation of the colours of which the stimuli is actually composed. We should also note that this reduction in attractors is not always the case, as illustrated by the central row of figure 3.18. Here we can clearly observe two different classes of grey for the restricted case in comparison to the single class for the unrestricted stimuli.

This separation of classes is explainable if we again consider the stacking properties of Gaussians. When an unrestricted stimuli is applied, values across a wider range are encountered, this results in the stimuli forming one minimum with a larger basin. In the case where stimuli are restricted, in-between stimuli are omitted, this results in two distinct minima forming. If we consider that this model should contain close to eight classes (listed in table 3.1), then the best fit for this assertion is observed in figures 3.18j,k,l, when $\Gamma = 40$ and $\sigma = 0.01$. We should be wary of this however, as the colours seen by the camera do not exactly match the colours that it is shown. It is interesting to note the effect that selecting parameters has on the development of the systems attractors. We have only considered a small subset here, however it is very apparent that adjustments can have a dramatic impact on the development of the system.

The further two examples shown in figures 3.19 and 3.20 corresponds to stimuli collected from the cartoons Futurama and Southpark respectively. As we have no control over their colour compositions we cannot speculate about the appropriateness of the number of classes formed. We do however observe that the distribution of inputs emulates the final potential in both cases. Again we may consider the classes of the one-dimensional model against the classes of the three-dimensional models shown in figures 3.12 and 3.9. What becomes apparent is that the transformations applied have a considerable effect on the classes of the system. This is hardly surprising if we consider that the 16,581,375 positions of the RGB space are mapped onto a meagre 1060 points of a spectrum. It is difficult to compare the different colours of the three-dimensional model to those of the one-dimensional due to the transform that has been applied. Certainly some similarities can be drawn, for instance the one-dimensional model shown in figure 3.19 appears to contain many of the classes of colour shown in figure 3.12, however, there are some glaring differences. One obvious example is the pink class, this is appreciated by the three-dimensional scheme but is not replicated in the one-dimensional case. It may be the case that this class has been replaced with a purple colour as a result of the transform. If we consider the line R^*, G^*, B^* , then the omission of pink is easy to account for as the spectrum transform does not permit this colour. Issues such as this suggest that although the one-dimensional approach is less computationally expensive, the restrictions of the scheme mean that it is ill equipped to recognize colour in an unrestricted sense. We may devise certain environments where the colours observed are restricted to this smaller palette, however, then we are no longer considering the system operating in a natural setting. In contrast, the shortcomings of the one-dimensional model can be seen to validate the three-dimensional approach. It may be postulated that the three-dimensional RGB scheme is a minimal requirement for the learning and recognition of colours in a natural environment. Another concern that may be raised about the one-dimensional approach is highlighted if we compare figures 3.19l and 3.20l. The differences between the classes shown in the bottom rows of these colour spectrum's is somewhat negligible. The relative sizes of each class are different, however both examples contain many of the same constituent colours. Notable differences include a small regions of yellow and dark-grey that appear for Southpark but not for Futurama. Many similar classes are realized despite the radically different appearances of the two potentials (shown in 3.19j and 3.20j). If we compare the classes shown in figures 3.19l and 3.20l to the classes of the three-dimensional systems shown in 3.12c and 3.9c, then we see that there is a much greater difference between the colours contained within the two three-dimensional representations than there is for the one-dimensional models shown here. This may suggest that a one-dimensional spectrum is simply too basic and does not adequately portray the adjustment of colour in a continuous, gradual way.

Putting these concerns to one side we should also mention the successes of the system. These

become apparent if we neglect the transformation from RGB to $R^*G^*B^*$ and instead only consider stimuli in the form of $R^*G^*B^*$ triplets. Here we may draw an analogy to a person with a severely restricted ability to perceive colour. We may consider that the range of colours that they are able to interpret corresponds to the values defined in table 3.2. Such a limitation is comparable with human visual limitations, for instance suffers of monochromacy may only discern approximately 100 colours. These, of course, do not correspond to our line $R^*G^*B^*$. Restricting our attention to the colours $R^*G^*B^*$ we can compare the relationship between a stimulus and the class that it is attributed to by considering the middle and bottom rows of each diagram in the third and sixth columns of figures 3.19 and 3.20. It is clear from these representations that a strong relationship exists between the colour of the stimuli and the classes that form. Again we should highlight that the number of classes is influenced by the parameter σ . The system builds a representation of the world that it has observed via an on-line, unsupervised, flexible approach. The downside of the one-dimensional system lies not with the learning model, but with the transformation of three-dimensional coordinates onto the line. We may consider the three-dimensional model to possess a superior capacity to learn colours. The one-dimensional alternative may be thought of as a less sensitive system, capable of drawing fewer distinctions between colours, but still fulfilling the principal target; automatic classification and recognition of a wealth of observed colours.

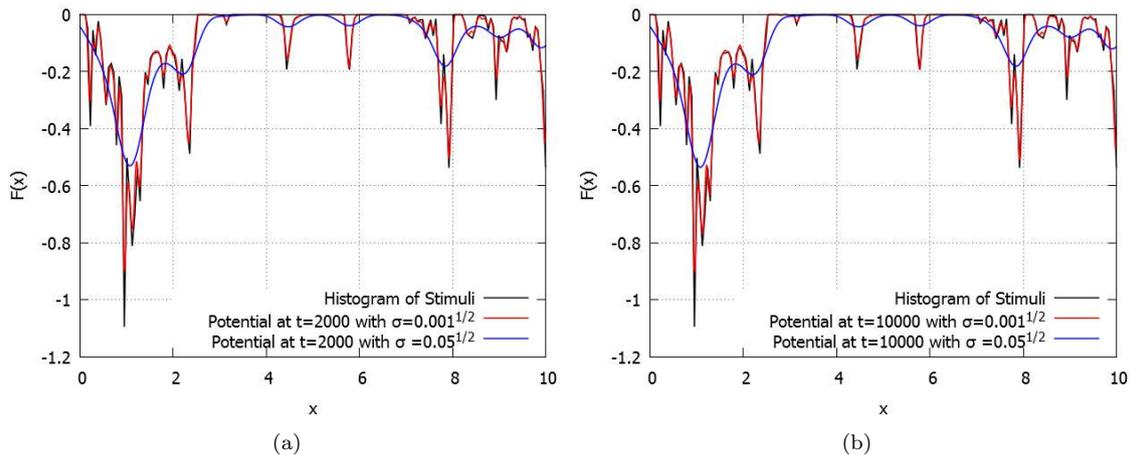


Figure 3.17: **(a)**: Comparison between the PDD of inputs and the system's potential $V(t, x)$ at $t = 2000$ for two different values of σ . Note that a narrow Gaussian width parameter ensures convergence between the PDD of inputs and the potential. In the case where σ is larger the system is not able replicate the distribution of the stimuli. **(b)** demonstrates that this is not due to insufficient evolution time as the potential shaped with $\sigma = \sqrt{0.05}$ has still failed to imitate the profile of the PDD of inputs by $t = 10000$. Note that the statistical properties of the stimuli remain the same in each illustration.

3.6 Summary and Conclusion

Here we have considered the utilization of the memory foam model for the complex tasks of colour learning and recognition. This autonomous dynamical system utilizes a plastic vector field to create a representation of the experienced stimuli. Unknown stimuli are recognized in accordance with this landscape, while no distinction between the learning and recognition of colours is required. Here we consider a stimuli as as set of RGB triplets that are prescribed to the model via a web-

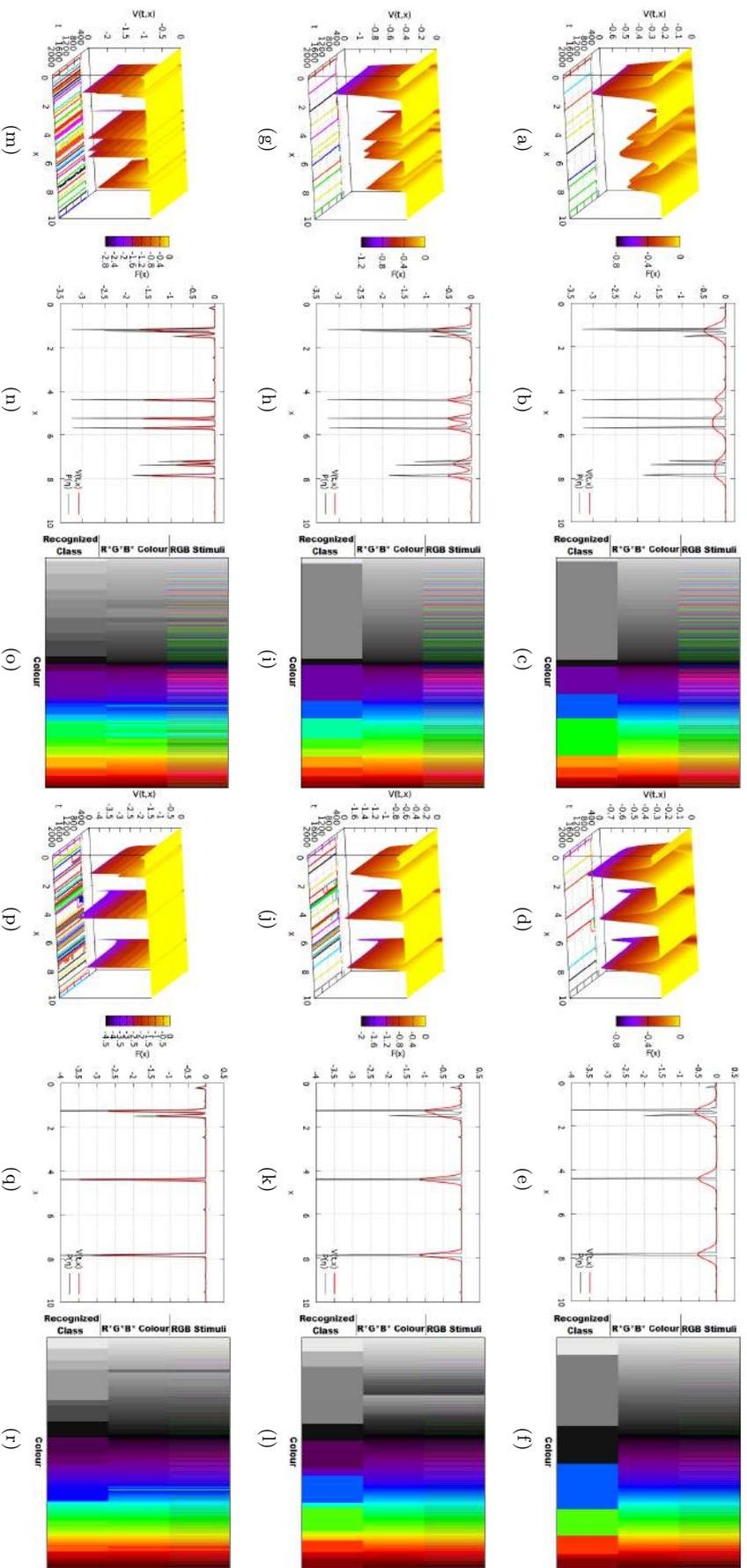


Figure 3.18: In each figure the stimuli is a looping video of colours as previously utilized in section 3.3. Columns 1, 2 and 3 correspond to an unrestricted stimuli transformation, whereas, columns 4,5 and 6 illustrate the case where the stimuli is restricted to within 40 units of the line R^* , G^* , B^* . We also consider the effect of σ on the number of classes that emerge. The first row demonstrates the system when $\sigma = \sqrt{0.05}$, for the second row $\sigma = \sqrt{0.01}$ and the third row corresponds to $\sigma = \sqrt{0.001}$. Figures (a),(d),(g),(j),(m),(p) demonstrate the evolution of the potential $V(t, x)$. Also included in these illustrations is the recognition of 2200 initial conditions, this is shown on the base of each graphic. The initial conditions are evenly interspersed across the range of x and the descent rate parameter $\gamma = 10$. These recognition trajectories are represented in (c),(f),(i),(l),(o),(r) in terms of colour. The top row of each figure shows the initial R, G, B colour, the middle is the transformed R^*, G^*, B^* colour and below this is the recognized colour. In figures (b),(e),(h),(k),(n),(q) the potential $V(t, x)$ at $t = 2000$ is compared to the PDD of the permitted stimuli.

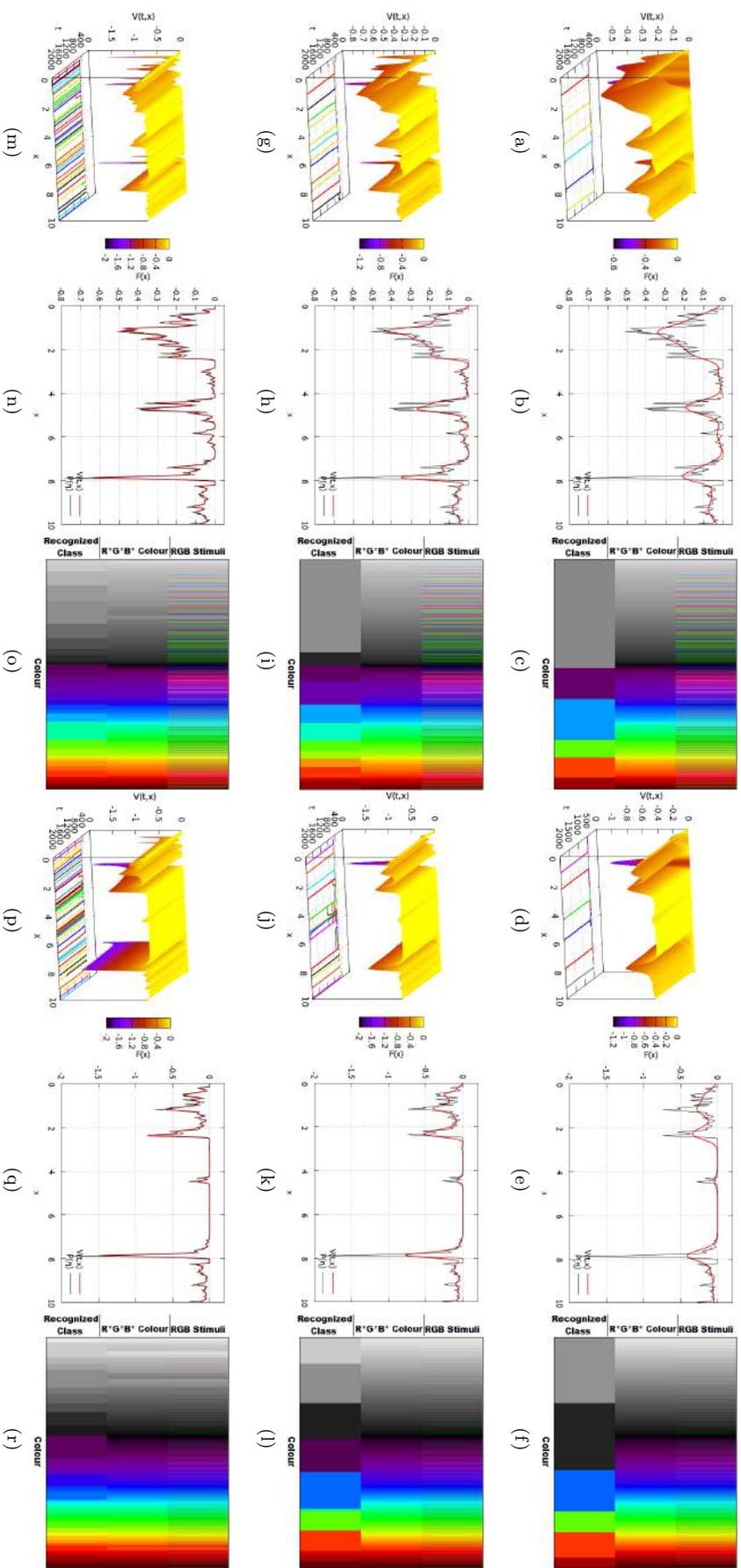


Figure 3.19: In each figure the stimuli is a looping video of the cartoon Futurama. Columns 1, 2 and 3 correspond to an unrestricted stimuli transformation, whereas, columns 4,5 and 6 illustrate the case where the stimuli is restricted to within 40 units of the line R^*, G^*, B^* . We also consider the effect of σ on the number of classes that emerge. The first row demonstrates the system when $\sigma = \sqrt{0.05}$, for the second row $\sigma = \sqrt{0.01}$ and the third row corresponds to $\sigma = \sqrt{0.001}$. Figures (a),(d),(g),(j),(m),(p) demonstrate the evolution of the potential $V(t,x)$. Also included in these illustrations is the recognition of 2200 initial conditions, this is shown on the base of each graphic. The initial conditions are evenly interspersed across the range of x and the descent rate parameter $\gamma = 10$. These recognition trajectories are represented in (c),(f),(i),(l),(o),(r) in terms of colour. The top row of each figure shows the initial R, G, B colour, the middle is the transformed R^*, G^*, B^* colour and below this is the recognized colour. In figures (b),(e),(h),(k),(n),(q) the potential $V(t,x)$ at $t = 2000$ is compared to the PDD of the permitted stimuli.

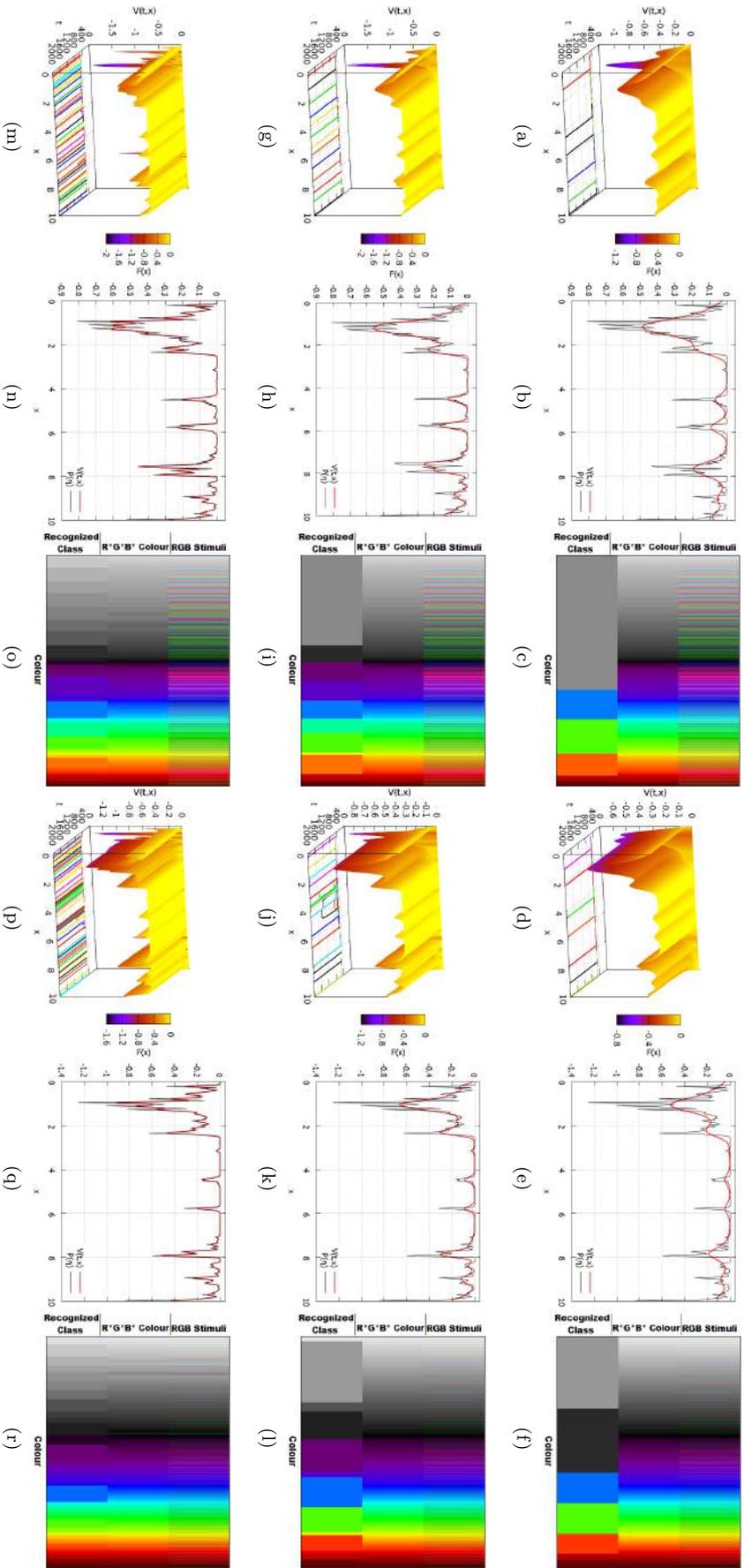


Figure 3.20: In each figure the stimuli is a looping video of the cartoon Southpark. Columns 1, 2 and 3 correspond to an unrestricted stimuli transformation, whereas, columns 4,5 and 6 illustrate the case where the stimuli is restricted to within 40 units of the line R^*, G^*, B^* . We also consider the effect of σ on the number of classes that emerge. The first row demonstrates the system when $\sigma = \sqrt{0.05}$, for the second row $\sigma = \sqrt{0.01}$ and the third row corresponds to $\sigma = \sqrt{0.001}$. Figures (a),(d),(g),(j),(m),(p) demonstrate the evolution of the potential $V(t, x)$. Also included in these illustrations is the recognition of 2200 initial conditions, this is shown on the base of each graphic. The initial conditions are evenly interspersed across the range of x and the descent rate parameter $\gamma = 10$. These recognition trajectories are represented in (c),(f),(i),(l),(o),(r) in terms of colour. The top row of each figure shows the initial R, G, B colour, the middle is the transformed R^*, G^*, B^* colour and below this is the recognized colour. In figures (b),(e),(h),(k),(n),(q) the potential $V(t, x)$ at $t = 2000$ is compared to the PDD of the permitted stimuli.

camera. The three-dimensional vector field of the system is manipulated by these inputs to form attractors of the system that coincide with the distribution of colours. Some of the research of this Chapter is summarized in a yet to be published paper, this is included in the appendix.

By experimenting with colour we have been able to demonstrate that the system is applicable to more intensive learning tasks than previously encountered. It may be argued that vision is the most significant source of input for many biological learning system. It is certainly encouraging that we have been able to replicate the learning of stimuli representing part of this faculty. The ability to visually perceive our environment is integral to the way that humans interact and learn from our surroundings. Here we have taken the first step towards demonstrating this artificially. Modelling the learning of more general visual inputs such as objects, people or landscapes would require an even more complex system, with a much higher level of dimensionality. We may consider this experiment as an important part of the chain leading from the representation of individual stimuli, to much more complex situations. Ultimately we may imagine a high-dimensional system that retains knowledge from every imaginable source. The sensory information utilized by the model may in fact exceed that of a human. Features of a stimuli such as: colours, sound, size, position and temperature may all be incorporated to shape the vector field of the system.

The limitations of the system's sensor (web-camera) are discussed and quantified in section 3.2.1. Despite these limitations the system is shown to be able to automatically form classes representing different colours when introduced to the popular cartoons Futurama and Southpark. As external human influence is avoided, we may stimulate that the model demonstrates unsupervised learning. We have previously stated that the majority of human learning occurs in this fashion, an assertion that would appear particularly relevant to learning colours. As children we may be taught the label to assign to a specific colour, however we naturally learn to distinguish between colours. The formation of new classes and associations made between different colours are instilled without the need for labels, these only really exist as a necessity of language [93, 164].

The RGB colour scheme is demonstrated to be a viable construct to define colours that are applied to our learning model. The memory foam model's groups colours into distinct classes based on the distribution of inputs and the parameters of the system. One difficulty that is encountered is the illustration of the models dynamics. We should be clear that this has no relevance to the actual functioning of the model. Because the potential is a function of R,G,B and time, it cannot be fully visualized in a single diagram. To convince readers that the system does indeed function as hypothesised, we presented aspects of the shaping via various alternative illustrations. These include presenting projections of the potential against two of the three-dimensional components at multiple time instances, visualizing the evolution of the local minima in an RGB space, and demonstrating recognition from a large mesh of initial conditions. Considering the amalgamation of these methods a clear description of the memory foam model's behaviour is accumulated.

By selecting cartoon based stimuli we have ensured that the potential comes to represent a diverse cross-section of colours. This is not a requirement, but should help readers to associate with the results. We may also consider how the system may behave in less diverse conditions. Consider for instance a desert, or the north/south pole, places where colour contrasts are limited. In such locations we may expect that the system will only experience a narrow range of colours. At the poles the system may become an expert in differentiating between shades of 'white' and 'blue', but will have little/no knowledge of 'red' or 'green'. To learn finer details the system will require σ to be small. We may anticipate that a person who lives in a similar environment, where the contrast between colours is low, may learn in a similar manner. We cannot learn without experience, this

should be kept in mind when considering the results of this model. The development of this paradigm may involve extending our trials to other real world environments, but before this is considered we may suggest that superior hardware is acquired.

Although the three-dimensional colour representation is able to portray a wealth of colours, the integration of the systems variables and the collection of stimuli are computationally expensive. This means that an extensive delay exists between the observation of a stimulus and the model utilizing this information. As a method to improve processing time we may first transform the stimuli from a three-dimensional RGB triplet onto a one-dimensional spectrum. In section 3.5.1 we have presented one method of making such a transform. We may then progress with a one-dimensional potential in a manner similar to that discussed in Chapter 2. Although this limits the number of distinct colours that may be perceived by the model, the categorization of colours still appears viable. If we were to consider the system in an artificial environment that only contained the limited colours of our one-dimensional spectrum then this one-dimensional approach would certainly provide a sufficient alternative to the three-dimensional model. Unfortunately it is shown that the mapping between some RGB values and the colours of this one-dimensional spectrum are not particularly close. An example of this problem is the colour pink which becomes associated with purple via this approximation. As a means to counter such issues we have introduced a restriction on the stimuli. If the mapping between an RGB colour and spectral colour exceeds a given threshold, then the stimulus is not utilized by the system. Here we may make an analogy to a person with a restricted set of receptors such as a person suffering from monochromatism. It should be noted that monochromatism is unlikely to effect a person in this way, but does have the effect of dramatically reducing the number of distinguishable colours. We may conclude that when RGB colours are close to the one-dimensional spectrum defined by the line $R^*G^*B^*$, the system utilizes this information to form classes that are relevant to the input. We are able to demonstrate that unknown stimuli are recognized as part of relevant classes in this situation.

In the next Chapter we begin to look at the alternatives to gradient descent for modelling recognition. We focus on the van der Pol oscillator as a paradigm of self-sustained oscillation that may be utilized to elicit higher order dynamics. Applying non-linear dynamical principles to the concept of recognition should allow the model to become more flexible, permit a more biologically relevant description of recognition, allow more information to be conferred by the vector field and possibly permit hierarchy to be conveyed.

Chapter 4

Representation of Class Recognition via Limit Cycles

In this Chapter we turn our attention away from the classification of stimuli and instead focus on recognition. A major success of this new approach is that recognition and classification can be simultaneously realized. Classification deals with utilizing stimuli to form a representation of everything the system has learnt. This is described by a flexible, evolving potential. Recognition is concerned with comparing a stimuli to the information stored by the system. Let us consider the example of facial recognition, this is a task that traditional computers typically perform very poorly at due to way in which they linearly process information. A human observer may recognize a person almost instantaneously despite minor differences in the persons appearance from the last time they were observed. In this context it is clear what is meant by recognition, it is identifying whether the person is known to the learner and identifying the label (name) associated with that stimuli. The same concept may be inferred by our system. A learnt face may be considered as a class represented by a local minimum of the potential. Recognition is considered as identifying the correct/most relevant class to associate a stimuli with. It is important to realize that the perceived stimuli is unlikely to exactly resemble a known class, the learning system must therefore make an appropriate comparison between retained information and stimuli. Returning to our facial recognition analogy, the system should be resilient to minor changes in stimuli such as lighting conditions, hair style, facial hair or ageing. The basin of attraction of any attractor representing a class should be sufficiently large that stimuli encoding minor differences are still attributed to the same class. Utilizing the system to actually perform facial recognition would be a complex task. This description is intended to merely provide a basis for conceptualizing the process of recognition. To employ this model for facial recognition the potential characterizing the memory of the system would require a considerable number of dimensions to contain sufficient information about each individual. Beyond this issue, the system may be suitably trained and should perform recognition much faster than a standard database search. At this juncture we only use this analogy to describe the concept of recognition, however, there may be some scope for such development in the model's future. The most basic premise underpinning recognition lies in the idea of comparing a stimuli to retained information, from this the system infers which known class the stimuli most closely resembles.

The original model, presented in [74], utilizes gradient descent to locate a local minimum which is associated with a class. Each local minimum of the potential corresponds to an attracting fixed points of the system and the topology of the potential in each points vicinity dictates the rate at which a stimuli ‘descends’ to the attractor. As such, the phase space of the system is split into a set of basins of attraction, each containing an attracting fixed point. These basins remain smooth due to the properties of the Gaussian function (equation 2.2). Gradient descent utilizes the vector field of the potential to locate these minima, hence, any position $\eta = x$ will converge towards an attracting fixed point, corresponding to a local minimum of the potential. In this fashion any stimulus can be recognized as part of a known class.

The primary issue with the gradient approach is that it lacks the ability to represent recognition by anything other than a fixed point attractor. Recalling our discussion of non-linear dynamics presented in section 1.4, the stable fixed point may be considered the most basic object that organises a phase space. All trajectories within the attractor’s basin converge towards it. Here we wish to explore a more general phase space with various types of attractor. This will allow us to represent additional aspects of a class such as its size and relation to other classes. A class may still be represented by a fixed point, however, the possibility of limit-cycles, quasi periodic orbits and even chaotic motions would instigate greater scope to the way the system represents recognition. Ideally different regions of the phase space should be occupied by different distinct attractors. A certain region of the potential may be affiliated with a limit cycle, whilst another, with different topological properties, may be described by a quasi-periodic orbit. The ability to represent recognition with differing phase space attractors may form the basis of a hierarchical approach.

4.1 A Hierarchical Representation of Class Recognition

The notion of hierarchy is intended to convey a structure that relates the different classes of the system. It is clear that any relation between different classes is not representable with a fixed point attractor, but may be approached by demonstrating the recognition of a stimulus with limit cycles and multi-dimensional quasi-periodic orbits. The knowledge of the system may be considered to form classes and subclasses. Let us consider the hierarchy of biological classification as an analogy for the type of structure that we may wish to identify. This classification method has the following branches: Life→Domain→Kingdom→Phylum→Class→Order→Family→Genus→Species. These form a pyramid structure where constituents with features in common all appear at the same level. The same concept may be inferred from a potential such as that shown in figure 4.1. In this figure we may consider there to be 3 classes, each divided into smaller and smaller subclasses. It would be advantageous if this relation between classes and subclasses could be inferred from the realizations of recognition. As stated, this is clearly not possible utilizing gradient dynamics, but may be achieved if the system is able to utilize higher order representations such as limit cycles. At this stage comments regarding hierarchical representation are purely speculative and are only intended to motivate our move away from gradient approaches towards more robust methods. One perceivable route to relate classes and subclasses may be to associate the lowest class level with a fixed point, the next step may be associated with a limit cycle and yet wider classes may be represented by quasi-periodic orbits. Where the system has many layers of classes and subclasses we may consider that there will be many different, but related quasi-periodic cycles. The dimensionality of the associated tori may depend on the number of levels of hierarchy that exist for

the given potential. It may also be considered that at some stage the upper bound of the class representation is represented by a chaotic trajectory. This concept is illustrated in figure 4.2.

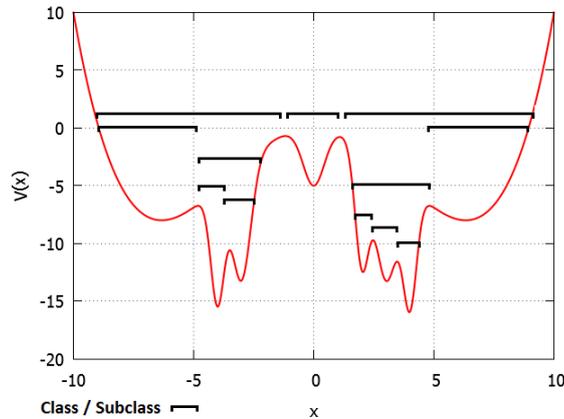


Figure 4.1: Classes and sub-classes may be illustrated for an arbitrary potential as shown by this red curve. As marked, the profile may be split into a number of smaller and smaller wells. The black brackets indicate the approximate size of each sub-class. Here we have identified several layers of classification. This is just one example, the topology of the evolving system may be very different but we anticipate that similar relationships between local minima may be observed.

The first stage of our investigation should focus on generating limit cycle type dynamics, this is the logical extension from the fixed point representation. The limit cycle should be a superior representation of recognition as it is able to represent more information about the system than merely the location of the local minima (fixed point corresponding to the most prevalent stimuli that resides within a class). The limit cycle has both frequency and amplitude that may be defined by the state of the system in order to create a unique representation of any class. These components are important if we are to build a non-linear system that illustrates the interaction of the various phase space objects. This notion will later become clear in Chapter 5 when we consider synchronization dynamics. In this Chapter we shall consider the first step toward generating these higher order dynamics, representing recognition via a limit cycle. Towards this goal we focus on the van der Pol oscillator, a system identified and discussed in section 1.4.2. We relate the van der Pol oscillator to a potential $V(t, x)$ created by a one-dimensional stimuli. Here we focus on how recognition may be conceptualized by oscillatory motions.

4.2 Utilizing the van der Pol System to Describe Recognition

The van der Pol oscillator is widely utilized as a basis for biological models that incorporate oscillatory dynamics [156, 39, 151, 136, 51]. Due to its pedigree in this domain we shall continue to pursue this convention. We have already discussed the critical aspects of this system in section 1.4.2 and highlighted the parameter range over which limit cycle behaviour is prevalent. At this juncture we are required to relate the well established foundations of the VDP model to the new venture of characterising the recognition of stimuli for this new learning system. The basic premise behind the memory of our learning model is that the energy potential $V(t, \mathbf{x})$ possesses local

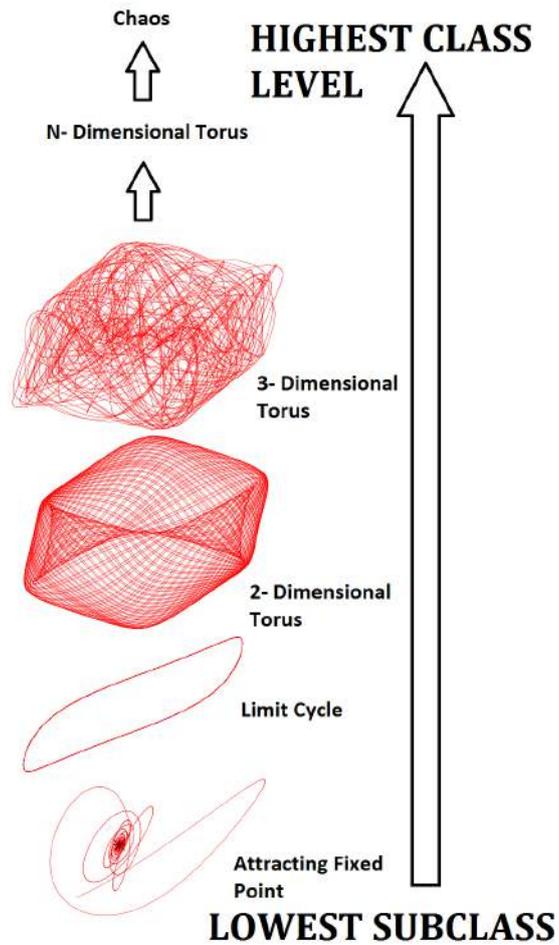


Figure 4.2: Illustration of the possible organisation of phase space attractors which may characterize the relation between different classes of the system utilizing a hierarchical framework. The lowest subclass is represented by a fixed point, higher/wider classes are represented firstly by limit cycles and then by quasi-periodic orbits.

minima formed as a result of the repeated application of stimuli. These local minima represent the most prevalent stimuli, and hence the most important retained information. At this stage we wish to relate the potential of this learning model to the potential of the van der Pol system.

As shown in section 1.4 the total energy of the VDP dynamical system may be represented by a kinetic energy plus potential energy function, the minimum of the potential function corresponds to the fixed point of the system. For the VDP oscillator the potential $V(x)$ is described by a symmetric parabola scaled by the constant $\frac{\omega^2}{2}$.

$$V(x) = \int g(x) dx = \int \omega^2 x dx = \frac{1}{2} \omega^2 x^2 \quad (4.1)$$

We may consider the behaviour of the VDP system if the potential is no longer parabolic. We want the potential of the oscillatory system to take any arbitrary topology defined by the distribution of stimuli. In this way we may relate our learning system to the van der Pol oscillator. The potential utilized by the van der Pol scheme would become a time dependent function, with as many local minima as dictated by the distribution of the stimuli. We postulate that a Hopf bifurcation about any fixed point of the system (defined by $V(t, x)$) will result in a unique limit cycle that represents a corresponding class. Before advancing to this complex case, it is prudent to begin with a simplified test example. Let us consider the dynamics of the VDP oscillator when the potential is described by the fourth order double well polynomial:

$$V(x) = d \left(\frac{1}{4} x^4 - \frac{1}{3} x^3 (x_1 + x_2 + x_3) + \frac{1}{2} x^2 (x_1 x_2 + x_1 x_3 + x_2 x_3) - x (x_1 x_2 x_3) \right) \quad (4.2)$$

This polynomial, with appropriate choices of x_1 , x_2 and x_3 may be perceived as a representation of a potential containing two local minima. The system will contain three fixed points, that correspond to x_1 , x_2 and x_3 . Such a polynomial may be considered to possess the minimum number of fixed point to define a double well profile. Our choice of a double well is intended to be the simplest possible complication beyond a parabola. It is speculated that the system will possess two limit cycles centred about x_1 and x_3 , corresponding to the two wells of the polynomial. If this is the case we shall need to consider the extension of this method to an N^{th} order polynomial, before then considering the time evolving potential defined by the distribution of stimuli. Defining the potential with the polynomial (4.2) allows us to easily define and manipulate its derivative $\frac{dV}{dx}$. This is required to simulate van der Pol dynamics. Utilizing this double well potential our dynamical system becomes:

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \varepsilon (1 - x^2) y - \frac{dV}{dx} = \varepsilon (1 - x^2) y - d(x - x_1)(x - x_2)(x - x_3) \end{aligned} \quad (4.3)$$

It is apparent that the fixed points of the system are $(x_1, 0)$, $(x_2, 0)$, $(x_3, 0)$. Recall that fixed points correspond to locations where $\dot{x} = 0$ and $\dot{y} = 0$. If we take the Jacobian of the system we arrive at the characteristic equation:

$$\lambda^2 - \varepsilon (1 - x^2) \lambda + 2\varepsilon xy + d((x - x_1)(x - x_2) + (x - x_2)(x - x_3) + (x - x_1)(x - x_3)) \quad (4.4)$$

inserting initial conditions the eigenvalues of the fixed points are given by:

$(x_1, 0)$	
$\lambda_{1,2} =$	$\frac{\varepsilon(1-x_1^2) \pm \sqrt{\varepsilon^2(1-2x_1^2+x_1^4) - 4d(x_1-x_2)(x_1-x_3)}}{2}$
$(x_2, 0)$	
$\lambda_{1,2} =$	$\frac{\varepsilon(1-x_2^2) \pm \sqrt{\varepsilon^2(1-2x_2^2+x_2^4) - 4d(x_2-x_1)(x_2-x_3)}}{2}$
$(x_3, 0)$	
$\lambda_{1,2} =$	$\frac{\varepsilon(1-x_3^2) \pm \sqrt{\varepsilon^2(1-2x_3^2+x_3^4) - 4d(x_3-x_2)(x_3-x_1)}}{2}$

Table 4.1

By studying the eigenvalues of the systems fixed points presented in table 4.1, we can make some inferences about the behaviour of each fixed point for certain regions of the parameter space. To aid in this analysis we need to make the assumption that $x_1 < x_2 < x_3$. To further simplify this discussion we will also restrict ourselves to the case where $\varepsilon > 0$. If this were not the case the conditions for stability of $(x_1, 0)$ and $(x_3, 0)$ would simply be inverted (i.e stable when $|x_1|, |x_3| < 1$ and unstable when $|x_1|, |x_3| > 1$). This should become evident from the following discussion.

Fixed point with coordinate $(x_1, 0)$

The implication of the assumption that $x_1 < x_2 < x_3$ is that $4d(x_1 - x_2)(x_1 - x_2) < 0$ if $d < 0$. As $\varepsilon^2(1 - 2x_1^2 + x_1^4) > 0$ for any x_1 , we can assume that the eigenvalues of this fixed point are real if $d < 0$. We should also note that $\varepsilon(1 - x_1^2) = \sqrt{\varepsilon^2(1 - 2x_1^2 + x_1^4)}$. Then denoting $A = \varepsilon(1 - x_1^2)$ and $B = 4d(x_1 - x_2)(x_1 - x_2)$ we have $2\lambda_{1,2} = A \pm \sqrt{A^2 + |B|}$. Clearly $A < \sqrt{A^2 + |B|}$, we therefore have one real positive, and one real negative eigenvalue. Hence when $d < 0$ the fixed point x_1 is a saddle point. For the case where $d > 0$ there are two options. The term $4d(x_1 - x_2)(x_1 - x_2) > 0$ and so the eigenvalue will be complex if $4d(x_1 - x_2)(x_1 - x_2) > \varepsilon^2(1 - 2x_1^2 + x_1^4)$. In this circumstance the fixed point will be attracting or repelling depending on $\varepsilon(1 - x_1^2)$. Where $|x_1| > 1$ the fixed point will be an attracting spiral, alternatively if $|x_1| < 1$ we will observe a repelling spiral. Finally there is the possibility that $d > 0$ and $4d(x_1 - x_2)(x_1 - x_2) < \varepsilon^2(1 - 2x_1^2 + x_1^4)$. In this situation the eigenvalues will be real. We may then consider the eigenvalue denoted as $2\lambda_{1,2} = A \pm \sqrt{A^2 - |B|}$. Evidently $A > \sqrt{A^2 - |B|}$ and so the eigenvalue will be real with a sign determined by A . Hence if $|x_1| > 1$ we will encounter a stable node and if $|x_1| < 1$ the fixed point will be an unstable node.

Fixed point with coordinate $(x_2, 0)$

Again the function $\varepsilon^2(1 - 2x_2^2 + x_2^4) > 0$ for all x_2 . The function $4d(x_2 - x_1)(x_2 - x_3)$ will be positive if $d < 0$ or negative if $d > 0$. Therefore, if $d > 0$, $\varepsilon^2(1 - 2x_2^2 + x_2^4) - 4d(x_2 - x_1)(x_2 - x_3) > 0$ and the eigenvalues are real. We can again consider the eigenvalues to be of the form $2\lambda_{1,2} = A \pm \sqrt{A^2 + |B|}$ as $\varepsilon(1 - x_2^2) = \sqrt{\varepsilon^2(1 - 2x_2^2 + x_2^4)}$. Evidently $A < \sqrt{A^2 + |B|}$, hence the fixed point behaves as a saddle. When $d < 0$ the eigenvalues may still be real, this is the case when $\varepsilon^2(1 - 2x_2^2 + x_2^4) > 4d(x_2 - x_1)(x_2 - x_3)$. In this situation we should consider the eigenvalues represented as $2\lambda_{1,2} = A \pm \sqrt{A^2 - |B|}$. Clearly $A > \sqrt{A^2 - |B|}$ and so the fixed

point will be either a stable or unstable node. This again depends on $\varepsilon(1 - x_2^2)$, if $|x_2| > 1$ the fixed point will be stable, conversely the point will be unstable. Finally is the situation where the eigenvalue is complex. This occurs if $d < 0$ and $\varepsilon^2(1 - 2x_2^2 + x_2^4) < 4d(x_2 - x_1)(x_2 - x_3)$. The stability of the fixed point is again dependent on $\varepsilon(1 - x_2^2)$, however, now if $|x_2| > 1$ the point is a stable spiral and if $|x_2| < 1$ the point is an unstable spiral.

Fixed point with coordinate $(x_3, 0)$

Following the same reasoning used for the previous two fixed points the eigenvalues of this point will be real if $d < 0$ or $\varepsilon^2(1 - 2x_3^2 + x_3^4) > 4d(x_3 - x_1)(x_3 - x_2)$. If the latter is realized the stability will depend up x_3 . Where $|x_3| > 1$ the point will be characterized as a stable node, alternatively, the point will be an unstable node. Returning to the real eigenvalues that present when $d < 0$, we may obtain eigenvalues of the form $2\lambda_{1,2} = A \pm \sqrt{A^2 + |B|}$, hence the point becomes a saddle. Where $\varepsilon^2(1 - 2x_3^2 + x_3^4) < 4d(x_3 - x_1)(x_3 - x_2)$ the eigenvalue solutions are complex. The fixed point will be a stable spiral if $|x_3| > 1$, or, if $|x_3| < 1$, an unstable spiral will be observed.

$(x_1, 0)$				
$d > 0$				$d < 0$
$\varepsilon^2(1 - 2x_1^2 + x_1^4) > 4d(x_1 - x_2)(x_1 - x_3)$		$\varepsilon^2(1 - 2x_1^2 + x_1^4) < 4d(x_1 - x_2)(x_1 - x_3)$		Saddle
$ x_1 > 1$	$ x_1 < 1$	$ x_1 > 1$	$ x_1 < 1$	
Stable Node	Unstable Node	Stable Focus	Unstable Focus	
$(x_2, 0)$				
$d > 0$	$d < 0$			
Saddle	$\varepsilon^2(1 - 2x_2^2 + x_2^4) > 4d(x_2 - x_1)(x_2 - x_3)$		$\varepsilon^2(1 - 2x_2^2 + x_2^4) < 4d(x_2 - x_1)(x_2 - x_3)$	
	$ x_2 > 1$	$ x_2 < 1$	$ x_2 > 1$	$ x_2 < 1$
	Stable Node	Unstable Node	Stable Focus	Unstable Focus
$(x_3, 0)$				
$d > 0$				$d < 0$
$\varepsilon^2(1 - 2x_3^2 + x_3^4) > 4d(x_3 - x_1)(x_3 - x_2)$		$\varepsilon^2(1 - 2x_3^2 + x_3^4) < 4d(x_3 - x_1)(x_3 - x_2)$		Saddle
$ x_1 > 1$	$ x_3 < 1$	$ x_3 > 1$	$ x_3 < 1$	
Stable Node	Unstable Node	Stable Focus	Unstable Focus	

Table 4.2: Classification of the fixed points $(x_1, 0)$, $(x_2, 0)$ and $(x_3, 0)$ of equation (4.3). The eigenvalues of each fixed point as shown to depend on ε , d , x_1 , x_2 and x_3 . The relationship between eigenvalues and the behaviour of a fixed point is detailed in section 1.4.2.

The possible behaviours of each of the fixed points are summarized in table 4.2. We should remember to be cautious as the overall dynamics of the system may be radically different to the local fixed point behaviour. For non-linear dynamical systems there is typically an interplay between the local dynamics surrounding the fixed point and the global behaviour. The Hopf bifurcation is a type of local bifurcation that is typically observed in van der Pol systems. This bifurcation occurs when the pair of eigenvalues of a fixed point form a complex conjugates pair, which crosses the line $Re\lambda_{1,2} = 0$. Considering the fixed points of our system we can see that a Hopf bifurcation may occur for any of these points as a focus type attractor losses stability. Further, we may expect to see bi-stability when $d > 0$, $|x_1| > 1$ and $|x_2| > 1$. We may anticipate that the phase space will be separated into two basins of attraction encapsulating the fixed points $(x_1, 0)$ and $(x_2, 0)$. There is also the possibility that one of these fixed points may be attracting whilst

the other meets the criteria for a Hopf bifurcation. In this circumstance we may again expect to observe bi-stability, one attractor being a stable cycle, the other a fixed point. In this case we may speculate about a further cycle/orbit that acts as a separatrix. The possibilities indicated by the eigenvalues of the fixed points demonstrate that even in this contrived, double well potential form, there are some interesting dynamics to understand. Although we can understand each fixed point in turn, how trajectories interact over the global phase space is less clear. We should consider the topology of any limit cycles and understand how these interact with the fixed points of the system. Of particular interest for further analysis is bi-stability. As such we should restrict our attention to the case where $d > 0$. This is also necessary if we require our potential to have two local minima (rather than two maxima). To validate our assertions and investigate the global behaviour of the system we may utilize numerical methods and simulation. At this stage we turn to XPPAUT continuation software [44] to illustrate the trajectories of the system and to track bifurcations.

4.2.1 Numerical Calculations

It is already becoming apparent that utilizing a potential with just two wells yields a system that is relatively complex. Such a system will have five bifurcation parameters $(x_1, x_2, x_3, d, \varepsilon)$. The standard approach to observe Hopf bifurcations in the VDP system is to adjust the parameter ε , a Hopf bifurcation then occurs when ε crosses the line $\varepsilon = 0$. If we combine this oscillatory system with the memory foam model potential, the parameters x_1, x_2 and x_3 will be determined by the distribution of stimuli. They will therefore be outside of our control. It is clear that they have a significant impact on the bifurcation structure of the system. This may result in difficulties as dynamic behaviour may be based on position rather than controllable parameters. We may also be sceptical about extending this approach to a potential with N fixed points.

It is logical to impose the constraints $x_1 < x_2 < x_3$, $\varepsilon > 0$ and $d > 0$ and consider the system's behaviour for fixed values of ε and d . A Hopf bifurcation is predicted around $(x_3, 0)$ when $|x_3| = 1$ and $\varepsilon^2 (1 - 2x_3^2 + x_3^4) < 4d(x_3 - x_1)(x_3 - x_2)$. The same can be said for $(x_1, 0)$ when $|x_1| = 1$ and $\varepsilon^2 (1 - 2x_1^2 + x_1^4) < 4d(x_1 - x_2)(x_1 - x_3)$. Note that both of these conditions may be met simultaneously. To validate our assertions about the fixed points of the system we may consider the phase space dynamics for the various parameter values listed in table 4.3. Each parameter set corresponds to a different arrangement of fixed point types. Because $d > 0$ the fixed point $(x_2, 0)$ is always a saddle. A sample of phase space trajectories for each of the parameters listed in table 4.3 are illustrated in figure 4.3. It is apparent that the saddle point plays an important role in partitioning the phase space. Each diagram shows the stable (blue) and unstable (yellow) manifolds of the saddle point, as well as including a cross-section of trajectories that converge to the fixed point $(x_1, 0)$ (black lines), fixed point $(x_3, 0)$ (red lines) or stable limit cycle (cyan lines). The locations of each fixed point are marked by coloured squares, cyan denotes a saddle point, red denotes a stable node, green a unstable node, blue a stable focus and yellow denotes an unstable focus.

It is clear that the fixed points behave as expected, however, some important comments should be made. Firstly we may comment on the case where both fixed points are attracting (nodes or focus), in this instance the basins of attraction of the two fixed points are separated by the stable manifold of the saddle point. Due to the aforementioned Poincaré - Bendixson theorem, trajectories cannot intersect, the phase space is therefore neatly separated into two distinct basins. This may be appreciated from figures 4.3a,c,i,k. It would also appear that if the system contains

an unstable fixed point (node or focus) the stable manifold of the saddle point will always interact with this point. Most important to our study is the occurrence of limit cycles. If both points $(x_1, 0)$ and $(x_3, 0)$ are unstable, the system will converge to a stable limit cycle. This cycle exists even when both fixed points are unstable nodes (figure 4.3f). This suggests that the bifurcation is not always a Hopf bifurcation as the eigenvalues of the fixed point do not contain an imaginary part. In each case this stable cycle is approached by the unstable manifold of the saddle point. In all circumstances the system appears to contain no more than a solitary limit cycle that encapsulates all three fixed points. Our speculation that a Hopf bifurcation about either fixed point (whilst the other remains stable) could lead to multiple distinct cycles appears to be unfounded (as shown by figures 4.3d,l,m,o). Instead, a limit cycle bifurcation always results in a cycle that encircles all three fixed points, a feature that does not easily lend itself to a robust recognition scheme. Despite not realizing the co-existence of multiple cycles, we do observe bi-stability in the form of a stable focus co-existing with a limit cycle (figure 4.3o). The stable manifold of the saddle point encapsulates the attracting focus, forming a separatrix. One unstable manifold of the saddle interacts with the stable focus whilst the other unstable manifold organises an attracting limit cycle trajectory.

Fixed Point		Parameter Values				
$(x_1, 0)$	$(x_3, 0)$	ε	d	x_1	x_2	x_3
Stable Node	Stable Node	2	0.05	-1.2	-0.1	1.2
Stable Node	Unstable Node	2	0.1	-1.2	-0.2	0.7
Stable Node	Stable Focus	2	0.05	-1.2	0.1	1.1
Stable Node	Unstable Focus	2	0.05	-1.2	0.1	0.9
Unstable Node	Stable Node	2	0.1	-0.7	0	1.3
Unstable Node	Unstable Node	2	0.3	-0.5	0.1	0.7
Unstable Node	Stable Focus	1.5	0.3	-0.5	0.1	1.1
Unstable Node	Unstable Focus	1.5	0.3	-0.5	0.1	0.9
Stable Focus	Stable Node	3	0.1	-1.1	0.6	1.2
Stable Focus	Unstable Node	3	0.1	-1.1	0.6	0.8
Stable Focus	Stable Focus	0.5	0.5	-1.2	0.1	1.1
Stable Focus	Unstable Focus	0.5	0.5	-1.3	0.1	0.7
Unstable Focus	Stable Node	2	0.1	-0.7	0.9	1.1
Unstable Focus	Unstable Node	2	0.04	-0.9	0.5	0.9
Unstable Focus	Stable Focus	0.5	0.5	-0.8	0	1.2
Unstable Focus	Unstable Focus	0.5	0.5	-0.8	0	0.8

Table 4.3: The types of fixed point behaviour demonstrated by $(x_1, 0)$ and $(x_3, 0)$ for the listed parameter values of ε , d , x_1 , x_2 and x_3 . These parameters are chosen to demonstrate every possible combination of fixed point behaviour. In figure 4.3 we use this list to simulate the dynamics of the system in order to numerically validate our analytical analysis.

Due to the high number of bifurcation parameters a complete bifurcation analysis would prove to be very complex. What we have observed is that a bifurcation occurs at $|x_1| = 1$ and $|x_3| = 1$. If either $|x_1| > 1$, $|x_3| < 1$ or $|x_1| < 1$, $|x_3| > 1$ this bifurcation may result in bi-stability. One fixed point remains attracting whilst a stable limit cycle encapsulates the three fixed points. Basins of attraction are partitioned by a separatrix formed by the stable manifold of the saddle point. It is also possible that despite these conditions the stable cycle does not appear. Instead all trajectories may converge to the stable fixed point. The bifurcation to a stable cycle is therefore dependent on the interaction of all parameters ε , d , x_1 , x_2 and x_3 . What would appear evident is that a stable limit cycle is the only attractor when both fixed points $(x_1, 0)$ and $(x_3, 0)$ are unstable. This

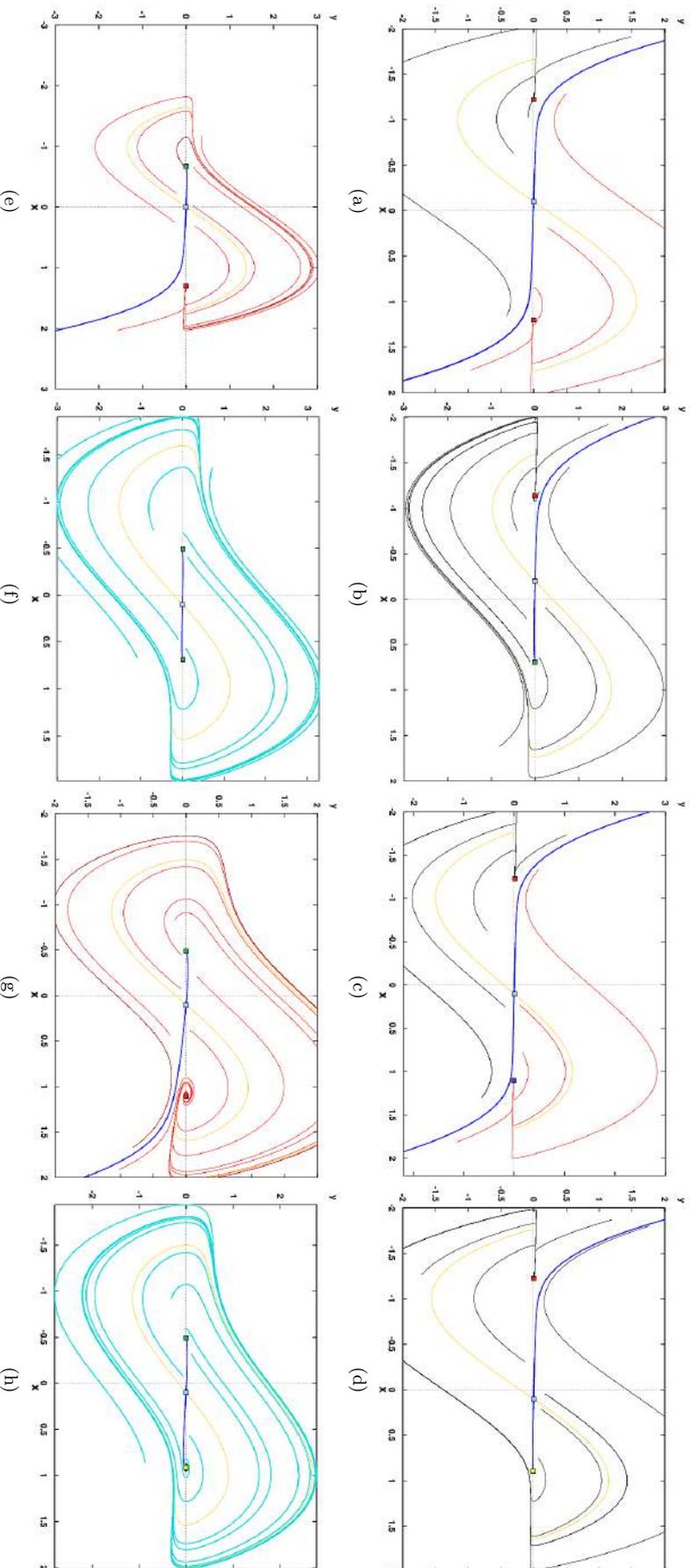


Figure 4.3: Demonstrations of various trajectories on the phase space when parameters correspond to those detailed in table 4.3. These parameters are chosen to demonstrate the range of possible behaviours that the system may illicit. Note that the fixed point $(x_2, 0)$ is always a saddle point. In each diagram the stable (blue) and unstable (yellow) manifolds of the saddle point are demonstrated. It is apparent that these manifolds play an important role in partitioning the phase space. Trajectory that converge to the fixed point $(x_1, 0)$ are marked in black whilst trajectories that converge to $(x_3, 0)$ are red. Cyan trajectories converge to a stable limit cycle that encapsulates the 3 fixed points. The three fixed points of the system are marked by coloured squares. Cyan denotes a saddle point, red denotes a stable node, green a unstable node, blue a stable focus and yellow denotes an unstable focus. **(a):** The fixed points $(x_1, 0)$ and $(x_3, 0)$ are both stable nodes. **(b):** $(x_1, 0)$ is a stable node, $(x_3, 0)$ is an unstable node. **(c):** $(x_1, 0)$ is a stable node, $(x_3, 0)$ is a stable focus. **(d):** $(x_1, 0)$ is a stable node, $(x_3, 0)$ is an unstable focus. **(e):** $(x_1, 0)$ is an unstable node, $(x_3, 0)$ is a stable node. **(f):** $(x_1, 0)$ and $(x_3, 0)$ are both unstable nodes. **(g):** $(x_1, 0)$ is an unstable node, $(x_3, 0)$ is a stable focus. **(h):** $(x_1, 0)$ is an unstable node, $(x_3, 0)$ is an unstable focus.

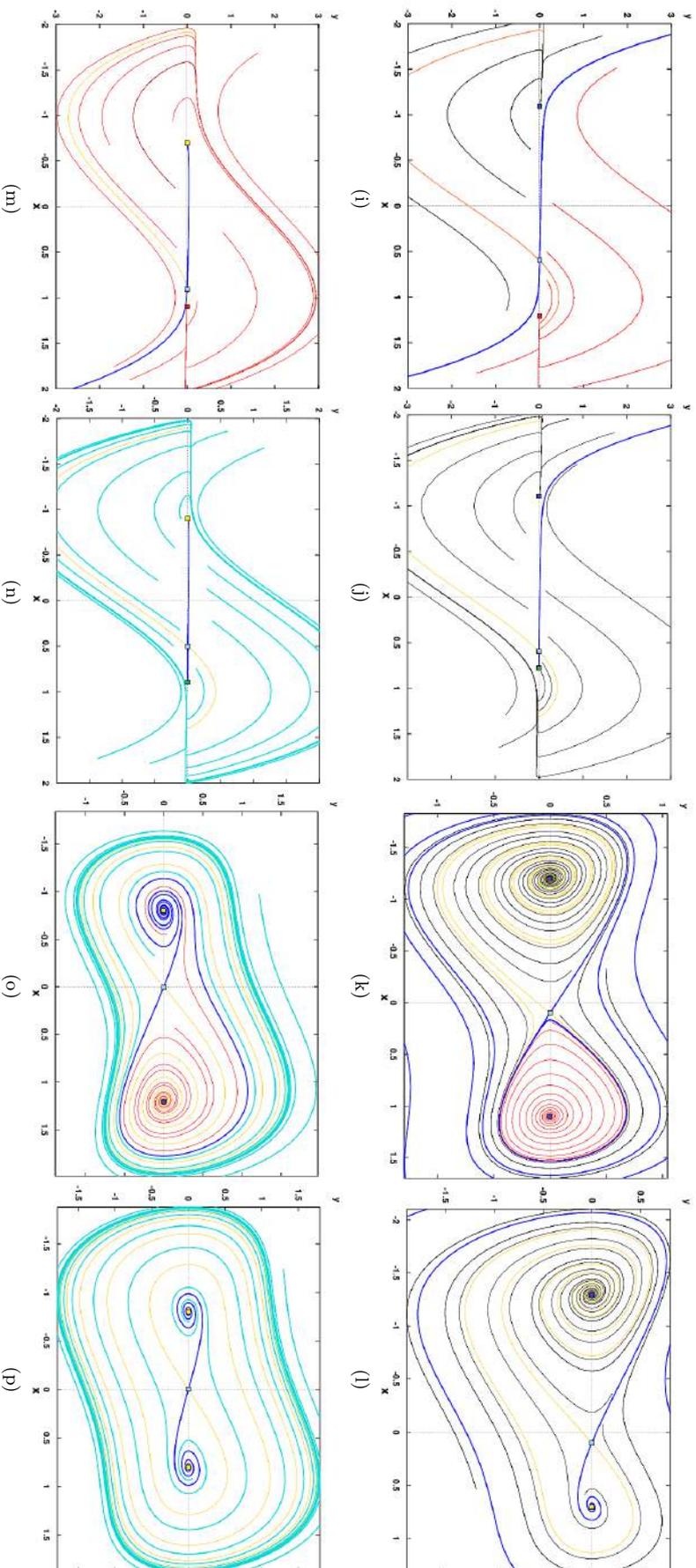


Figure 4.3: **(i)**: $(x_1, 0)$ is a stable focus, $(x_3, 0)$ is a stable node. **(j)**: $(x_1, 0)$ is a stable focus, $(x_3, 0)$ is an unstable node. **(k)**: $(x_1, 0)$ and $(x_3, 0)$ are both stable foci. **(l)**: $(x_1, 0)$ is a stable focus, $(x_3, 0)$ is an unstable focus. **(m)**: $(x_1, 0)$ is an unstable focus, $(x_3, 0)$ is a stable node. **(n)**: $(x_1, 0)$ is an unstable focus, $(x_3, 0)$ is an unstable node. **(o)**: $(x_1, 0)$ is a stable focus, $(x_3, 0)$ is a stable focus. **(p)**: $(x_1, 0)$ and $(x_3, 0)$ are both unstable foci.

behaviour appears to endure even when both points are unstable nodes. This suggests that the Hopf bifurcation may not be the only route by which a stable limit cycle emerges in this system.

Considering the dynamics that we have thus far observed, we may be perturbed from further consideration of a system of this type. The main deficiency that has become apparent is the lack of distinct local bifurcations leading to new unique limit cycles. The system seems to possess an inherent stable cycle that acts as a global attractor for certain parameter choices. Although the dynamics of this system seem interesting, they do not endear themselves to the recognition scheme that we wish to formulate and apply to the memory foam model. Foremost, we have not observed cycles characterising individual local minimum, this is a clear requirement from both the perspective of recognition and instigating higher order dynamical behaviour.

The standard VDP oscillator utilizes a parabolic potential, our deviation away from this to a double well does not illicit the results desired. We must now attempt to devise an alternative strategy to generate a series of unique attracting cycles. We may consider a system where each local minima of $V(t, x)$ is characterized by a parabolic potential with characteristic's dependent on the local shape of the potential $V(t, x)$. A VDP oscillator that utilized such a potential, would, given an appropriate value of ε , demonstrate oscillatory dynamics with a frequency relevant to the local topology. By only utilizing local information about a minimum, the potential $V(t, x)$ can be modelled as a set of parabolas. As such, oscillations should be predictable and controllable, whilst meeting our requirement for every local minimum to be associated with a distinct cycle.

4.3 Modelling the Foam Profile with Parabolas

A one-dimensional profile that contains several local minima may be approximately modelled as a series of parabolas, each centred about a minimum with a width dependent on each minima's local topology. This is advantageous as the traditional formulation of the van der Pol system (equation 1.19) can then be employed. If the potential of the van der Pol system is parabolic, then it follows that the dynamics of the system will be similar to those discussed in section 1.4.2. For $\varepsilon > 0$ we may anticipate that each minimum will be associated with a limit cycle. Because the potential of the van der Pol system reflects the local topology of the one-dimensional profile we may expect that each limit cycle has a frequency that relates to the shape of the profile. This concept is visualized in figure 4.4. The black line indicates an arbitrary potential and the three coloured lines are appropriate parabolas that may be used to approximately model this potential. Advancing from this arbitrary case, we may postulate a system where our energy potential, modulated by the repeated application of stimuli, is modelled as a set of parabola's, which in turn act as the potential of a van der Pol system characterizing recognition. By relating the energy potential of the learning system to the VDP potential we should be able to infer the structure of a class from a phase space trajectory. If we consider each minimum to describe a class, then we may surmise that the curvature of the potential, in the vicinity of the local minimum, provides an estimate of the size of the class.

It is possible to construct these parabolas about any point $x_* \in x$ associated with the profile $V(t, x)$ using a Taylor expansion. In order to centre the parabola about x_* the first order term of the expansion is neglected, we also omit all higher order terms. Hence we obtain:

$$U(V, x) = V(x_*) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} (x - x_*)^2 \quad (4.5)$$

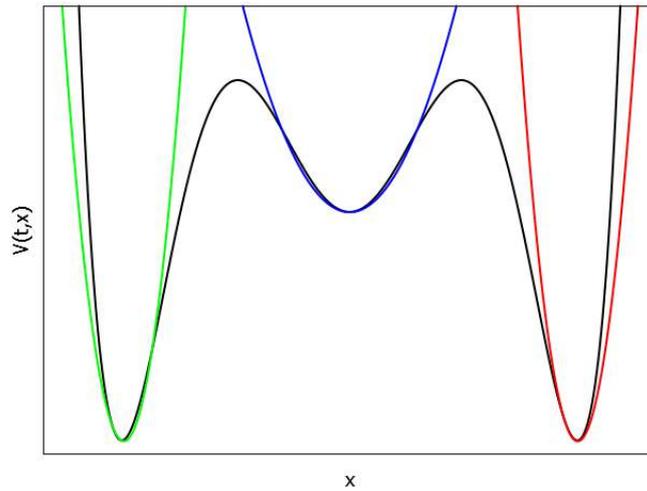


Figure 4.4: An arbitrary potential (black) with parabolas (coloured lines) centred about the local minimas. Correct scaling of the parabolas width ensures that a series of parabolas can vaguely represent the primary features of the potential.

$U(V, x)$ is a parabola centred about the point x_* , its width is dependent on $\frac{\partial^2 V}{\partial x^2}$ and as such will

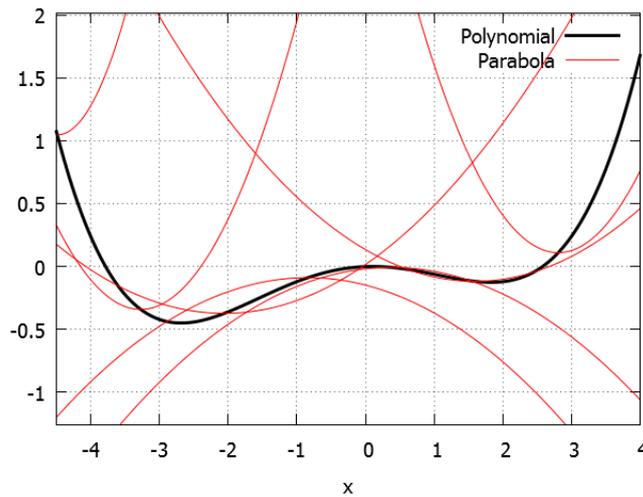


Figure 4.5: Parabolas (red lines) constructed at various points of a polynomial (black line) by utilizing equation (4.5). The width of each parabola is dependent on the curvature of the polynomial (at the intersection point between the polynomial and a parabola), this is defined by $\frac{d^2 V}{dx^2}$. Where this derivative is negative the associated parabola is in a undesirable orientation for application to a VDP system.

be dependent on the distribution of stimuli received by the system. Note that the potential remains time dependent and as such $U()$ will share this property. Figure 4.5 demonstrates this transform applied to a basic double well profile. The same procedure is applicable to represent any continuous potential. There remains a critical oversight of this transform. The typical potential of the VDP oscillator, as described by equation (4.1), contains the term $\frac{\omega^2}{2}$, this is always positive. This ensures that the potential is always orientated with the ‘cup’ facing upwards. It is evident from figure 4.5 that $\frac{\partial^2 V}{\partial x^2}$ may be negative and so the parabola would be inverted when compared to the shape of the potential of the VDP oscillator. Application of the transform (4.5) would lack smoothness

and continuity when applied to an evolving system at the point where the slope changes sign. This obstacle can be overcome if we instead consider the modulus of the second derivative. The downside of such an approach is that the resultant function would then lack uniqueness, positive and negative slopes would be considered equivalent, hence the information encoded by frequency would lose its purpose. The integral of the sigmoid function, described by equation (4.6) possesses the following properties and therefore provides a viable solution to overcoming this issue.

- For negative values of $\frac{\partial^2 V}{\partial x^2}$ the function $g(V'')$ asymptotically tends to a small positive constant. As a result, all parabolas in this region have maximal width and are upright.
- For positive values of $\frac{\partial^2 V}{\partial x^2}$ the transformation asymptotically tends to $\frac{\partial^2 V}{\partial x^2}$, the frequency is coded by the curvature of the potential.

We must clearly scale and shift this transform so that it is relevant to the scale of our system. The transform presented in equation (4.6), and depicted graphically in figure 4.6a includes a scaling parameter s and a shift parameter φ .

$$g(V'') = \frac{1}{s} \left(\log \left(1 + e^{(sV'' - \varphi)} \right) + \varphi \right) \quad (4.6)$$

All points associated with negative $\frac{d^2 V}{dx^2}$ now tend towards a small positive constant. This is not an issue as these points will not typically reside in close proximity to the fixed point of the system. As the local minimum is approached, $\frac{d^2 V}{dx^2}$ will become positive and the frequency of oscillation will become dependent on the shape of the potential ($g\left(\frac{d^2 V}{dx^2}\right) \rightarrow \frac{d^2 V}{dx^2}$ for $\frac{d^2 V}{dx^2} > 0$).

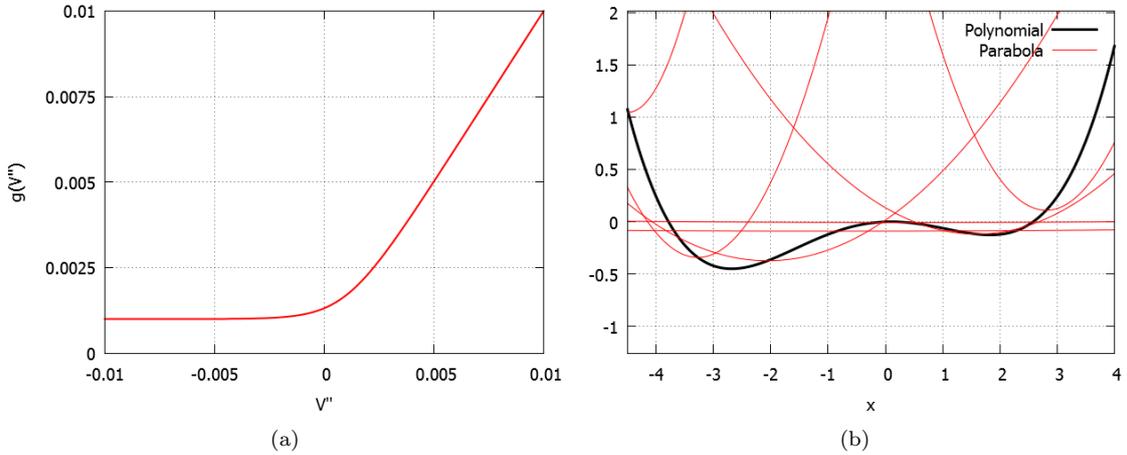


Figure 4.6: **(a)**: Transformation $g()$ of the second derivative as defined by equation (4.6). This ensures that all constructed parabolas are in the correct orientation. Here we utilize the parameter values $s = 1000$ and $\varphi = 1$. For negative values of $\frac{d^2 V}{dx^2}$ the function tends to a small positive constant, whereas, when $\frac{d^2 V}{dx^2} > 0$, $g\left(\frac{d^2 V}{dx^2}\right) \rightarrow \frac{d^2 V}{dx^2}$. **(b)**: Parabolas constructed at various points of a polynomial by utilizing equation (4.5). These parabolas are all in the same orientation as the potential of a typical VDP system.

Utilizing this function, equation (4.5) becomes:

$$U(V, x) = V(x_*) + \frac{1}{2} g(V'') (x - x_*)^2 \quad (4.7)$$

Applying this transform to the same double well profile as shown in figure 4.5 this addition ensures that all parabolas take the requisite upright form. This is demonstrated in figure 4.6b. These parabolas are completely dependent on the profile $V(t, x)$ and offer the advantage that they allow the potential to be represented by a curve comparable to that normally characterising the potential of the VDP oscillator. For the VDP oscillator it is known that limit cycles emerge from a Hopf bifurcation, these cycles encapsulate the fixed point. For limit cycles to encapsulate the fixed points corresponding to our local minimum we must shift the fixed points of the VDP system. This may be achieved by exchanging the term x with $(x - x_*)$. Here x_* is a point that approaches the local minimum. There still remains the issue of locating this local minima. To overcome this we can continue to apply gradient descent and describe the recognition of a stimulus by three first order differential equations as described in equation (4.8). For appropriate values of ε we expect all trajectories on the plane x, y to be cyclic. The shape and frequency of each orbit will be determined by the properties of the learnt classes. The system should converge towards a stable limit cycle trajectory as the potential $V(t, x)$ approaches the probability density distribution of the stimuli.

$$\begin{aligned}\frac{dx_*}{dt} &= -\gamma \frac{dV}{dx} \\ \frac{dx}{dt} &= y \\ \frac{dy}{dt} &= \varepsilon \left(1 - (x - x_*)^2\right) (y - y_*) - \frac{dU}{dx}\end{aligned}\tag{4.8}$$

The additional term y_* arises from the inclusion of x_* . Note that $\dot{x}_* = y_*$.

4.3.1 Parabola width and Frequency

We have previously presented the issues of parabola approximation in reference to improving smoothness and ensuring correct classification. A further adjustment that can be made involves scaling the frequency. By increasing the frequency of oscillation we ensure that the time scale of oscillation is greater than the time scale on which $V(t, x)$ evolves. In the event that the two are similar, it may not be possible to visibly observe oscillations. In this circumstance oscillations will only become evident once the potential approaches a more stationary profile. Conversely, this may be seen as an advantageous guide to judging the certainty of recognition. Regardless of this, we should consider how frequency may be described. Typically the frequency of an oscillator is described by:

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{\partial^2 U}{\partial x^2}}\tag{4.9}$$

We can increase the frequency by introducing the parameter α to equation (4.8) as follows:

$$\ddot{x} - \varepsilon \left(1 - (x - x_*)^2\right) (\dot{x} - \dot{x}_*) + \alpha \frac{\partial U}{\partial x} = 0\tag{4.10}$$

The frequency of oscillation will now be given by:

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi} \sqrt{\alpha \frac{\partial^2 U}{\partial x^2}}\tag{4.11}$$

Because α only acts as a scaling parameter each position on $V(x, t)$ is still associated with a unique frequency of oscillation (when U'' is positive) that is related to the curvature of the potential. Next

we shall demonstrate limit cycles in the systems phase space and show how these orbits relate to the development of classes.

4.4 Limit Cycles Recognition Scheme

In figures 4.7 and 4.8 six examples of recognition represented by limit cycles are illustrated. The stimuli are generated utilizing the method introduced in section 2.2.1 and are described by equation (2.8). The utilized functions again correspond to the those described in equation (2.9). These stimuli distributions are intended to be suitably different so as to highlight the versatility of this approach. In each visualization $\sigma = \sqrt{0.05}$, $\alpha = 1$, $D = 5$ and $\gamma = 1$. As predicted, when $\varepsilon > 0$ limit cycle oscillations are guaranteed. The time scales on which these oscillations emerge depends upon the vector field of the systems potential. If the potential changes configuration rapidly then the frequency and location of oscillations will also be adjusted at a similar rate. The frequency of oscillation can be considered to present additional information about the class of recognized stimuli. A potential well that is narrow will have a higher $\frac{\partial^2 U}{\partial x^2}$ term, this class will therefore be realized by an oscillator with a higher frequency. A narrow potential well implies that the the system has experienced a set of stimuli all within a narrow bandwidth, as such the class is highly distinct. We may also consider that oscillations with higher frequency reflect the identification of a stimulus that the system is more certain about. A lower level of identification accuracy may be implied when a stimulus is recognized by a trajectory with lower frequency.

Each row of figure 4.7 and 4.8 pertains to a different stimuli. This is evident from the different distributions of potential shown in the first column. Included for each stimuli is a demonstration of the oscillatory dynamics against time, a plot of the phase plane (x, y) after a semi-stable state has been reached, and an indication of the instantaneous frequency of each cycle. The colour of each cycle in the phase space (x, y) (column 3) is coded to the corresponding colour for the instantaneous frequency demonstration. As should be expected, larger limit cycles are indicative of higher frequency. Recognition trajectories are introduced at $t = 0$ and evolve in conjunction with the shaping of the potential. These ten trajectories take initial conditions evenly interspersed across the range of x . They are chosen like this to ensure that every minimum of the potential is demonstrated by a cycle. The gradient aspect of the system is illustrated at the base of each plot in the first two columns of figures 4.7/4.8. We have chosen to retain aspects of gradient descent to adjust the location of the limit cycle to coincide with the local minima of the potential, the additional knowledge regarding frequency yields useful information about the size of each basin of attraction surrounding a local minimum. An alternative approach would be to neglect the shifting approach introduced by gradient descent and utilize only frequency to represent a class. Stimuli that induce oscillations of equal frequency being considered as part of the same class. We should be aware that such an approach may fail if two classes are described by potential wells with very similar topologies. This issue is highlighted in figures 4.8k,l. Although four distinct cycles are observed in 4.8k, two of these cycles have very similar frequencies as shown in 4.8l. Without considering the location of the cycle in the phase space we would struggle to differentiate between two of these different classes. The same number of stimuli and the same time period ($t = 4000 \rightarrow 5000$) is used to generate the trajectories shown in the third column of each example. It is clear from these examples that the potential of the system is not completely stationary in some of these illustrations. The limit cycles shown in the third column are not always smooth closed curves. We are aware that these cycles must be limit cycles due to the restrictions of the systems dimensions.

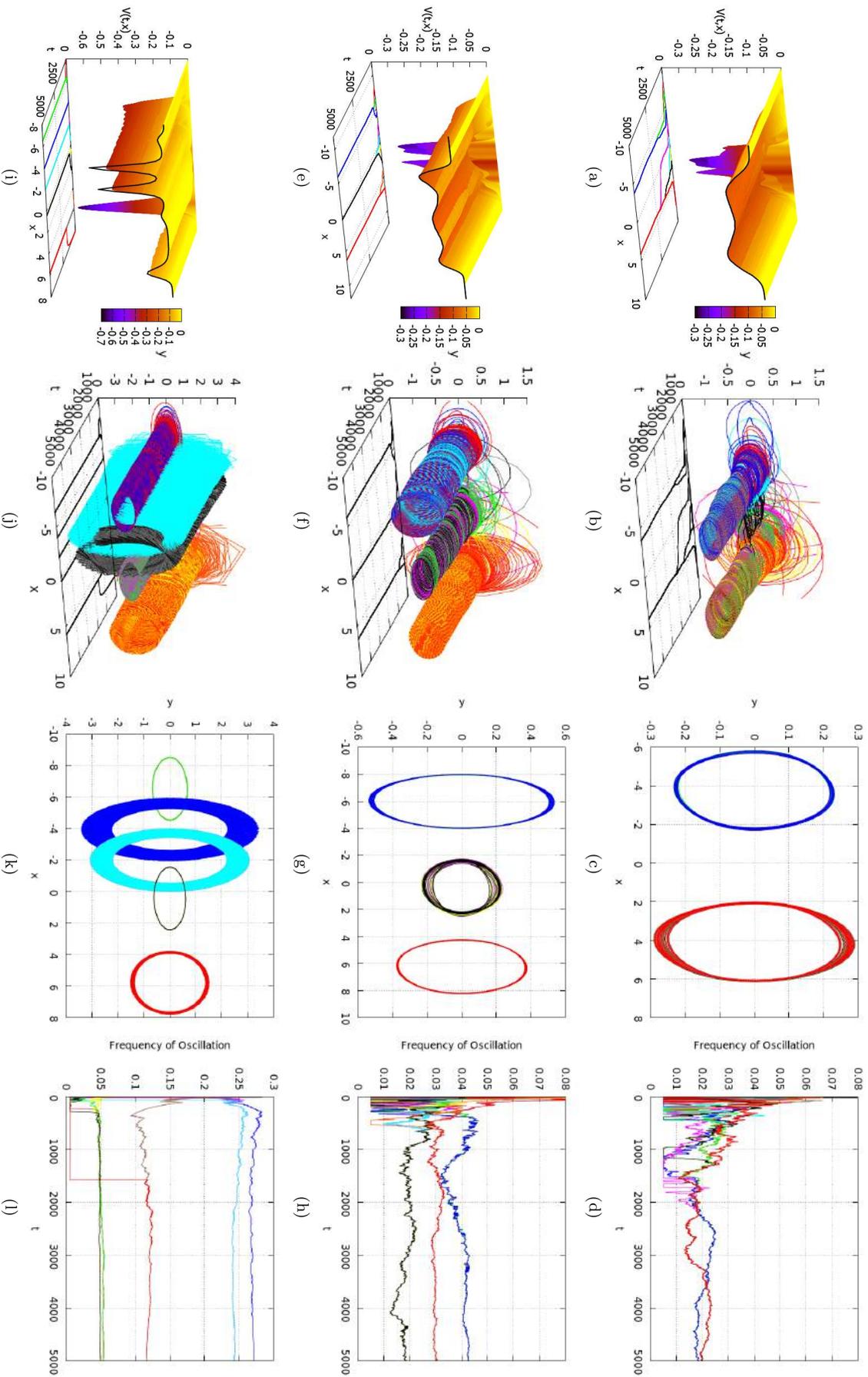


Figure 4.7: Demonstrations of limit cycles characterising the recognition of stimuli. Each row corresponds to a differently distributed stimulus that is simulated using equation (2.8). In each case the stimuli noise intensity $D = 5$. The effect of the different stimuli is evident from the shapes of the potentials shown in the first column. In the first row $\frac{dF}{d\eta}$ is given by equation (2.9a), in the second row $\frac{dF}{d\eta}$ is associated with equation (2.9b) and in the third row $\frac{dF}{d\eta}$ is given by equation (2.9c). The second column depicts the recognition of ten initial conditions as they converge to stable limit cycle trajectories. Gradient descent trajectories are included at the base of each figure to illustrate how the differential equations (4.8) influence each other. The stable behaviour of the attracting limit cycles is better illustrated in third column, here the final 1000 time iterations are overlaid on the phase plane x, y . These phase dynamics can only be observed when the system has reached a stationary state. The final column introduces the frequency of each trajectory as a function of time. Note that colours correspond to column 3. In each example the Gaussian width $\sigma = \sqrt{0.05}$, the recognition rate parameter $\gamma = 1$, the VDP parameter $\varepsilon = 0.01$ and the frequency scaling parameter $\alpha = 1$.

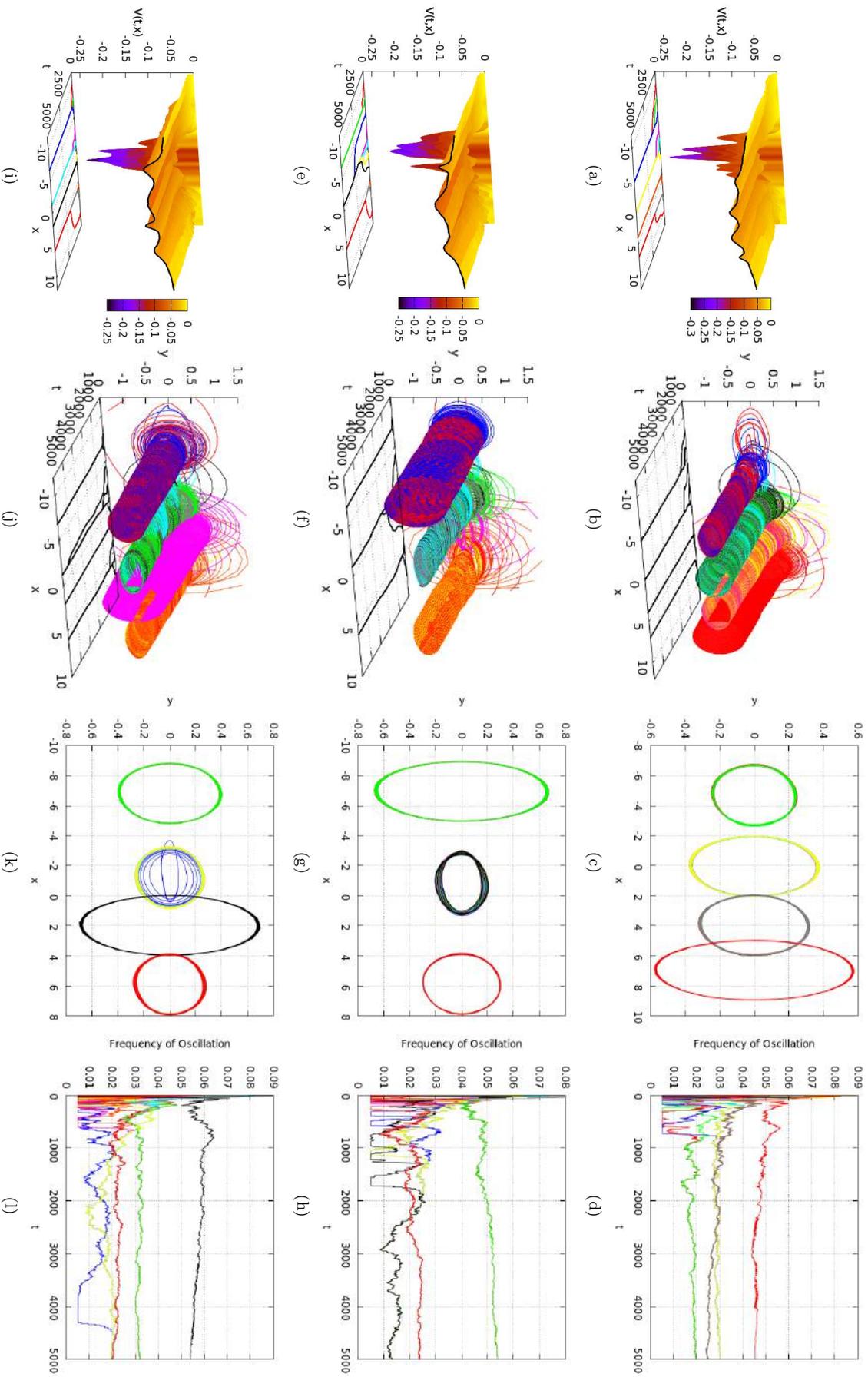


Figure 4.8: Demonstrations of limit cycles characterising the recognition of stimuli. Each row corresponds to a differently distributed stimulus that is simulated using equation (2.8). In each case the stimuli noise intensity $D = 5$. The effect of the different stimuli is evident from the shapes of the potentials shown in the first column. In the first row $\frac{dF}{d\eta}$ is given by equation (2.9d), in the second row $\frac{dF}{d\eta}$ is associated with equation (2.9e) and in the third row $\frac{dF}{d\eta}$ is given by equation (2.9f). The second column depicts the recognition of ten initial conditions as they converge to stable limit cycle trajectories. Gradient descent trajectories are included at the base of each figure to illustrate how the differential equations (4.8) influence each other. The stable behaviour of the attracting limit cycles is better illustrated in third column, here the final 1000 time iterations are overlaid on the phase plane x, y . These phase dynamics can only be observed when the system has reached a stationary state. The final column introduces the frequency of each trajectory as a function of time. Note that colours correspond to column 3. In each example the Gaussian width $\sigma = \sqrt{0.05}$, the recognition rate parameter $\gamma = 1$, the VDP parameter $\epsilon = 0.01$ and the frequency scaling parameter $\alpha = 1$.

It is certainly not possible that they are quasi-periodic. Increasing the duration of the stimuli and improving the resolution of our numerical method would cause the unusual trajectories (4.7g,k for example) to collapse onto a conventional limit cycle.

Because the dynamics of the VDP oscillator are well understood, we have not placed any emphasis on the bifurcation parameter ε . This parameter may be utilized if we wish the system to incorporate both fixed point and limit cycle dynamics. If we set $\varepsilon < 0$ oscillatory dynamics will subside and the fixed point $(x_*, 0)$ will become stable. This fact may be exploited if we require recognition to be described by both/either a stable point and/or a cycle.

4.4.1 One Dimensional Colour Recognition with Oscillatory Dynamics

Here we shall consider the simple task of applying this approach to the evolving potentials presented in figures 4.9a and 4.10a. These potentials are generated from the same stimuli considered in section 3.5.2 and correspond to Southpark and Futurama respectively. Here we consider the application of the stimuli to the foam without any restriction and where $\sigma = \sqrt{0.01}$. As this new approach to recognition retains a gradient aspect, we again demonstrate the convergence of several trajectories towards the local minima of the potential. This is illustrated by a set of black lines at the bases of figures 4.9a and 4.10a. The rate parameter in these cases is $\gamma = 10$. The limit cycles of the system can be seen to oscillate around these trajectories. Figures 4.9b and 4.10b demonstrate a set of limit cycles that characterize the recognition of a cross-section of stimuli that are encountered by the system at $t = 500$. In both examples these limit cycles quickly converge to a position in phase space that may be associated with the minima of the potential and therefore a recognisable colour learnt by the system. The colours corresponding to these limit cycle trajectories are shown in figures 4.9e and 4.10e respectively. Working from left to right the colours appear in the same order as the cycles along the x-axis. Returning to figures 4.9b and 4.10b it is clear that each cycle has a unique amplitude. This attribute may be easier to observe in figures 4.9c and 4.10c which illustrate the semi-stable states of the cycles towards the latter stages of the time domain. For each cycle the parameter $\varepsilon = 0.01$. The amplitude of the cycle is therefore only dependent on the curvature of the potential. If we were to consider a more complex scenario, we could insist that ε become an additional flexible variable of the potential. In Chapter 5 we begin to consider a few possibilities for this. The frequency of each oscillator is shown in figures 4.9d and 4.10d. Here the colour of the lines correspond to the colours of the cycles shown in figures 4.9c and 4.10c. It is clear that larger cycles are associated with higher frequencies. As the frequency of oscillation is determined by the curvature of the potential in the vicinity of a local minimum, it is clear that larger cycles relate to higher values of curvature. The frequency is therefore high when the potential well that characterizes the class is narrow. It should be noted that frequency does not necessarily indicate the size of a limit cycle's basin of attraction as the topology of the potential in close proximity to a minimum may be very different from shape of the more general potential well. We may consider that amplitude only serves as a rough estimate of class width. Comparing the range of colours recognized by the systems, as shown in 4.9e and 4.10e, suggests that both sets of stimuli are very similar. If we were to rely solely on a gradient technique then there would be no evidence to contradict this statement. In this case the limit cycle approach allows us to refute this. It can be seen from the shape of the potentials that the formation of the systems memories are indeed very different, this difference would be lost from recognition if not for the inclusion of oscillatory dynamics.

The potential shown in figures 4.9a and 4.10a possesses many local minima. We have therefore only represented a subset of the limit cycles that may exist in the phase space of the system. Selecting further stimuli to recognize we may expect to observe several more distinct cycles.

4.5 Summary and Conclusion

In this Chapter we have considered oscillatory dynamics and demonstrated how, by taking an appropriate function of the potential $V(t, x)$, we can utilize the VDP system to stimulate oscillatory recognition dynamics relevant to Janson and Marsden’s learning system. Recognition is introduced as the assignment of a stimulus to a class that has formed based on the information that the system has learnt. Perhaps we observe an unfamiliar animal, we may consider its features (compare it to the information we have stored) and identify the species. This comparison and assignment occurs autonomously. The phase space attractors of the system are controlled by the potential. The advantage of utilizing a limit cycle to represent a class is that more information may be conveyed. Let us consider another animal example. Identifying an animal as a bird is all the information that may be represented by a gradient approach. We remain unaware of how many species of bird there are. This information may be inferred by the shape of the potential, where the local minimum has a large basin we may consider the class to be broad. A fixed point attractor would suggest that all birds are the same. Presenting this information as a limit cycle we are able to appreciate the size of the class. The location of the cycle in the phase space identifies the animal, however, the frequency of oscillations provides information about the size of the class. Where the frequency is low we should be aware that the class is relatively broad, this is equivalent to identifying that this is one variety of bird, however, many more exist.

Continuing with this idea, frequency may also be utilized by the system to gauge how ‘expert’ it has become in a a specific domain. We may appreciate that an expert classifier would have lots of small classes relating to different learnt items, whilst a novice may only possess a few broad classes. It is possible that both systems may have received the same stimuli, however, due to factors such as the correlation between stimuli, or the choice of Gaussian width parameter σ , their energy potentials $V(t, x)$ may be significantly different. If whilst recognising stimuli the novice system is able to appreciate that the classes are wide, it may then take measures to focus its learning within this region. This may include simple procedures such as narrowing the Gaussian width parameter σ , a feature of the system that we previously explored in section 2.3. In this way recognition may be utilized to stimulate deeper, more focussed learning. For the ‘expert’ system it may be beneficial to recognize a new stimulus faster, this is only plausible if the system is aware that it has developed a detailed comprehension of the environment. As stated previously this concern may be associated with the frequency of a limit cycle. Recognition will proceed quicker if the rate parameter γ of equation (4.10) is increased. It may be advantageous to modulate the recognition rate in accordance with the class width. Where the system is a ‘novice’, slower recognition rates would conceivably prevent stimuli being assigned to a less appropriate class, greater time would potentially permit more suitable attractors to emerge.

It should also be stipulated that creating a model of recognition represented by limit cycles is arguably the first step towards realizing a hierarchical framework in which different recognized stimuli may be associated with each other. Identifying classes and subclasses of stimuli certainly remains outside of the scope of a simple fixed point dynamical representation, but may be achieved utilizing limit cycles and quasi-periodic orbits. At the start of this Chapter we speculated about

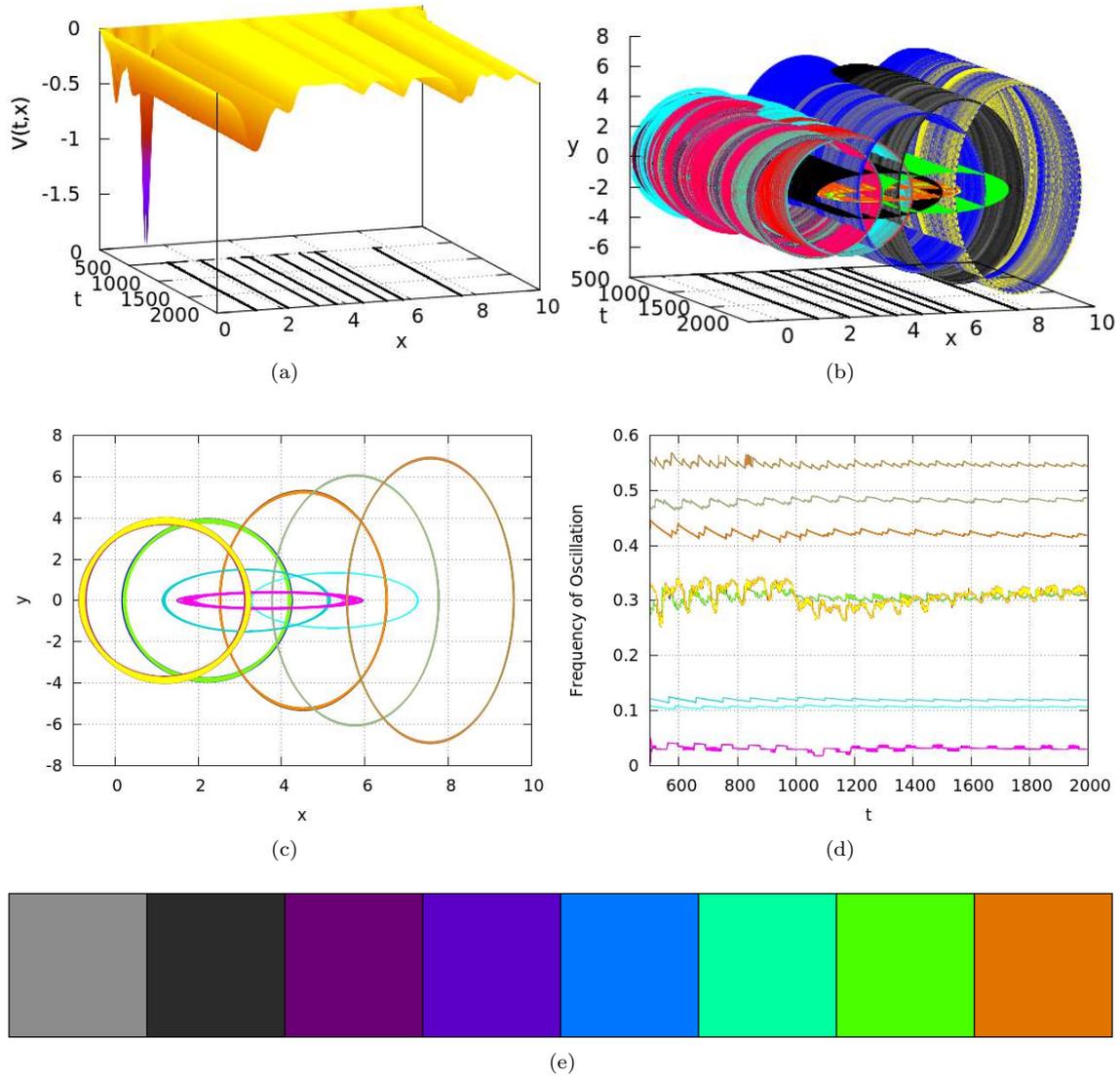


Figure 4.9: **(a)**: The evolution of the potential $V(t,x)$ for a stimuli generated by observing the colours found within an episode of the cartoon Southpark. The initial RGB triplet that identifies each colour is mapped onto a one-dimensional colour spectrum scalar x . The Gaussian width parameter in this instance is $\sigma = \sqrt{0.01}$. At the base of the plot are several gradient descent trajectories indicating the recognition of stimuli from a cross-section of initial conditions. The recognition rate parameter $\gamma = 10$, hence the rapid convergence towards each local minimum. **(b)**: Limit cycles representing the recognition of the cross-section of stimuli. The size and shape of each limit cycle adjusts with time as new stimuli shape the potential. The parameter ε that influences the oscillatory dynamics is set to 0.01. **(c)**: As the potential approaches a stable composition and recognition trajectories locate the local minima of the system, the limit cycles demonstrate a stable behaviour. Here we show this stable behaviour by overlaying trajectories for the time interval $t = 1500 \rightarrow 2000$. **(d)**: Each limit cycle may be associated with a distinct frequency, due to the time dependency of the system this will also be a function of time. The colour of each cycle in (c) corresponds to the colours of these visualizations of the frequency. **(e)**: The set of colours recognized by the limit cycle orbits.

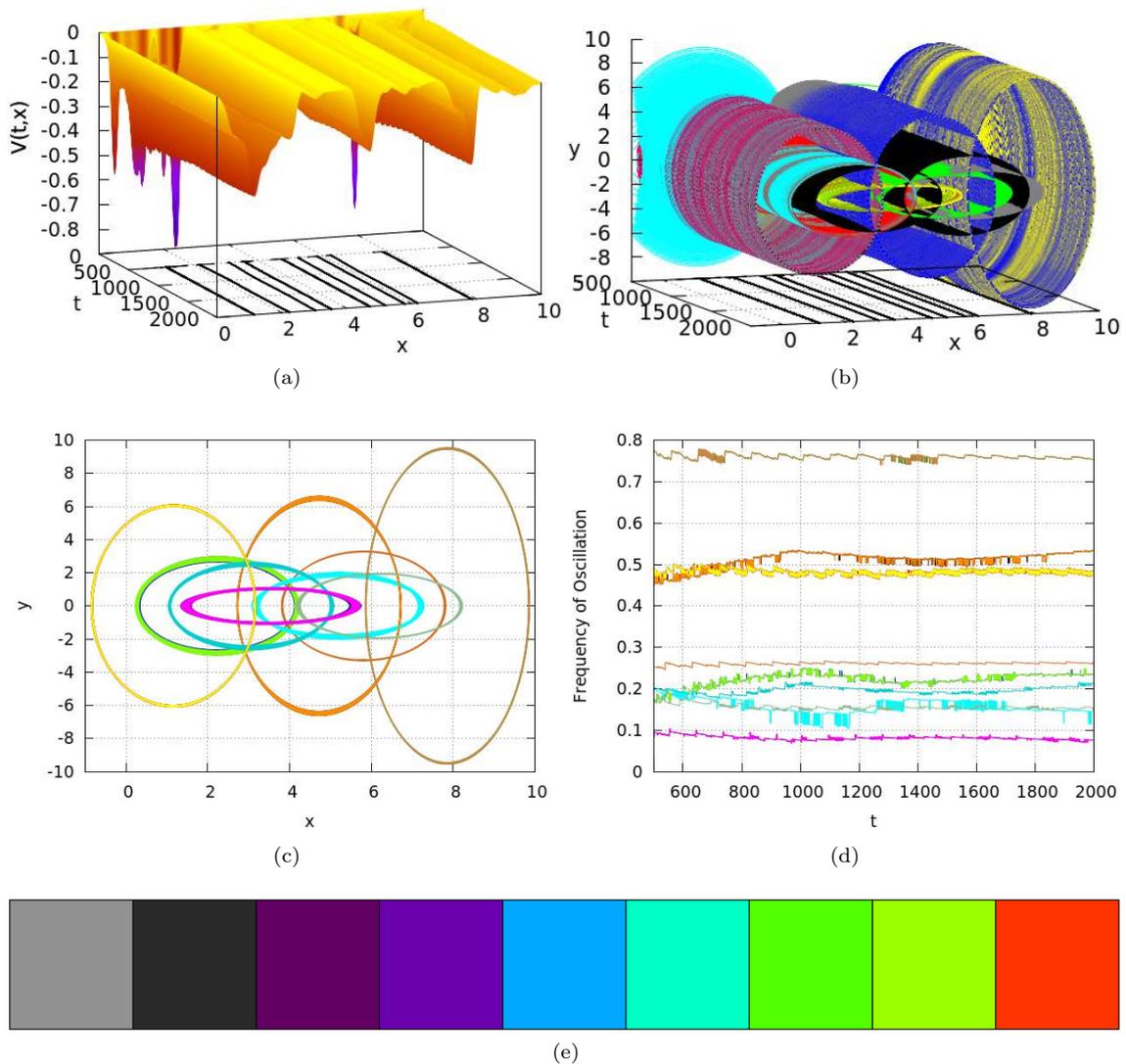


Figure 4.10: **(a)**: The evolution of the potential $V(t,x)$ for a stimuli generated by observing the colours found within an episode of the cartoon Futurama. The initial RGB triplet that identifies each colour is mapped onto a one-dimensional colour spectrum scalar x . The Gaussian width parameter in this instance is $\sigma = \sqrt{0.01}$. At the base of the plot are several gradient descent trajectories indicating the recognition of stimuli from a cross-section of initial conditions. The recognition rate parameter $\gamma = 10$, hence the rapid convergence towards each local minimum. **(b)**: Limit cycles representing the recognition of the cross-section of stimuli. The size and shape of each limit cycle adjusts with time as new stimuli shape the potential. The parameter ε that influences the oscillatory dynamics is set to 0.01. **(c)**: As the potential approaches a stable composition and recognition trajectories locate the local minima of the system, the limit cycles demonstrate a stable behaviour. Here we show this stable behaviour by overlaying trajectories for the time interval $t = 1500 \rightarrow 2000$. **(d)**: Each limit cycle may be associated with a distinct frequency, due to the time dependency of the system this will also be a function of time. The colour of each cycle in (c) corresponds to the colours of these visualization's of the frequency. **(e)**: The set of colours recognized by the limit cycle orbits.

representing sub-classes as limit cycles, whilst larger classes become associated with quasi-periodic trajectories. Where there are many layers of classes and subclasses we may hypothesise about a phase space containing many higher order tori. To achieve this vision we must consider the interaction of limit cycles representing the different sub-classes.

In this Chapter we initially considered a VDP type system with a double well potential, this was considered the most simple extension from a VDP model with a parabolic potential. This approach was expected to generate multiple limit cycles within the phase space of the system, one for each local minimum of the potential. This behaviour was predicted due to the existence of fixed points that could meet the requirement for a local Hopf bifurcation. Analysis of this simple system showed that although this Hopf criteria was met, instead of local cycles encapsulating each fixed point, a bifurcation would only lead to a single cycle that contained all the fixed points of the system. Although some interesting dynamics were presented, this method did not offer an appropriate routine to generating relevant, desirable oscillatory behaviour. As such we did not extend this approach to consider a potential with more local minima. Faced with this issue we returned to the standard VDP system. Although the potential of the learning system may, in general, contain N local minima, we can model such a profile using a Taylor expansion and represent any multi-well potential as N distinct parabolas. In order to ensure that each parabola remains in the desired orientation a further transform must be applied. This transform is the derivative of the sigmoid function, chosen to ensure smoothness as discussed in section 4.3. Using these parabolas (that are directly related to the potential $V(t, x)$) the VDP system will generate limit cycle trajectories that have a frequency dependent on the potential. By retaining gradient descent we can position each cycle at the corresponding local minimum. Recognition can therefore be described in terms of frequency as well as position. One factor that we have not considered is the amplitude of the generated cycle, this is influenced by the parameter ε and indirectly by $V(t, x)$. This attribute of the cycle may be useful to characterize more properties of the system such as the depth of the potential or the similarity between a stimulus and a fixed point. The parameter ε is important as it ensure that the dynamics of the system follow a limit cycle, however, we may not have exploited it fully within the current formulation of the model. A concern may be expressed about the application of the method presented in section 4.3 when the dimensionality of the potential is increased. It is unclear how to associate the van der Pol system with an N -dimensional potential. If x is an N -dimensional vector quantity we may consider the oscillations in a $2N$ -dimensional phase space. How trajectories will behave in this space requires further investigation.

By advancing from simple gradient descent we permit future development of the model. Later, in Chapter 5 we discuss coupling and synchronization as a method to model the interaction of stimuli with the learning system. These concepts may also help us to model recognition with regards to hierachy. Coupling is known to generate a wealth of different dynamical regimes such a quasi-periodicity or even chaos. Oscillation is paramount as a first step if we wish to demonstrate these further behaviours. Understanding the bifurcations permissible within the model is clearly of critical importance. Because our VDP recognition scheme utilizes a time dependent potential we may expect that the oscillations of the system may appear noisy due to pertubations of both frequency and position. This may be particularly eminent during the early formation of the classification profile, when each new stimuli may have a significant impact. Clear limit cycles may only become apparent after a sufficient amount of time has passed. In order to achieve a smoother transition to a stable cycle the frequency of oscillation can be intentionally increased by

a constant parameter, this is demonstrated in sub-section 4.3.1. We should note that there remains an argument for and against such an approach. This apparent noisy, transient behaviour may be utilized as a measure of certainty, a stimulus may only be considered to be adequately recognized when a stable cycle is obtained. While the cycles remain susceptible to transient changes it may be considered that the system is still learning and has not yet obtained enough relevant information to correctly recognize/optimally recognize the stimulus.

Although we have considered a method to generate oscillatory dynamics when the potential is characterized as a one-dimensional profile, it remains unclear how higher order dynamical behaviour may be instigated in a model with a potential consisting of two or more spacial dimensions. A new approach is required to enable us to robustly model the state of the system in this circumstance. Incorporating oscillatory behaviour to characterize the recognition of a stimulus for a multi-dimensional memory should allow us to create a better representation of the memory's state. Ultimately the application of non-linear principals may allow us to create a hierarchical representation of recognition. This would relate classes and sub-classes of information via dynamic properties.

The next Chapter seeks to develop our current concept of recognition so that limit cycles and higher order phase trajectories may represent classes of information when the plastic vector field of the system is N-dimensional. The van der Pol oscillator is a two-dimensional system, this infers that the trajectories of the system are restricted to the limit cycle or the fixed point. When increasing the spacial dimension of the model we should ensure that the dimensionality of the phase space is also increased. The advantage of this is that quasi-periodic and chaotic trajectories may then become a possibility. This increases the possibilities for characterising the state of the potential and may form the basis of our hierarchy. We may conceive of a representation of classes and sub-classes where each sub-class is characterized by a cycle and the wider class is inferred by a quasi-periodic orbit. The coupling of several van der Pol systems offers one method to increase the number of phase space dimensions of our model. A consideration of such methods may offer a valuable insight into how our general model of non-linear recognition may be composed.

Chapter 5

Coupled Oscillators and higher order Phase Dynamics

In Chapter 4 we discussed a method to describe the recognition of a stimulus with oscillatory dynamics. It would seem apparent that alternatives to a fixed point representation of a class are necessary if we wish to describe the characteristics of a class in greater detail and within a hierarchical structure. This concept of hierarchy is explained in section 4.1. We postulate that the formation of non-linear attractors such as limit cycles and quasi-periodic orbits should form the basic constituents of this recognition scheme. The model derived in Chapter 4 presents the most simple case of oscillatory dynamics, related to the case where the stimulus consists of a single component. This is certainly a restricted case, as, in general, a stimulus may consist of many components that relate to environmental factors. We may therefore visualize a stimulus $\eta = (\eta_1, \eta_2, \dots, \eta_N)$ as an N-dimensional vector that relates to a multitude of sensory information. If we compare this to the inputs received by a human, we may consider these inputs to relate to sound, sight, touch, taste and smell. For the model, each one of these attributes may be described by a number of different components. This notion should be clear if we recall the three-dimensional RGB construct used to define colours in Chapter 3. In such circumstances it is clear that the dimensionality of the stimuli must be imitated by the potential $V(t, \mathbf{x})$. In this chapter we explore how higher order attractors may be generated and related to the system when the knowledge represented by the model takes this multi-dimensional form.

The dimensionality of the potential $V(t, \mathbf{x})$ has important implications on the vector field of the system. Where \mathbf{x} is a N-dimensional vector we require a recognition scheme that is capable of representing a class associated with the N-dimensional space. In the previous discussion presented in Chapter 4, we explored the one-dimensional case using a van der Pol system to represent recognition with a limit cycle. In that circumstance we applied a relation of the van der Pol system to demonstrate the recognition of a stimulus on a two-dimensional phase plane. The extension of this method for the case where the stimuli is N-dimensional is not obvious. Certainly we wish to retain the ability of the system to portray limit cycles, however, it would be beneficial if we could extend the dimensionality of the phase space to permit quasi-periodic and chaotic trajectories. Different dynamic attractors may serve to characterize the recognition of stimuli when the corresponding classes demonstrate different topological properties. These properties may

relate to the size of a class or the curvature of the potential. We may also consider the depth of a potential-well or the relation between several classes. A system where the potential contains many local minima in close proximity, embedded within a larger depressed region is conceivable. Such a scenario is shown in figure 4.1, clearly the same relation may be observed for an N-dimensional profile. It would be useful in this circumstance to relate the attractors of the system to the organization of the minima constituting the potential. We have previously hypothesised about representing local minima as limit cycles, whilst wider classes of information are associated with quasi-periodic orbits. This concept may be utilized to convey a hierarchical structure relating the different classes of the system. This notion is explained in section 4.1.

Before we address such ideas we must first extend the limit cycle concept to characterize the recognition of stimuli when the potential has many dimensions. It is clear that this situation may become complex as we wish to consider both the extension of the model to an N-dimensional potential and also permit the demonstration of hierarchy. We have previously speculated that both of these goals may be approached by investigating the dynamics of coupled oscillators. As it appears necessary to first model the recognition of an N-dimensional stimulus, we shall first consider synchronization within the framework of an N-dimensional potential. A study of the synchronization dynamics of coupled oscillators appears to be an appropriate place from which to begin our investigation as it is known that sets of coupled van der Pol oscillators possess various different attractors depending on the choices of parameters [10]. These attractors include quasi-periodic orbits, fixed points and chaos. Coupling also permits the existence of limit cycles when the systems synchronize.

In this Chapter we discuss the possible methods of coupling sets of van der Pol oscillators. We then move on to a consideration of the bifurcation structure of a pair of dissipatively coupled oscillators, before applying this knowledge to a system where the potential is two-dimensional. We subsequently consider the regimes of a set of three mutually coupled oscillators and demonstrate the difficulties of such analysis. Our analysis of a three oscillator system and the implications that this has on recognition remain incomplete. Here we only begin to explore the possible dynamics of the system in relation to the concept of learning. We highlight some of the dynamics that can be expected and apply this knowledge to the stationary three-dimensional potentials of the learning model that were created via the application of colour stimuli in Chapter 3.

5.1 Coupled Oscillators

The inclusion of a coupling term describes the case where the motions of an oscillating system are perturbed by the influence of an ensemble of other interacting oscillators. In many situations this coupling may be regarded as a multi-directional driving force. Such system possess a much richer variety of dynamical attractors than uncoupled systems, these attractors include quasi-periodic cycles, limit cycles, fixed points and chaotic orbits. The behaviour of the system is generally different depending on the method of coupling. Most prevalently considered methods include dissipative and reactive coupling [126], however, more unusual non-linear methods have also been presented in the literature [124]. Any higher order differential equation may be written as a set of linearised equation as shown in section 1.4.1. Reactive coupling is associated with the difference between the variables x of interacting linearised systems, whereas dissipative coupling concerns the difference between the variables y . This is demonstrated for a pair of van der Pol oscillators in equation (5.2).

Reactively coupling a pair of van der Pol oscillators presents the possibility of quasi-periodic motions as well as permitting a pair of bi-stable limit cycles to emerge [27]. This bi-stability is observed for a small portion of the parameter space, which may be partitioned into quasi-periodic, suppression and phase locking regions.

Dissipative coupling results in different types of phase space trajectories compared to reactive coupling. The emergence of a fixed point attractor due to oscillation death is apparent, as is the omission of the bi-stable pair of limit cycles. Where a pair of dissipatively coupled van der Pol oscillators are considered, the dynamics depend on the frequency of each oscillator, the non-linearity parameter, and the coupling strength. The effect of strong dissipative coupling is considered in [146], whilst weak is discussed in [132]. Dissipative coupling permits the emergence of stable limit cycle trajectories and quasi-periodic orbits, however, it can be shown that the possibility of chaotic motions is excluded for a pair of oscillators [86].

The number of oscillators included in the ensemble also plays a significant role in the dynamics observed. A large portion of research is devoted to the study of pairs of oscillators [10, 126], however, systems consisting of larger ensembles have also been explored. We may be required to consider many interacting systems if we are to model an N-dimensional potential. We should highlight that the complexity of such analysis increases many fold as the number of oscillators increases. We should also be aware that including more units may permit a greater range of dynamical attractors. Coupled systems of many oscillators are typically introduced with coupling between nearest neighbours rather than across the whole system. Following this approach oscillators may be organised in a ring [42, 121] or chain [40, 41]. Such approaches may have implications for the hierarchical representation of recognition in one-dimension, however seem insufficient to model the interaction of each dimension of an N-dimensional potential.

Recognition of stimuli by a system composed of many related categories of information (such as the RGB example discussed in Chapter 3 or the two-dimensional simulated examples discussed in Chapter 2) may be approached by considering a system composed of many self-oscillatory parts. When the knowledge of the system is represented by an N-dimensional potential, recognition may be characterized by associating each dimension of \mathbf{x} with a van der Pol oscillator. This oscillator may be coupled to the $N - 1$ oscillators representing the other dimensions. The dynamics of the system would be described by a trajectory within a $2N$ -dimensional phase space that is completely dependent upon the architecture of the potential. The frequency of each individual oscillator will remain dependent on the local configuration of the potential in the relevant dimension. It is evident that the analysis of such a system would be strongly reliant on numerical methods as general analytic solutions to van der Pol equations can only be derived in limited cases [122]. Large systems of coupled oscillators have previously been considered [43, 102, 148], however not in the same composition as we wish to explore. The treatment of a van der Pol system with ‘all to all’ coupling and varying frequency relations has not, to our knowledge, been systematically studied.

5.2 Coupled van der Pol Oscillators

We have briefly considered some of the broad array of previous works regarding the coupling of van der Pol systems [124, 27, 146, 132, 86, 10, 42, 121, 40, 41]. Although there appears to be a plethora of methods we should now focus our attention. We wish to discern a suitable method for instigating oscillatory behaviour for an N-dimensional system. The realistic choice we are presented with is between dissipative and reactive coupling [126]. We have already proposed that

each dimension of the potential may be represented by a van der Pol oscillator that is coupled to every other dimension, however, we have not yet concluded the appropriate form of this coupling.

A pair of coupled van der Pol oscillators are described by equation (5.1). Dissipative coupling is described by the term $B_D(\dot{x}_1 - \dot{x}_2)$, where B_D is a constant referred to as the dissipative coupling strength. Reactive coupling is described by $B_R(x_1 - x_2)$ where B_R is the reactive coupling strength constant.

$$\begin{aligned}\ddot{x}_1 - \varepsilon_1(1 - x_1^2)\dot{x}_1 + \omega_1^2 x_1 + B_R(x_1 - x_2) + B_D(\dot{x}_1 - \dot{x}_2) &= 0 \\ \ddot{x}_2 - \varepsilon_2(1 - x_2^2)\dot{x}_2 + \omega_2^2 x_2 + B_R(x_2 - x_1) + B_D(\dot{x}_2 - \dot{x}_1) &= 0\end{aligned}\tag{5.1}$$

The addition of the coupling terms adds complexity to the system. The dynamics that may be observed within the four-dimensional phase space $(x_1, \dot{x}_1, x_2, \dot{x}_2)$ are more varied as a result. As a basis for applying the concept of coupling to our learning model we should explore the possible phase space attractors of system (5.1) and consider its bifurcation structure. For convenience we may rewrite equation (5.1) as four first order differential equations (equation (5.2)).

$$\begin{aligned}\dot{x}_1 &= y_1 \\ \dot{y}_1 &= \varepsilon_1(1 - x_1^2)y_1 - \omega_1^2 x_1 - B_R(x_1 - x_2) - B_D(y_1 - y_2) \\ \dot{x}_2 &= y_2 \\ \dot{y}_2 &= \varepsilon_2(1 - x_2^2)y_2 - \omega_2^2 x_2 - B_R(x_2 - x_1) - B_D(y_2 - y_1)\end{aligned}\tag{5.2}$$

It has previously been shown that a pair of coupled van der Pol oscillators demonstrate a wealth of different phase-space attractors [10]. Given suitable parameter choices the system may demonstrate limit cycle behaviour, quasi-periodicity, chaos or attraction to a fixed point. The pathways to these attractors vary depending on the type of coupling and the choice of parameter values. A full analysis of a coupled van der Pol system similar to the system described by equation (5.2) is presented in [10]. As we have a slightly different formulation, it is apparent that the parameter values at which bifurcation events may occur should be somewhat different. However, as our system can be transformed into this alternative form (see section 1.4.2) the qualitative dynamics are the same. Needless to say, the ideas compiled within [10] have motivated and informed the direction of this approach. We begin by considering the amplitude and phase of the system as a function of its various parameters. These variables may be described by truncated equations, a derivation is included in appendix A.4.

5.2.1 The Truncated Equations for a Pair of van der Pol Oscillators

The choice to consider the truncated equations corresponding to a pair of coupled van der Pol oscillators is motivated by the need to study the parameter conditions that determine limit cycle dynamics. Limit cycle oscillations are characterized by a constant phase relation between the two systems when the amplitude is non-zero. Studying the evolution of phase and amplitude will allow us to identify appropriate parameter choices that permit the model to represent different classes with different phase space attractors. Generally, the solutions of the van der Pol system must be solved by numerical methods. The truncated equations make several assumptions that allow the evolution of the system to be approached from an analytical perspective. We are able to derive the conditions for limit cycle behaviour using the truncated approach and then compare these

assertions to numerical simulations of the full equations (equation (5.2)) to validate our results.

The truncated equations describe the amplitude of oscillation for each oscillator and the phase difference between oscillators. By considering the phase of the system we can derive conditions for synchronization. Synchronization of oscillation in this case infers limit cycle type behaviour. Equation (5.3) is a differential equation pertaining to the evolution of the amplitude and phase. Note that both these variables are time dependent however we have omitted the script (t). $A_{1,2}$ refers to the amplitudes, whilst $\theta = \varphi_2 - \varphi_1$ is the phase difference between oscillators. Here $\varphi_{1,2}$ is the phase of oscillators 1 and 2 respectively.

$$\begin{aligned}\dot{A}_1 &= -\frac{\varepsilon_1}{8}(A_1^3 - 4A_1) + \frac{B_D}{2}(A_2 \cos(\theta) - A_1) + \frac{B_R}{2\omega}A_2 \sin(\theta) \\ \dot{A}_2 &= -\frac{\varepsilon_2}{8}(A_2^3 - 4A_2) + \frac{B_D}{2}(A_1 \cos(\theta) - A_2) - \frac{B_R}{2\omega}A_1 \sin(\theta) \\ \dot{\theta} &= \Delta - \frac{B_D}{2} \sin(\theta) \left(\frac{A_2}{A_1} + \frac{A_1}{A_2} \right) + \frac{B_R}{2\omega} \cos(\theta) \left(\frac{A_2}{A_1} - \frac{A_1}{A_2} \right)\end{aligned}\tag{5.3}$$

Equation (5.3) utilizes the condition that $\omega \approx \omega_1 \approx \omega_2$. We must also define the parameter Δ which describes the detuning between oscillations, this is described by equation (5.4).

$$\Delta = \frac{\omega_2^2 - \omega_1^2}{2\omega} \approx \omega_2 - \omega_1\tag{5.4}$$

At this stage we must make a choice between dissipative and reactive coupling. To persist with both terms adds considerable complexity to the system which is likely to be greatly amplified when more than two oscillators are considered. It would seem prudent to begin from the simplest possible arrangement. The coupling of a pair of van der Pol oscillators by both reactive and dissipative terms is discussed in [8]. In this study Aronson et al. demonstrates that such a system is characterized by four different periodic cycles. It is also shown in [10] that reactive coupling permits the possibility of bi-stable limit cycles. These situations are interesting, however, the ambiguity that may arise is not desirable. A further implication of reactive coupling is that the frequency for a synchronized pair of oscillators does not lie in-between the systems two natural frequencies. Rather, the frequencies of both oscillators increase as B_R increases. The frequencies finally coincide at a value larger than either natural frequency, only here does synchronization occur. In contrast, the dissipative coupling model possess a single, stable limit cycle for certain parameter values. This limit cycle is characterized by a frequency that lies between the two natural frequencies. The possibility of oscillation death for certain parameter values is an additional behaviour that is not observed in reactively coupled systems. We may be able to exploit this additional behaviour to help describe additional aspects of stimuli recognition. Having indicated our primary motives to pursue a dissipative coupling model, it is essential that we consider the parameter ranges that correspond to synchronization.

5.2.2 Dissipative Coupling

To consider dissipative coupling in the absence of reactive coupling we set $B_R = 0$ in equation (5.3). At this stage we shall also make the simplification that $\varepsilon_1 = \varepsilon_2 = \varepsilon$ in order to reduce the complexity of the analysis. These conditions lead to the truncated equations given in equation (5.5).

$$\begin{aligned}
\dot{A}_1 &= \frac{\varepsilon}{2} \left(A_1 - \frac{A_1^3}{4} \right) + \frac{B_D}{2} (A_2 \cos(\theta) - A_1) \\
\dot{A}_2 &= \frac{\varepsilon}{2} \left(A_2 - \frac{A_2^3}{4} \right) + \frac{B_D}{2} (A_1 \cos(\theta) - A_2) \\
\dot{\theta} &= \Delta - \frac{B_D}{2} \sin(\theta) \left(\frac{A_2}{A_1} + \frac{A_1}{A_2} \right)
\end{aligned} \tag{5.5}$$

This equation possess both symmetric and asymmetric solutions for the condition $\dot{\theta} = 0$. These correspond to the synchronization of oscillations. An asymmetric solution to equation (5.5) requires $A_1 \neq A_2$. A method to determine this solution is presented in [8, 7]. It is shown that this solution is always unstable for dissipative coupling. These cycles can therefore be neglected from our considerations. The only stable solution to arise is characterized by a symmetric amplitude relation.

A phase locked symmetric solution corresponds to $\dot{A} = 0$ ($A \neq 0$) and $\dot{\theta} = 0$. This means that the the two oscillators retain a constant relation relative to each other throughout time. If the system is phase locked, oscillations will take the form of limit cycles rather than quasi-periodic orbits. If we search for symmetric solutions, where $A_1 = A_2 = A$ then $\dot{\theta}$ of equation (5.5) does not depend on the amplitude. Therefore we obtain solutions $\theta_1 = \sin^{-1} \frac{\Delta}{B_D}$ and $\theta_2 = \pi - \sin^{-1} \frac{\Delta}{B_D}$, which exist as long as $|\Delta| \leq B_D$. The borderlines of the phase-locking region in the parameter space (Δ, B_D) are therefore given by $\Delta = \pm B_D$. Inserting the solutions for $\theta_{1,2}$ into equation (5.5), solutions for A can be found:

$$\begin{aligned}
\dot{A} = 0 &= \frac{\varepsilon}{2} \left(A - \frac{A^3}{4} \right) + \frac{B_D}{2} (A \cos(\theta) - A) \\
\text{hence} & \\
A \left(\frac{\varepsilon}{2} - \frac{A^2 \varepsilon}{8} - \frac{B_D}{2} \right) &= -\frac{B_D}{2} A \cos(\theta)
\end{aligned} \tag{5.6}$$

This can then be written in the form:

$$\frac{A^2 \varepsilon}{4B_D} + 1 - \frac{\varepsilon}{B_D} = \cos(\theta) = \sqrt{1 - \sin^2(\theta)} = \sqrt{1 - \frac{\Delta^2}{B_D^2}} \tag{5.7}$$

Taking the square and collecting terms we arrive at the characteristic polynomial

$$A^4 + 8A^2 \left(\frac{B_D}{\varepsilon} - 1 \right) + 16 \left(\frac{\Delta^2}{\varepsilon^2} - 2 \frac{B_D}{\varepsilon} + 1 \right) = 0 \tag{5.8}$$

This has solutions:

$$\begin{aligned}
\tilde{A}_1^2 &= 4 \left(1 - \frac{B_D}{\varepsilon} \right) + \frac{4}{\varepsilon} \sqrt{B_D^2 - \Delta^2} \\
\tilde{A}_2^2 &= 4 \left(1 - \frac{B_D}{\varepsilon} \right) - \frac{4}{\varepsilon} \sqrt{B_D^2 - \Delta^2}
\end{aligned} \tag{5.9}$$

Numerical Investigation of Two Dissipatively Coupled van der Pol Oscillators

To validate the assertions of the analytic truncated approach to this problem, the conditions for bifurcation can be obtained via numerical simulation. Here we consider the numerical study of the full equations are represented by equation (5.2). This allows us to validate the assertions we have made about parameter choices and enables us to illustrate the qualitative behaviour of the array

of attractors.

A bifurcation plot obtained from the numerical continuation software XPPAUT [44] is shown in figure 5.1. We cannot demonstrate the effect of all parameters ε , ω_1 , ω_2 and B_D in a single plot, hence we set $\omega_1 = 1$ and $\varepsilon = 0.2$ and consider the bifurcation structure on the plane (ω_2, B_D) . We have already demonstrated via the truncated equations that the synchronization region is a function of Δ (frequency detuning) and B_D . This bifurcation plot supports this inference. We observe the separation of the parameter space into five regions. A central Arnold tongue corresponds to 1:1 synchronization. This tongue is a ‘v’ shaped region which fills more of the parameter space as B_D increases. Such tongues are indicative of most systems capable of synchronization, in addition to this 1:1 region, many systems may also contain tongues that identify other n:m phase relation. The width of the tongue increasing with coupling strength. Outside of this tongue exists two quasi-periodic regions for low coupling strength and two oscillation death regions for higher coupling strengths. The parameter ε has significant implications for the systems dynamics, the line $B_D = \varepsilon$ is seen to separate the quasi-periodic region of the parameter space from the oscillation death region. Oscillation death is characterized by the fixed point of the system becoming stable. In addition to demonstrating the bifurcation structure of this system, we also present an array of trajectories corresponding to the various regions of the parameter space. These trajectories, shown in figure 5.2, should serve as evidence to the aforementioned bifurcation structure. Quasi-periodic orbits are shown in figures 5.2a and 5.2b. These are illustrated by a torus in the phase space of the model. A torus may be characterized by two or more frequencies, these are associated with the winding of trajectories in the different dimensions of the phase space. Figures 5.2c and 5.2d demonstrate the evolution of trajectories for parameter values within the oscillation death region. Here trajectories approach the fixed point of the system $(0,0,0,0)$. Finally, we have highlighted the dynamics of the synchronization region in figures 5.2e and 5.2f.

Having briefly discussed the dynamics of a pair of dissipatively coupled van der Pol oscillators, we shall now demonstrate how this knowledge can be applied to our learning system. Clearly, as we have only considered a system of two coupled oscillators, we may only extend our approach to a two-dimensional potential. Once we have shown that this routine is effective for a two-dimensional representation of learning, we may consider the dynamics of three or more coupled van der Pol oscillators and extend the dimensionality of our system accordingly.

5.3 Limit cycles for an two-dimensional Potential

Our previous consideration of limit cycles characterising the recognition of stimuli presented to our learning system was restricted to the simple case where the potential, containing the knowledge of the system, was a one-dimensional profile. It is apparent that this is a very limited case. A more desirable result would be to generate oscillatory dynamics where the potential is N-dimensional.

The extension of the approach detailed in Chapter 4 to elicit higher order phase behaviour in a system of more than one spacial dimension may be attempted via the application of coupling. We propose that each dimension of the system is treated as an individual oscillator in accordance with the methods of Chapter 4. However, each oscillator representing a dimension of the potential should interact with the ensemble of other oscillators via a dissipative coupling term. The parameters of this coupled system should be chosen so that the system possesses the ability to synchronize across its spacial dimensions. As such, any local minimum of the potential may be represented via

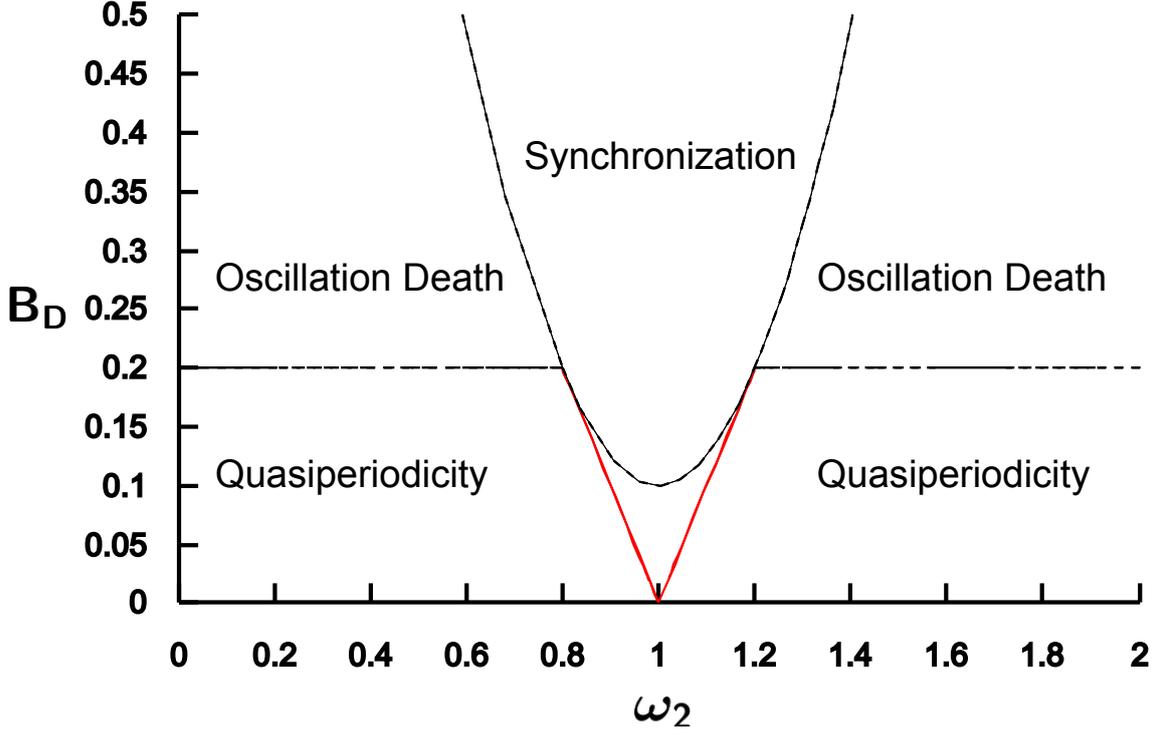


Figure 5.1: Bifurcation plot on the plane (ω_2, B_D) showing the parameter values at which transitions in behaviour occur for the equation (5.2). The parameter $\omega_1 = 1$ and $\varepsilon = 0.2$. An Andronov-Hopf bifurcation is shown by the dotted line separating the oscillation death regime from the synchronization zone. The red lines indicate a saddle node bifurcation, separating the quasi-periodic and synchronized areas. This plot was generating using the numerical continuation software XPPAUT [44].

a unique phase space trajectory. For a two-dimensional stimulus the recognition of a stimulus may be described by:

$$\begin{aligned}
\frac{dx_{*1}}{dt} &= -\gamma \frac{\partial V(x_1, x_2)}{\partial x_1} \\
\frac{dx_1}{dt} &= y_1 \\
\frac{dy_1}{dt} &= \varepsilon \left(1 - (x_1 - x_{*1})^2\right) y_1 - \alpha \frac{\partial U(V)}{\partial x_1} - B_D (y_1 - y_2) \\
\frac{dx_{*2}}{dt} &= -\gamma \frac{\partial V(x_1, x_2)}{\partial x_2} \\
\frac{dx_2}{dt} &= y_2 \\
\frac{dy_2}{dt} &= \varepsilon \left(1 - (x_2 - x_{*2})^2\right) y_2 - \alpha \frac{\partial U(V)}{\partial x_2} - B_D (y_2 - y_1)
\end{aligned} \tag{5.10}$$

Here the function $U(V)$ is described by equation (4.7). This ensures that the parabolas characterising the VDP system's potential remain in the required orientation, whilst still capturing the local topology of the potential $V(t, \mathbf{x})$. We again rely on a gradient term to identify the point x_{*1}, x_{*2} that converges to the local minima of $V(t, x_1, x_2)$. The system is composed of two VDP oscillators, considering them in isolation the fixed points are $(x_{*1}, 0)$ and $(x_{*2}, 0)$ respectively. When the system is coupled, the dynamics of the system should be viewed in a four-dimensional space

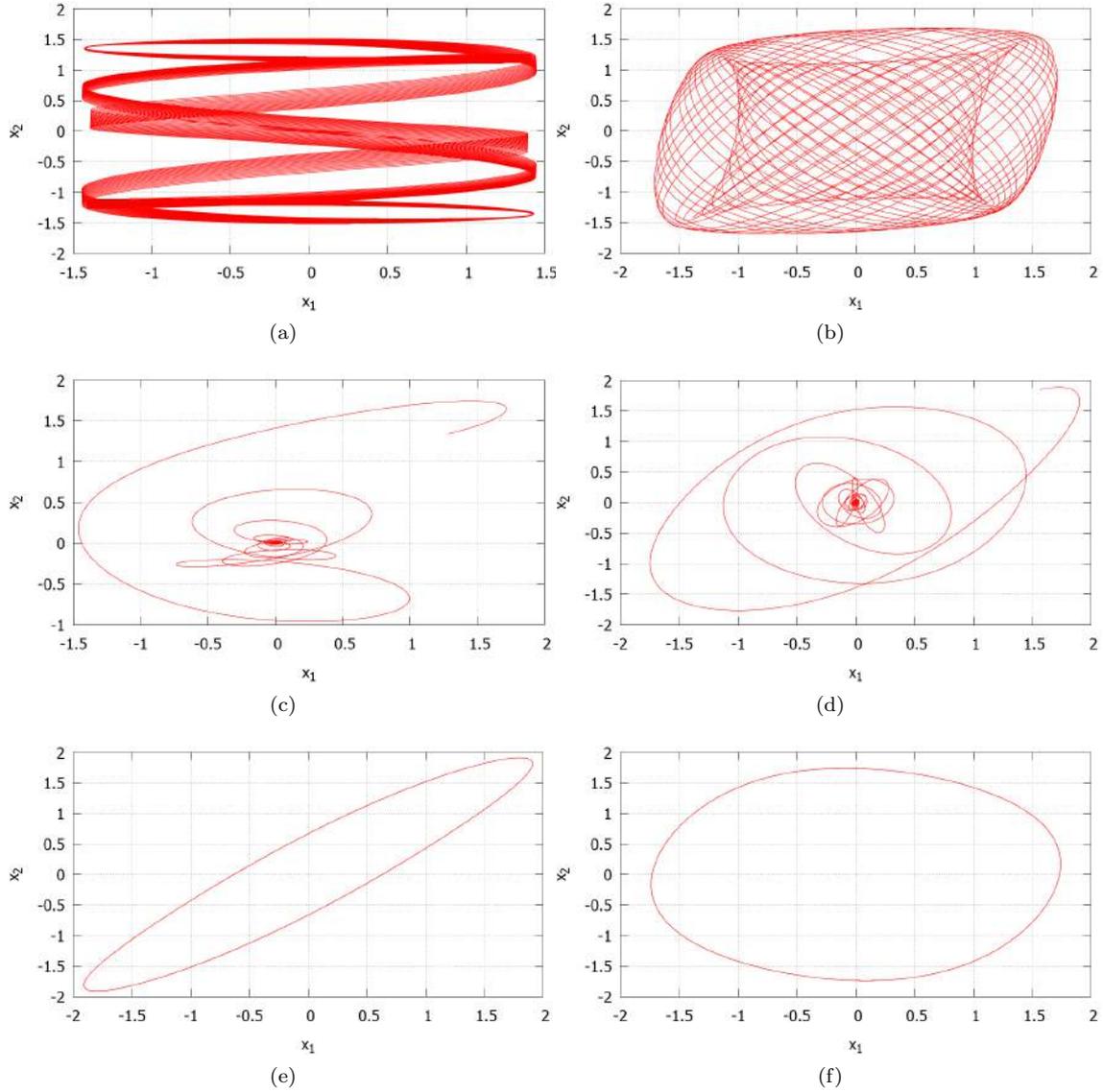


Figure 5.2: Projections of the phase space trajectories for a pair of coupled van der Pol oscillators described by equation (5.2) where $B_R = 0$. Parameters are chosen to demonstrate the various behaviours indicated by the bifurcation plot shown in figure 5.1. Note that in each figure $\varepsilon = 0.2$. A pair of coupled van der Pol oscillators has a four-dimensional phase space, hence we cannot visualize the entire space, we can however observe the behaviour in a restricted two-dimensional space. For the synchronized regions only the stable behaviour is shown. For the oscillation death region we have also included the behaviour of the system as the fixed point at $(0,0,0,0)$ is approached. **(a):-Quasiperiodic trajectory:** $\omega_1 = 1, \omega_2 = 0.2, B_D = 0.1$, **(b):-Quasiperiodic trajectory:** $\omega_1 = 1, \omega_2 = 1.2, B_D = 0.1$, **(c):-Oscillation death:** $\omega_1 = 1, \omega_2 = 0.4, B_D = 0.3$, **(d):-Oscillation death:** $\omega_1 = 1, \omega_2 = 1.4, B_D = 0.3$, **(e):-Synchronized (Limit cycle):** $\omega_1 = 1, \omega_2 = 1.1, B_D = 0.3$, **(f):-Synchronized (Limit cycle):** $\omega_1 = 1, \omega_2 = 1.05, B_D = 0.05$.

(x_1, y_1, x_2, y_2) . The point $(x_{*1}, 0, x_{*2}, 0)$ is then the fixed point of the coupled VDP system. The similarity between $\alpha \frac{\partial U(V)}{\partial x_1}$ and $\alpha \frac{\partial U(V)}{\partial x_2}$ and the values of ε and B_D will determine the dynamics of the system.

Let us consider how recognition trajectories may behave for the potential shown in figure 5.3a. This potential has been generated utilizing simulated stimuli and so provides a relevant medium on which to test proposed ideas. The stimuli are again generated using the method discussed in section 2.2.1 and are characterized by equation (2.8). The functions $\frac{dF_1}{d\eta}$ and $\frac{dF_2}{d\eta}$ that prescribe the distribution of the stimuli are:

$$\begin{aligned} \frac{dF_1}{d\eta} &= 0.06\eta^3 - 0.9\eta - 0.3 \\ \frac{dF_2}{d\eta} &= \frac{-4(\eta + 8)}{\left((\eta + 8)^2 + 0.75\right)^2} + \frac{-4(\eta + 5)}{\left((\eta + 5)^2 + 1\right)^2} + \frac{-4(\eta - 0.5)}{\left((\eta - 0.5)^2 + 4\right)^2} + \frac{-4(\eta - 4)}{\left((\eta - 4)^2 + 0.75\right)^2} \\ &\quad + \frac{-4(\eta - 8)}{\left((\eta - 8)^2 + 0.5\right)^2} + 2\eta e^{-3} \end{aligned} \quad (5.11)$$

In order to demonstrate the relative locations of the local minima and the surrounding contours of the potential we demonstrate the evolution of a cross-section of initial conditions via gradient descent in figure 5.3b. As an initial test we may consider the evolution of oscillatory trajectories when this potential remains stationary in time. Towards this aim we may allow trajectories to evolve from a mesh of initial conditions on the potential. It is apparent that the choice of coupling strength B_D and parameter ε will have important implications on the dynamics of the system. The synchronization tongue is partly governed by the difference between natural frequencies, as these are determined by the shape of the potential they are not directly controllable. We may however use the parameter α , introduced in section 4.3.1 to scale frequencies onto a more appropriate range. By scaling α we can ensure that the natural frequencies of the system are in close proximity to a synchronization tongue. This presents the opportunity that the system may characterize classes with various types of attractor. Figures 5.4c,d demonstrate the case where each local minimum of the potential shown in figure 5.3 is characterized by synchronous dynamics. This is the case when $\varepsilon = 0.6$, $\gamma = 50$, $\alpha = 1 \times 10^5$ and $B_D = 0.4$. For the sake of comparison figure 5.4a,b demonstrates the same initial conditions, with the same parameters, except $B_D = 0$. This highlights the effect of coupling. The figures 5.4a,c illustrate the evolution of projections of the trajectories as a function of time. In contrast figures 5.4b,d highlight the stable attractors that exist by demonstrating the projections of the systems attractors on the plane (x_1, x_2) .

Figure 5.1 shows that limit cycle behaviour is only guaranteed for certain regions of the parameter space. Outside of the synchronization tongue we may observe either oscillation death or quasi-periodic orbits. Which behaviour is observed is determined by the relationship between B_D and ε . We may utilize this information to demonstrate how limit cycles and either quasi-periodic orbits or fixed point attractors may be used to characterize the local minima of the potential. The behaviour observed will be dependent on the relation between $\omega_1 = \alpha \frac{\partial U(V)}{\partial x_1}$ and $\omega_2 = \alpha \frac{\partial U(V)}{\partial x_2}$. Where the basin surrounding the local minima is characterized by sufficiently different curvatures in each of its dimensions, the natural frequencies of the oscillators will be sufficiently different and synchronization will not occur. In figures 5.5a,b we have set $B_D > \varepsilon$. Here we observe that some minima are characterized by limit cycles when $|\omega_1 - \omega_2|$ is sufficiently small, and other minima

are characterized by quasi-periodic, two frequency cycles. The alternative case where $B_D < \varepsilon$ is shown in figures 5.5c,d. Here trajectories either converge to the fixed point via oscillation death or synchronize as demonstrated by stable limit cycles. In both of these cases we retain the parameter values $\gamma = 50$ and $\alpha = 1 \times 10^5$. To permit quasi-periodic solutions the parameter values $\varepsilon = 0.6$ and $B_D = 0.15$ are set. Conversely, oscillation death is realized when $\varepsilon = 0.1$ and $B_D = 0.15$. It should also be noted that when $\omega_1 = \omega_2$ the limit cycle lies on the diagonal of the phase space. As the frequencies become more distinct the observed trajectories become skewed. Figure 5.5e indicates the frequency relations for the ensemble of oscillators. This realization may be used to underpin the reasoning behind the dynamics that are observed.

Although constraints may be applied via the adjustment of α to ensure dynamics remain synchronous, it is more interesting to consider a system with various conceivable phase space attractors. Each isolated oscillator is related to a dimension of the potential, which in turn controls the frequency. The curvature of the potential in the vicinity of the minima therefore plays an important role in the dynamics of the system. For dissipatively coupled oscillators the natural frequencies of the systems are perturbed so that the synchronized system oscillates with a frequency that lies between the two characteristic frequencies [10]. We have demonstrated how alternative dynamics may be permitted outside of the synchronization tongue, this however requires a choice between oscillation death and quasi-periodicity. As a further development we may wish to control either B_D or ε through a function of the potential so that all three dynamical regimes may be appreciated by this model.

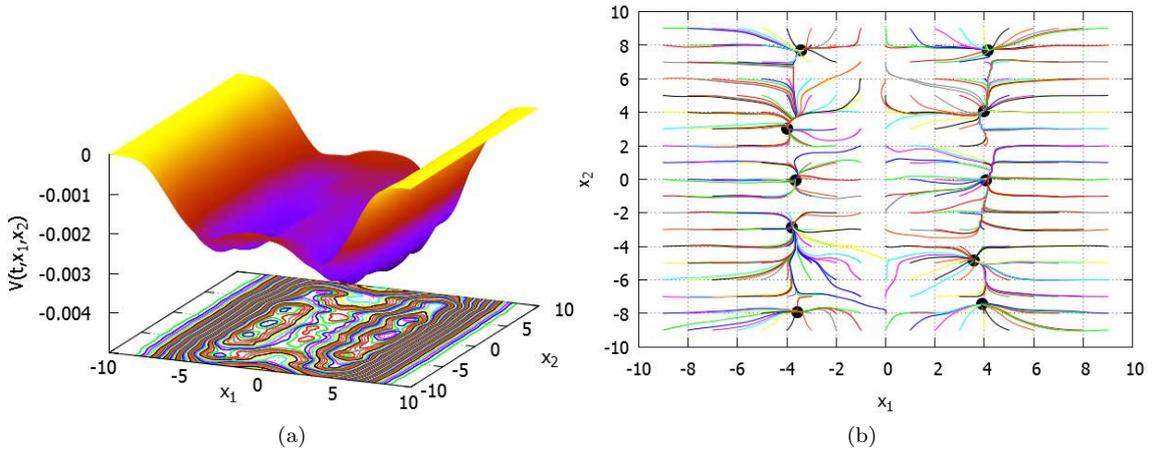


Figure 5.3: **(a)**: A potential $V(t, \mathbf{x})$ created by the application of 1×10^7 two-dimensional stimuli $\eta = \eta_1, \eta_2$. The stimuli are simulated with a prescribed PDD via equation (2.8). The distribution of stimuli are described by equation (5.11). The intensity of noise D used to simulate the stimuli is 5 and the Gaussian shaping parameter $\sigma = \sqrt{0.01}$. On the plane x_1, x_2 a contour mapping of the potential is included. **(b)**: Gradient descent for an ensemble of evenly interspersed initial conditions towards the local minima (marked as black circles) of the potential. This may be considered to intimate the general size and shape of each basin. The gradient descent rate parameter $\gamma = 5$.

5.3.1 Realizing Limit Cycles, Quasi-periodic orbits and Oscillation Death

If we wish to permit both quasi-periodicity and oscillation death in addition to synchronization then we must instigate some variation of either the parameter ε or B_D . It is useful to represent different classes of the system with these different attractors as the nature of the attractor can

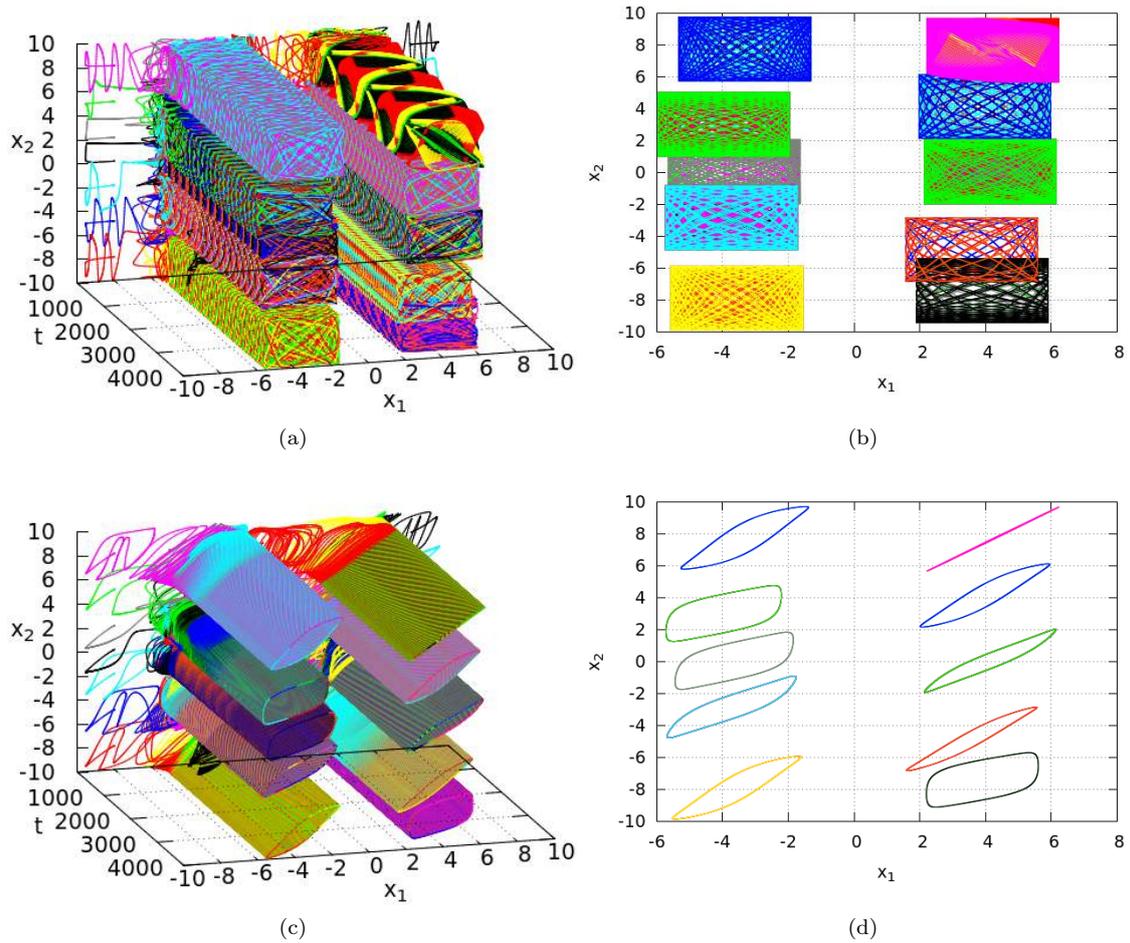


Figure 5.4: The behaviour of the system defined by equation (5.10) for the stationary potential shown in figure 5.3a. In each figure $\varepsilon = 0.6$, $\gamma = 50$ and $\alpha = 1 \times 10^5$. **(a)** and **(c)** illustrate the time dependent oscillations of the system for a sample of 21 stimuli with initial conditions from across the range of x_1, x_2 . **(b)** and **(d)** demonstrate the same trajectories but only after x_{*1}, x_{*2} has become stationary. **(a),(b)** reflect the dynamics of the uncoupled system ($B_D = 0$). For **(c)(d)** $B_D = 0.4$, as a result the oscillators synchronize and limit cycles are observed in the phase space x_1, x_2 .

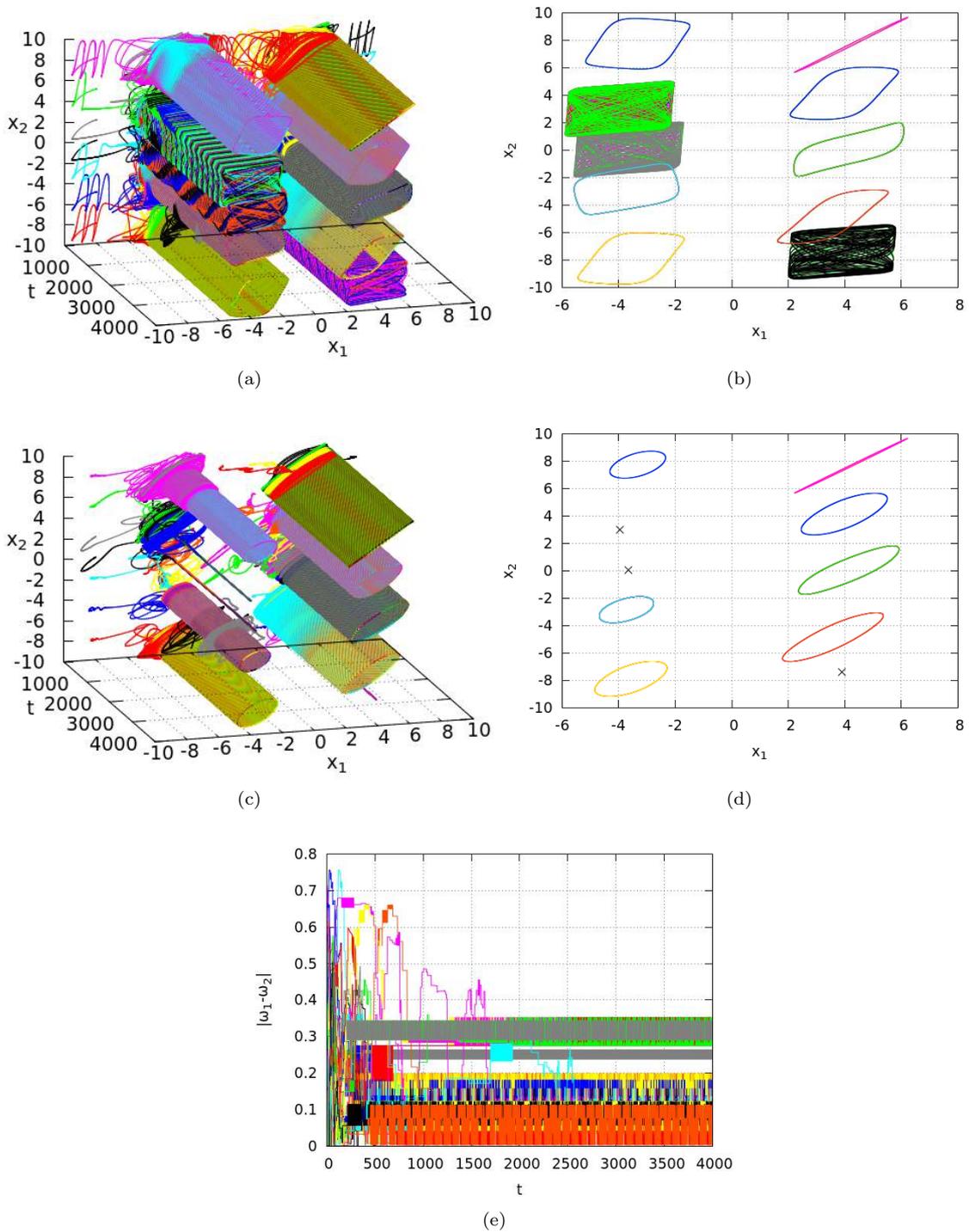


Figure 5.5: (a) and (c) illustrate the time dependent oscillations of the system for a sample of 21 stimuli with initial conditions from across the range of x_1, x_2 . (b) and (d) demonstrate the stable attractors of the system. (a),(b) reflect the dynamics of system when $\varepsilon = 0.6$ and $B_D = 0.15$, hence some trajectories as quasi-periodic. In (c),(d) $\varepsilon = 0.1$ and $B_D = 0.15$, this corresponds to the oscillation death region, some trajectories therefore converge to the fixed point of the system. (e): The difference between $\omega_1 = \alpha \frac{\partial U(V)}{\partial x_1}$ and $\omega_2 = \alpha \frac{\partial U(V)}{\partial x_2}$ as the trajectories of the system converge towards the local minima of the potential $V(t, x_1, x_2)$. The difference between these two frequencies variables and their relation to B_D determines the dynamics of the system (see figure 5.1). In each example $\gamma = 50$ and $\alpha = 1 \times 10^5$.

be used to convey additional information about the topology of the class and may even be able to demonstrate the relationships between different classes. It is important that we relate the parameters controlling the attractors of the oscillatory system to the characteristics of $V(t, \mathbf{x})$, which describes the distribution of stimuli and hence the classes of information known to the model. It would seem illogical to change the coupling strength B_D as each oscillator represents a different dimension of the same minimum. A logical approach may be to control ε . In keeping with our approach to the frequency, we should make ε dependent on some aspect of the potential. It may seem appropriate that we define ε by a function of the depth of $V(t, \mathbf{x})$. A function that may be appropriate is given in equation (5.12). This function will ensure that ε tends to a small constant for a deep potential well whilst increasing in size for larger values of $V(t, \mathbf{x})$.

$$\varepsilon = A \log \left(1 + \exp^{-B(-CV(\mathbf{x}=\eta))+D} + E \right) \quad (5.12)$$

The parameters of this function may require refinement depending on the choices made concerning the ranges of x_1 , x_2 and σ , however, for this initial consideration the parameters $A = 0.04$, $B = 0.35$, $C = 1000$, $D = 9.5$ and $E = 0.5$ are shown to scale the function onto an appropriate range to effectively control ε . This function is plotted in figure 5.6. This choice of this function is motivated by the distribution of $V(t, \mathbf{x})$, which ranges between 0 and some negative value. The function 5.12 asymptotically tends to a positive constant as $V(\mathbf{x}, t) \rightarrow -\infty$.

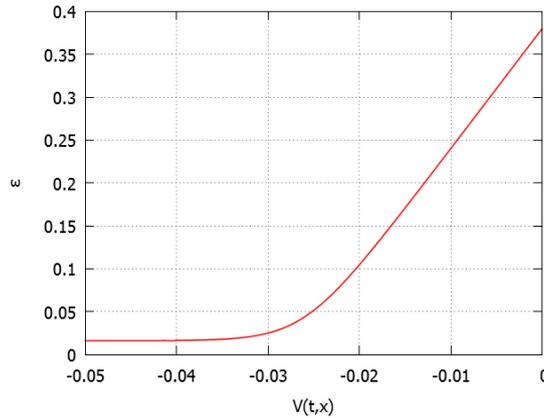


Figure 5.6: Function that adjusts ε as a function of $V(t, \mathbf{x})$. This function is described by equation (5.12). Here parameters are set such that $A = 0.04$, $B = 0.35$, $C = 1000$, $D = 9.5$ and $E = 0.5$.

The dynamics of recognition are directly derived from the topology of the potential. Here we relate a deeper potential to a larger value of ε . If the basin of the minimum has a similar curvature in both directions then the class will be characterized by a limit cycle, however, where the curvature is sufficiently different a deep well will be characterized by oscillation death. Similarly, a shallow local minimum may also be represented by a limit cycle when the curvatures are similar, the shape of this limit cycle will be distinguishable from a limit cycle corresponding to a deep minimum as ε will be small. In the event that the curvatures are sufficiently different, and the potential is shallow, the recognition of a stimulus will be demonstrated with a quasi-periodic motion. Most important to emphasise is the uniqueness of each dynamic trajectory. More distinctions may be drawn between recognized stimuli based on the knowledge contained within the system. The same cannot be said for a gradient approach which merely locates a local minimum and is incapable of providing any additional information about the classes of information contained.

To demonstrate the addition of ε scaling to the model, it is prudent to consider a potential containing minima characterized by significantly different curvatures and depths. The systems trajectories should only be dependent on the potential, in turn, this is dependent on the distribution of stimuli. To highlight each possible regime we may consider the trajectories associated with the potential shown in figure 5.7a. The change to consider this new potential is driven by the requirements that we have outlined for the topology. If we are to observe limit cycles, quasi-periodic orbits and oscillation death then certain topological constraints must be met. To simplify this demonstration we consider $V(t, x_1, x_2)$ to be stationary.

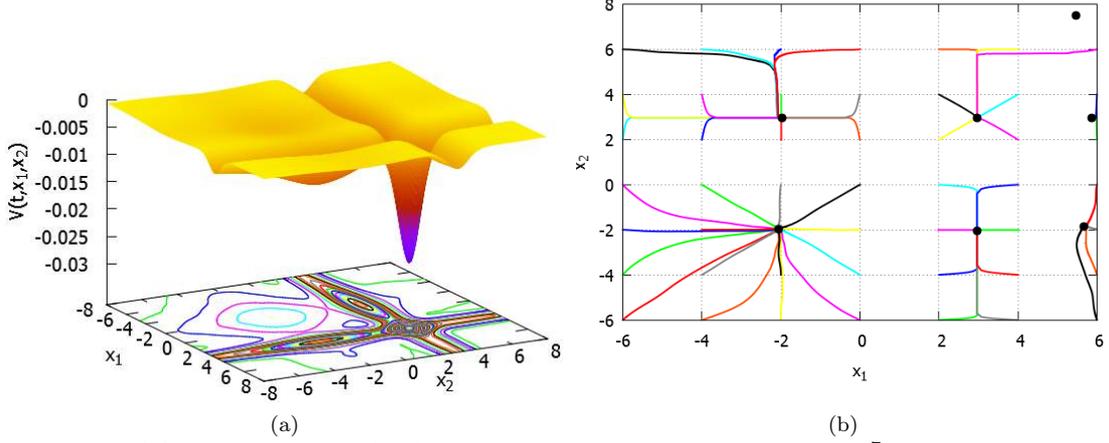


Figure 5.7: **(a)**: A potential $V(t, \mathbf{x})$ created via the application of 1×10^7 stimuli generated by equation (2.8) with distributions described by equation (5.13). **(b)**: Fixing the potential we may consider the basins of attraction of the potential via an appreciation of gradient descent from an ensemble of initial conditions.

The potential shown in figure 5.7a is created using the method introduced in section 2.2.1 with a Gaussian noise intensity $D = 25$. The function $\frac{dF_1}{d\eta} = \frac{dF_2}{d\eta} = \frac{dF}{d\eta}$ that prescribes the distribution of the stimuli is described by the function:

$$\frac{dF}{d\eta} = 0.01\eta^4 + 500e^{-1(\eta-3)^2} + 200e^{-0.1(\eta+2)^2} \quad (5.13)$$

The potential was shaped by Gaussians with a width parameter $\sigma = \sqrt{0.1}$. It takes 1×10^7 iterations of the stimuli to form this potential. The local minimas of the potential are marked in figure 5.7b along with a cross-section of initial conditions. Gradient descent is utilized here to emphasise the various basins of attraction for each attractor. Clearly the behaviour of a trajectory, associated with a basin, is dependent on the frequency $\omega_{1,2}$, the non-linearity parameter ε and coupling strength B_D . In this example we set $B_D = 0.3$. As is clear from figure 5.8a we have values of ε that are both above and below the line $\varepsilon = 0.3$. As shown in figure 5.8b we also have values of $|\omega_1 - \omega_2|$ either side of the line defining synchronization. These relations explain the dynamic regimes observed in figure 5.8c,d. Figure 5.8c demonstrates the time dependent evolution of the trajectories whilst figure 5.8d indicates the stationary nature of the attractors for the time interval $t = 1500 \rightarrow 2000$. It is shown that two of the minima are characterized by oscillation death, here $B_D > \varepsilon$ and $|\omega_1 - \omega_2| > B_D$, a further three are represented by limit cycles ($|\omega_1 - \omega_2| < B_D$) and the final local minimum is described by a quasi-periodic trajectory ($|\omega_1 - \omega_2| > B_D$, $B_D < \varepsilon$). This same approach may be applied for any prescribed stimuli. We should be aware however that we can only guarantee a

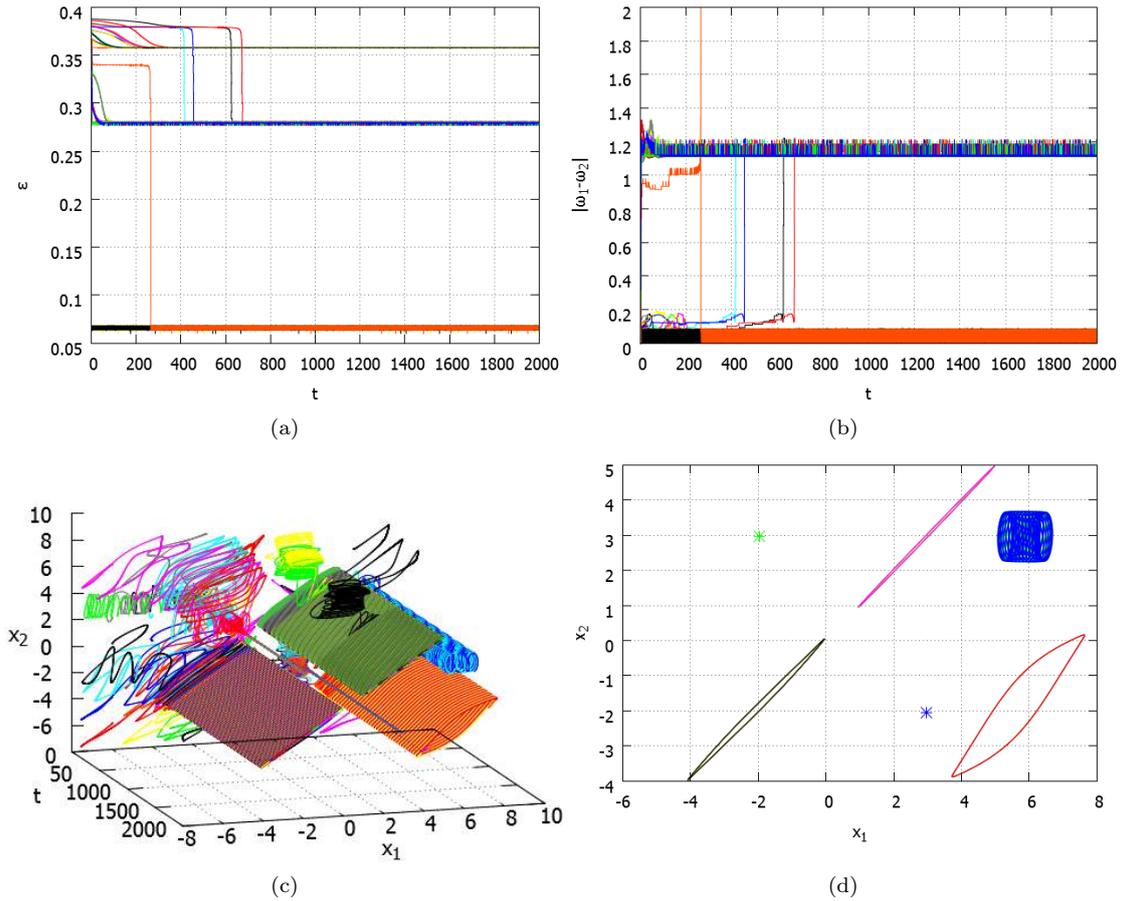


Figure 5.8: **(a)**: The parameter ε is introduced as a function of the potential. **(b)**: The evolution of the frequency relation $|\omega_1 - \omega_2|$ as an ensemble of initial conditions are allowed to evolve in conjunction with the shape of the potential. In this instance the frequency scaling parameter $\alpha = 10000$. **(c)**: The evolution of an ensemble of initial condition towards associated phase space attractors. Note that these trajectories are controlled by the parameters ε , ω_1 and ω_2 outlined in **(a)**,**(b)**, these are in turn dependent on the potential shown in 5.8a. In this illustration the coupling strength parameter $B_D = 0.3$. **(d)**: The phase space trajectories shown in **(c)** demonstrated on the plane (x_1, x_2) for the time interval $t = 1500 \rightarrow 2000$.

range of behaviours by manipulating the parameters of the system. During unsupervised learning the structure of the system's potential remains unknown and so appropriate parameter choices cannot always be inferred. The functions that we have assigned to manipulate the parameters of the system have been chosen to emphasise the various dynamic attractors that are possible. We also selected a stimuli distribution that would result in local minima with dramatically different depths and curvatures. Where the characteristics of the minima fall within a narrow range we may expect all classes to be identified by similar attractors.

We have demonstrated how ε may be related to the depth of a potential well and explained the implications that this may have for the recognition of a stimulus. Manipulating the parameter ε allows us to illustrate the features of a class as described by an attractor. Our notion of hierarchy also requires the existence of a range of different attractors, it is important to demonstrate a method to move between different types of attractor, relative to the topology of classes. We should keep in mind that we may also employ alternative strategies. A second line of investigation is that ε may be used to convey the difference between a stimulus and the class that it is assigned to. This would involve taking a measure between a stimulus's initial condition in the space \mathbf{x} and its position as it evolves in time. Let us explore this alternative concept.

5.3.2 Adjusting ε as a function of distance traversed

The local minima of the potential, corresponding to the fixed points of the system, represent the most prevalently observed stimuli. Considering our previous approaches to the problem of recognition, we have drawn no distinction between a stimulus that falls directly at the local minima or at the extreme of the attractors basin. Although it is true that we require both to be recognized as part of the same class, it may also be advantageous that we identify that two such stimuli are not exactly the same. This may be considered as a measure of certainty in the recognition, or may even be identified as a basic hierarchical representation. Separating stimuli that associate with the extremes of the class from those that compose the central regions may be equivalent to identifying different colours as 'grey' and 'grey-ish'. If we insist that ε is a continuous function of position, then we may consider the whole basin to be characterized by a single colour but with continuously changing levels of 'ish'.

This concept may be realized by a Gaussian function such as the one provided in equation (5.14). This takes as its argument a measure of the difference between the initial stimulus position η and the location $\mathbf{x}(t)$, as the trajectory evolves with time. Defining ε via this smooth continuous function ensures that ε remains within a defined range. We are able to control the sensitivity of the system by defining the variance of the Gaussian or by changing our parameter B_D .

$$\varepsilon = \frac{1}{A} \exp^{-B * z^2} + C \quad (5.14)$$

Here A controls the height of the Gaussian, B controls the width and C defines the minimum of the function. It is clearly important that ε tends to some positive constant C rather than zero. The variable z is our measure of distance. For a two-dimensional potential this may be defined as $z = \sqrt{(\eta_1 - x_1)^2 + (\eta_2 - x_2)^2}$. We may again utilize the stationary potential shown in figure 5.3 as a profile on which to illustrate this proposed approach. Selecting the parameter values $A = 0.6$, $B = 0.5$, $C = 0.1$, $\gamma = 50$ and $B_D = 1 \times 10^5$ this approach is demonstrated in figure 5.9.

It is evident from figure 5.9b that limit cycles, quasi-periodic orbits and fixed points can all be

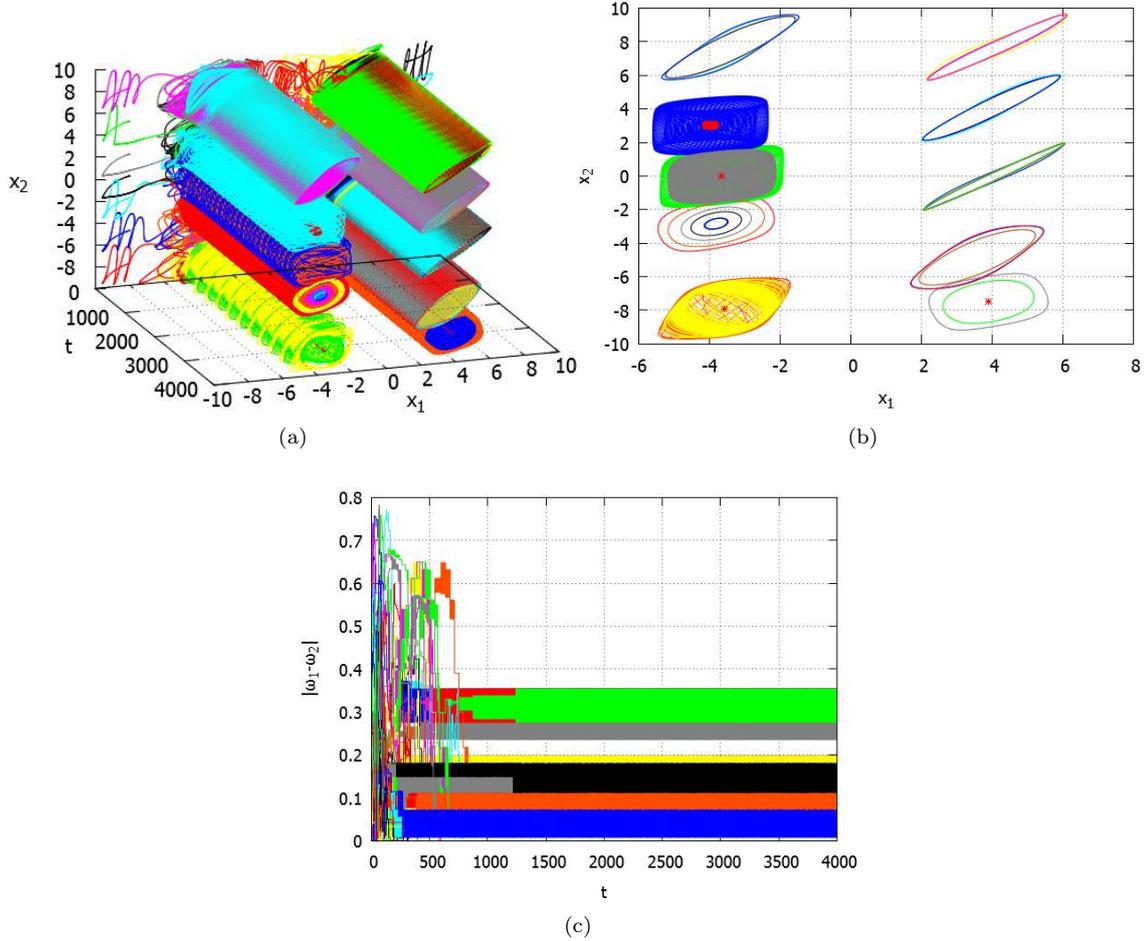


Figure 5.9: Illustration of the effect that ε has on the trajectories of the system when it is treated as a function of the difference between the minima of the potential and the stimulus. The potential utilized for this example is shown in figure 5.3. The parameters $\gamma = 50$, $\alpha = 1 \times 10^5$ and $B_D = 0.15$. **(a)**: Demonstration of recognition as described by trajectories in the phase space x_1, x_2 against time. The initial conditions for each trajectory influence the variable ε . **(b)**: The phase space behaviour of the system for the time interval $t = 3500 \rightarrow 4000$. The space contains quasi-periodic orbits, limit cycles and stationary points. **(c)**: The frequency relations for the oscillating variables shown in **(b)**. Note that the relation $|\omega_1 - \omega_2| = B_D$ separates the synchronization region within the parameter space.

realized within this system if parameter values are adjusted accordingly. As the synchronization region of the parameter space is defined by the line $\omega_1 - \omega_2 = \pm B_D$, the parameter ε does not influence whether the behaviour is a limit cycle or otherwise. The emergence of limit cycles as well as alternative dynamics are expected from an appreciation of the frequency relations shown in figure 5.9c. It is apparent from figure 5.9b that ε will effect the shape of any limit cycle. We can therefore distinguish between a cycle that is the result of a stimulus close to a local minimum and a cycle that is instigated by initial conditions that are a greater distance away. When the curvature of the potential $V(t, \mathbf{x})$ is sufficiently different with respect to the two characteristic directions of the system x_1 and x_2 , the trajectory of the system will converge to either a fixed point or a quasi-periodic orbit. For trajectories instigated further away from a minimum, ε is smaller, this condition is intended to correspond to the oscillation death region. Consultation of figures 5.9a,b highlight that any minimum represented as a quasi-periodic cycle may also elicit fixed point dynamics depending on the initial condition of the stimulus.

Both of the discussed methods of permitting different types of dynamical behaviour in addition to limit cycles rely on defining ε as a function of the systems variables. In the first instance we consider how ε can become a function of the potential's depth, hence making the dynamics completely dependent on the topology of the learning profile. Alternatively, we may be motivated to consider the difference between a stimulus and the minimum it is assigned to. Because the condition for a limit cycle is only dependent on the frequency and coupling strength, these additional considerations do not influence the exhibition of limit cycles. Cycles may appear for shallow or deep wells, or for stimuli that are close/far away from a local minimum. Dynamic behaviour is a function of the similarity between the potential's curvatures in the different dimensions of the profile. The value of ε does however impact upon the shape of the cycle and so it remains possible to distinguish between different characteristics of the potential.

Applying the methodology of subsection 5.3.1 various parameters of the class may be identified. Firstly, we may consider whether the curvature in each dimension is roughly symmetric, if this is the case the recognition trajectory will form a limit cycle. The frequency of this cycle will lie between the frequencies that characterize each dimension and so we can gain a unique estimate of the curvature of this class. As the depth depends on ε we may also infer some knowledge about the depth of the class. Where the curvature is sufficiently different in each dimension the dynamics of recognition will converge to either a fixed point or a quasi-periodic cycle. Deeper classes will be represented by a fixed point attractor, the result of oscillation death, whilst shallow classes are represented by quasi-periodicity. We can again obtain additional information about the nature of the system from the quasi-periodic trajectory. For a two-dimensional potential we may observe a two-dimensional torus, this has two characteristic frequencies that are associated with the curvature of the class.

The alternative approach, where ε is a function of the similarity between a stimulus and the centre of the class also has perceivable benefits. Again the manifestation of a limit cycle is a result of the curvature, however, this cycle is now effected by the similarity between a class and a stimulus. ε is larger for a stimulus that is close to a local minimum. This stimulus will be recognized with a larger limit cycle than a stimulus further from the same minimum. If the curvature is sufficiently different the dynamics of the system will be dependent on ε . Stimuli that are towards the extremes of a minimum's basin of attraction will be represented by oscillation death. Stimuli that lie in close proximity to the fixed point will be represented by quasi-periodic trajectories. It

is therefore possible to distinguish between stimuli that are close or distant from the centre of the class and to also gain an appreciation of the local architecture of the basin based on the frequency of oscillation. If the basin is sufficiently similar in each dimension we will observe a limit cycle with a frequency that lies between the natural frequencies corresponding to the curvatures in each dimension. If, alternatively, we observe a tori or a fixed point, we are aware that the curvature is sufficiently different in each dimension. Information about the shape of the potential and hence the class may be derived from the two frequencies that characterize the torus. A stimulus at the extreme of the attractors basin may not identify overly well with the centre of the class, the fixed point type classification offers restricted information in comparison to our other regimes. This may be seen to reflect the limited similarity between a stimulus at the extreme of the class and the minimum that acts as the centre of the class.

5.4 A System of 3 Diffusively Coupled van der Pol Oscillators

The extension from a pair of coupled oscillators to a system composed of many mutually coupled oscillators mirrors the extension of the learning systems potential as the dimensionality of the stimulus is increased. By mutual coupling we infer that each unit is coupled to every other unit. Where i refers to a single van der Pol oscillator and N is the set of mutually coupled units. The system may be described as:

$$\ddot{x}_i = \varepsilon (1 - x_i^2) \dot{x}_i - \omega_i^2 x_i - B_D \sum_{j=0}^{j=N} (\dot{x}_i - \dot{x}_j) \quad (5.15)$$

In this instance we shall consider the case where $N = 3$. This will allow us to model our *RGB* potentials, demonstrated in Chapter 3, as oscillatory systems in a six-dimensional phase space. The extension to three oscillatory units is significantly more complex than the previously discussed two unit system. This is apparent from an appreciation of the truncated model which may now be written as:

$$\begin{aligned} \dot{A}_1 &= \frac{\varepsilon}{2} \left(A_1 - \frac{A_1^3}{4} \right) + \frac{B_D}{2} (A_2 \cos(\theta_A) + A_3 \cos(\theta_B) - 2A_1) \\ \dot{A}_2 &= \frac{\varepsilon}{2} \left(A_2 - \frac{A_2^3}{4} \right) + \frac{B_D}{2} (A_3 \cos(\theta_C) + A_1 \cos(\theta_A) - 2A_2) \\ \dot{A}_3 &= \frac{\varepsilon}{2} \left(A_3 - \frac{A_3^3}{4} \right) + \frac{B_D}{2} (A_1 \cos(\theta_B) + A_2 \cos(\theta_C) - 2A_3) \\ \dot{\theta}_A &= \frac{\omega_1^2 - \omega_2^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_2}{A_1} + \frac{A_1}{A_2} \right) \sin(\theta_A) + \frac{A_3}{A_1} \sin(\theta_B) - \frac{A_3}{A_2} \sin(\theta_C) \right) \\ \dot{\theta}_B &= \frac{\omega_1^2 - \omega_3^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_3}{A_1} + \frac{A_1}{A_3} \right) \sin(\theta_B) + \frac{A_2}{A_1} \sin(\theta_A) + \frac{A_2}{A_3} \sin(\theta_C) \right) \\ \dot{\theta}_C &= \frac{\omega_2^2 - \omega_3^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_3}{A_2} + \frac{A_2}{A_3} \right) \sin(\theta_C) - \frac{A_1}{A_2} \sin(\theta_A) + \frac{A_1}{A_3} \sin(\theta_B) \right) \end{aligned} \quad (5.16)$$

A derivation of this relation is included in appendix A.5. As the system is comprised of three units, there are three phase differences that must be defined. Here we have $\theta_A = \varphi_1 - \varphi_2$, $\theta_B = \varphi_1 - \varphi_3$ and $\theta_C = \varphi_2 - \varphi_3$. The simplifications that arise for a pair of coupled oscillators cannot be introduced in the same way for a system of three coupled oscillators. We should again

consider the phase locking region, this corresponds to limit cycle trajectories and is identified by constant phase relations. Hence phase locked solutions require $\dot{\theta}_A = \dot{\theta}_B = \dot{\theta}_C = 0$. Solutions that correspond to this condition are not apparent, even if the restricted case of symmetric amplitude ($A_1 = A_2 = A_3$) is considered. Determining the synchronization region analytically appears to present a considerable challenge.

The existence of phase locked solutions clearly depends on the frequency differences and the coupling strength B_D . As an alternative to the analytic approach we may infer the synchronization region by numerically considering the interceptions of a Poincaré plane. This method will highlight not only period-1 orbits, but also indicate regions of the parameter space where n:m:j frequency relations exist. These relations result in higher order periodic cycles. Initially we shall consider the case where the frequency ω_1 is held at a constant value $\omega_1 = 1$ and investigate the effect of other parameter changes.

5.4.1 Poincaré approach to characterising the Dynamical regions

Attempting to characterize the system via analytical means such as considering the truncated equations highlights the complexity of the dynamics. An appreciation of the system as described by equations 5.16 does not present a simple method to arrive at the set of parameters that correspond to the various expected regimes. This does not exclude the possibility that analytical methods may provide a suitable approach to classifying this system, but deriving a full analytical understanding of this system will remain outside of the scope of this investigation. Faced with the difficulties of analytical methods, we may again utilize numerical schemes to visualize the behaviour of the system for certain parameter relations.

A simple method to gain an appreciation of the systems dynamics is to construct a Poincaré section. Taking a subset of the phase space we may consider the intersections of our system's flow with this region. The number and distribution of intersections illustrates the qualitative nature of the system's behaviour. An alternative way to compile such a section is to consider the state of the system after a time interval T , often T is related to the period of a driving force. Utilizing such a regime a limit cycle may be identified by a single point on the section, whilst a quasi-periodic orbit may be identified by a closed curve. By considering the intersection of the trajectory with a plane in the phase space, limit cycle trajectories may be associated with a finite number of intersections, these intersection points may be useful in indicating some of the qualitative features of the trajectory and may indicate the ratios of the phases between the synchronized oscillators. Quasi-periodic orbits are associated with a closed curve. We should point out that where the tori characterising the quasi-periodic behaviour is N-dimensional, the Poincaré approach must be applied N times in order to observe a closed curve. It is therefore not always apparent whether a trajectory is high order quasi-periodic or chaotic from the application of a single section. For a system of three oscillators we may anticipate partial synchronization where two of the oscillators synchronize whilst the final one does not. This will be indicated by a two-dimensional torus. A three-dimensional torus will characterize the absence of all synchronization. As the system is dissipative, chaotic trajectories may be demonstrated by a large diffuse set of points or by fractal structures.

Utilizing the Poincaré section concept we may compare the number of unique intersections of the plane $x_1 = 0$ for any combination of parameter choices. The choice of section is clearly

important. As the trajectories encapsulate the fixed point of the system located at $x_1 = 0, y_1 = 0, x_2 = 0, y_2 = 0, x_3 = 0, y_3 = 0$ the choice of plane $x_1 = 0$ is certainly appropriate. A sample of the different behaviours that the system may demonstrate are presented in figures 5.11, 5.12 and 5.13 along with corresponding sections. Gradually varying each parameter value and noting the number of unique intersections, diagrams such as the one presented in figure 5.10 can be compiled. As there are five different bifurcation parameters it is only possible to visualize a small subset. From the perspective of synchronization the relationships between the frequencies, characterized by ω_1, ω_2 and ω_3 are of critical importance. Because of this we may study the relationship between ω_1, ω_2 and ω_3 whilst keeping ε and B_D constant. Considering the number of different intersection points that any trajectory has with the Poincaré plane as a function of its parameters we are able to compose figure 5.10. We have numerically simulated the system's solution for the given parameters and recorded the number of intersections the flow has with a Poincaré plane once a stable solution has had sufficient time to be reached. Limit cycles may be identified by a low number of different intersection points, whilst quasi-periodic orbits will typically have a much higher number. Here we have restricted our numerical scheme to retain no greater than 200 intersections. It is apparent that a large synchronization region exists when all three parameters ω_1, ω_2 and ω_3 fall within a certain bandwidth. The large section at the centre of figure 5.10 corresponds to 1:1:1 synchronization. Other smaller regions of n:m:j synchronization are also obvious, these regions are separated by areas of quasi-periodicity. One relation that is not distinguished by this plot is partial synchronization. A partially synchronized system should be described as a two-dimensional tori in the phase space, whereas, a complete lack of synchronization would be implied by a three-dimensional tori. These features can be clearly distinguished by comparing figures 5.11(i)(j) with 5.11(o)(p). In 5.11(j) the Poincaré section is a closed loop, this is not the case with 5.11(n) and 5.11(p). If we take the tori shown in 5.11(m) and 5.11(o) and apply sections with respect to both $x_1 = 0$ and $x_2 = 0$ then their sections contain closed loops as may be observed in figure 5.14. Taking a section with respect to two-planes enables us to ascertain that the tori's shown in figures 5.11(m) and 5.11(o) are three-dimensional, these quasi-periodic cycles are therefore associated with three distinct frequencies. We should note that there is also the possibility that this system may be chaotic for certain regions of the parameter space. Although we have shown that some of the yellow region of figure 5.10 corresponds to two and three-dimensional quasi-periodic orbits, it may also be possible that regions of this space are chaotic.

The array of trajectories presented in figures 5.11, 5.12 and 5.13 are intended to validate the regions shown in figure 5.10. (a)-(l) of figure 5.12 illustrate the main 1:1:1 region. Interestingly, there are various other regions of synchronization, these are apparent from figure 5.10, the dynamics within these regions are portrayed in figure 5.12(m)-(t) and also in figure 5.13.

A concern that may be raised with figure 5.10 is that ω_1 will not remain fixed for our evolving, learning system. Plotting the difference between ω_1 and $\omega_{2,3}$ does not exclude the possibility that the system behaviour is not only a function of the difference between these parameters, but also dependent on the absolute values. To consider this point we may apply the same procedure again, with the same parameters ε and B_D , but with a different ω_1 . A comparison is shown in figure 5.15, demonstrating the 1:1:1 synchronization region when $\omega_1 = 0.5, \omega_1 = 1$ and $\omega_1 = 1.5$. This highlights that the dynamics of the system are dependent not only on the differences between $\omega_{1,2,3}$, but also on their absolute values. If we consider the wider bifurcation space for the parameter values $B_D = 0.2, \varepsilon = 0.6$ and $\omega_1 = 1.5$, as shown in figure 5.16, it is evident that the space is organised

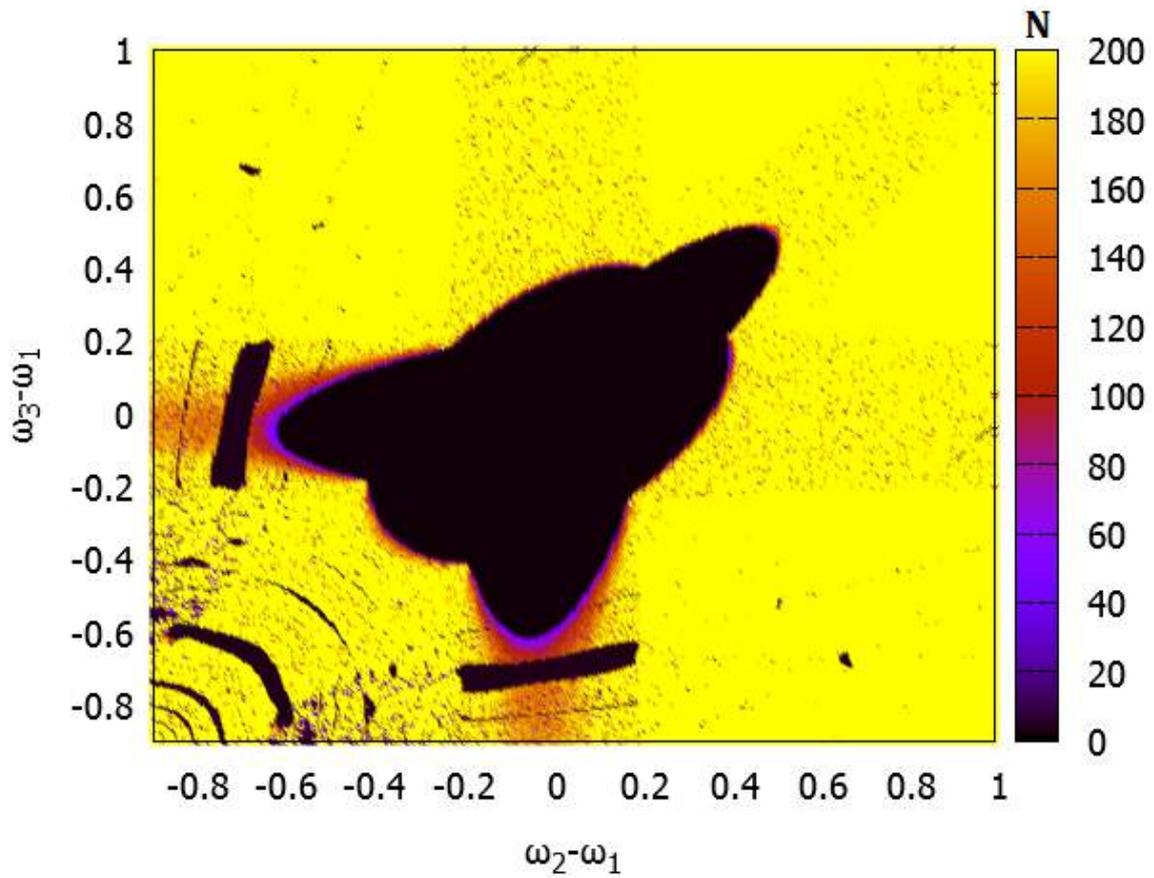


Figure 5.10: Bifurcation diagram indicating the behaviour of the system as a function of ω_2 and ω_3 . Colour indicates the number of intersections of a Poincaré plane located at $x_1 = 0$. The maximum number of permitted intersections is 200. Yellow regions therefore correspond to quasi-periodicity or chaos. For this illustration $\varepsilon = 0.6$, $B_D = 0.2$ and $\omega_1 = 1$.

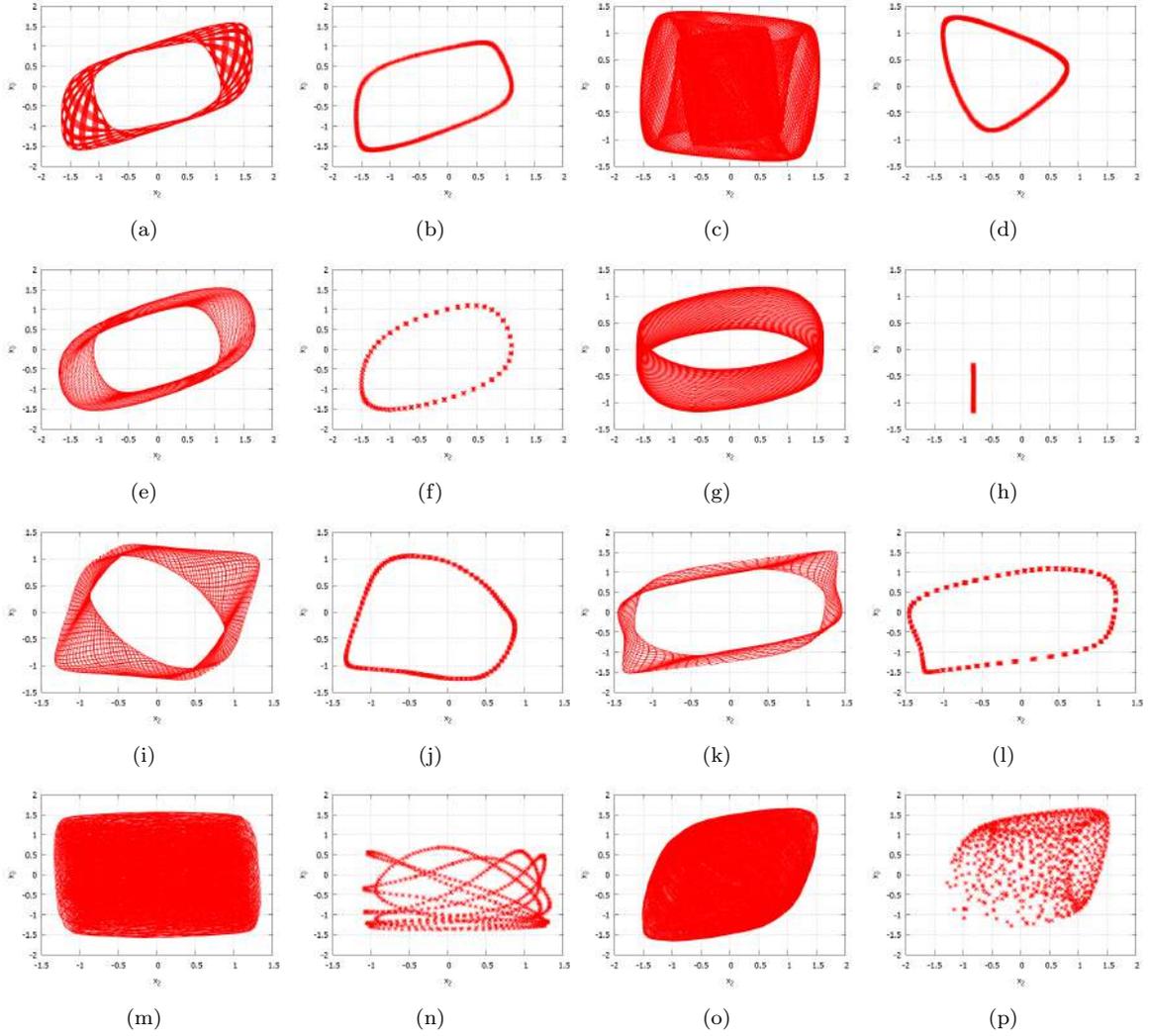


Figure 5.11: By selecting various relations for frequencies ω_2 and ω_3 , whilst keeping the parameters $B_D = 0.2$, $\varepsilon = 0.6$ and $\omega_1 = 1$ constant, the various dynamical regimes of three coupled van der Pol oscillators with dissipative coupling (equation(5.15) when $N = 3$) can be observed. The first and third columns demonstrate the systems trajectories in the phase space x_2, x_3 . The second and fourth columns demonstrate the intersection of these trajectories with a p Poincaré plane set at $x_1 = 0$. Here we observe a number of parameter values that correspond to quasi-periodic orbits. **(a),(b)**: $\omega_2 = 0.5$ and $\omega_3 = 0.32$, **(c),(d)**: $\omega_2 = 0.6$ and $\omega_3 = 1.4$,**(e),(f)**: $\omega_2 = 0.7$ and $\omega_3 = 0.5$, **(g),(h)**: $\omega_2 = 0.9$ and $\omega_3 = 0.35$, **(i),(j)**: $\omega_2 = 0.13$ and $\omega_3 = 0.36$, **(k),(l)**: $\omega_2 = 0.34$ and $\omega_3 = 0.14$, **(m),(n)**: $\omega_2 = 1.7$ and $\omega_3 = 0.7$, **(o),(p)**: $\omega_2 = 1.7$ and $\omega_3 = 1.4$. Quasi-periodic orbits are associated with a closed curve in the Poincaré section. **(a)-(l)** demonstrate two-dimensional tori, indicating partial synchronization. In contrast figures **(m)-(p)** correspond to a complete lack of synchronization. In these instances the tori are three-dimensional.

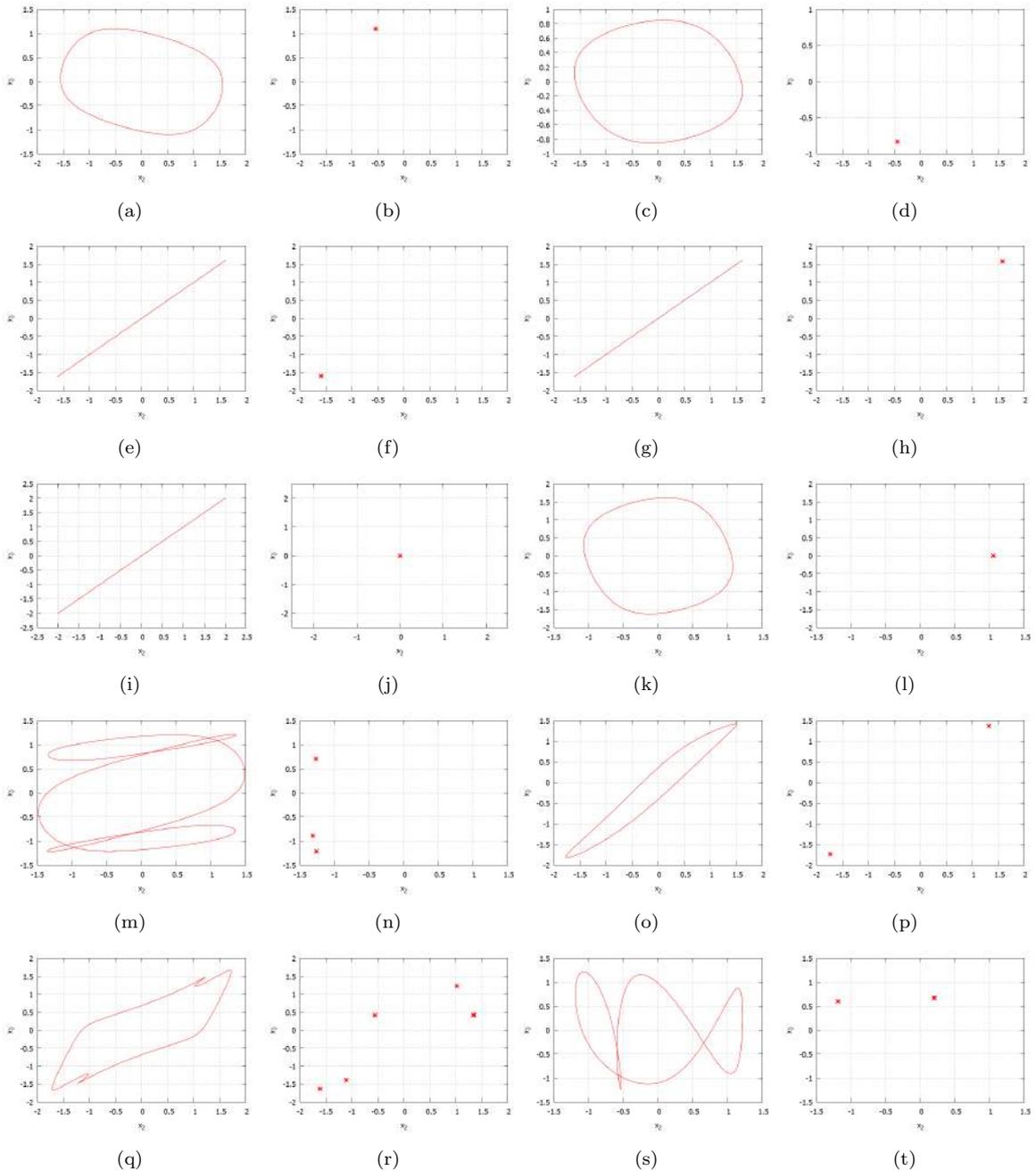


Figure 5.12: By selecting various relations for frequencies ω_2 and ω_3 , whilst keeping the parameters $B_D = 0.2$, $\varepsilon = 0.6$ and $\omega_1 = 1$ constant, the various dynamical regimes of three coupled van der Pol oscillators with dissipative coupling (equation(5.15) when $N = 3$) can be observed. The first and third columns demonstrate the systems trajectories in the phase space x_2, x_3 . The second and fourth columns demonstrate the intersection of these trajectories with a Poincaré plane set at $x_1 = 0$. Here we observe a number of parameter values that correspond to limit cycles. **(a),(b)**: $\omega_2 = 0.92$ and $\omega_3 = 1.28$, **(c),(d)**: $\omega_2 = 0.95$ and $\omega_3 = 0.4$, **(e),(f)**: $\omega_2 = 0.68$ and $\omega_3 = 0.68$, **(g),(h)**: $\omega_2 = 1.45$ and $\omega_3 = 1.45$, **(i),(j)**: $\omega_2 = 1.0$ and $\omega_3 = 1.0$, **(k),(l)**: $\omega_2 = 1.33$ and $\omega_3 = 1.0$, **(m),(n)**: $\omega_2 = 0.81$ and $\omega_3 = 0.27$, **(o),(p)**: $\omega_2 = 0.45$ and $\omega_3 = 0.49$, **(q),(r)**: $\omega_2 = 0.245$ and $\omega_3 = 0.16$, **(s),(t)**: $\omega_2 = 0.485$ and $\omega_3 = 1.52$. Limit cycle are indicated by single points in the Poincaré plane.

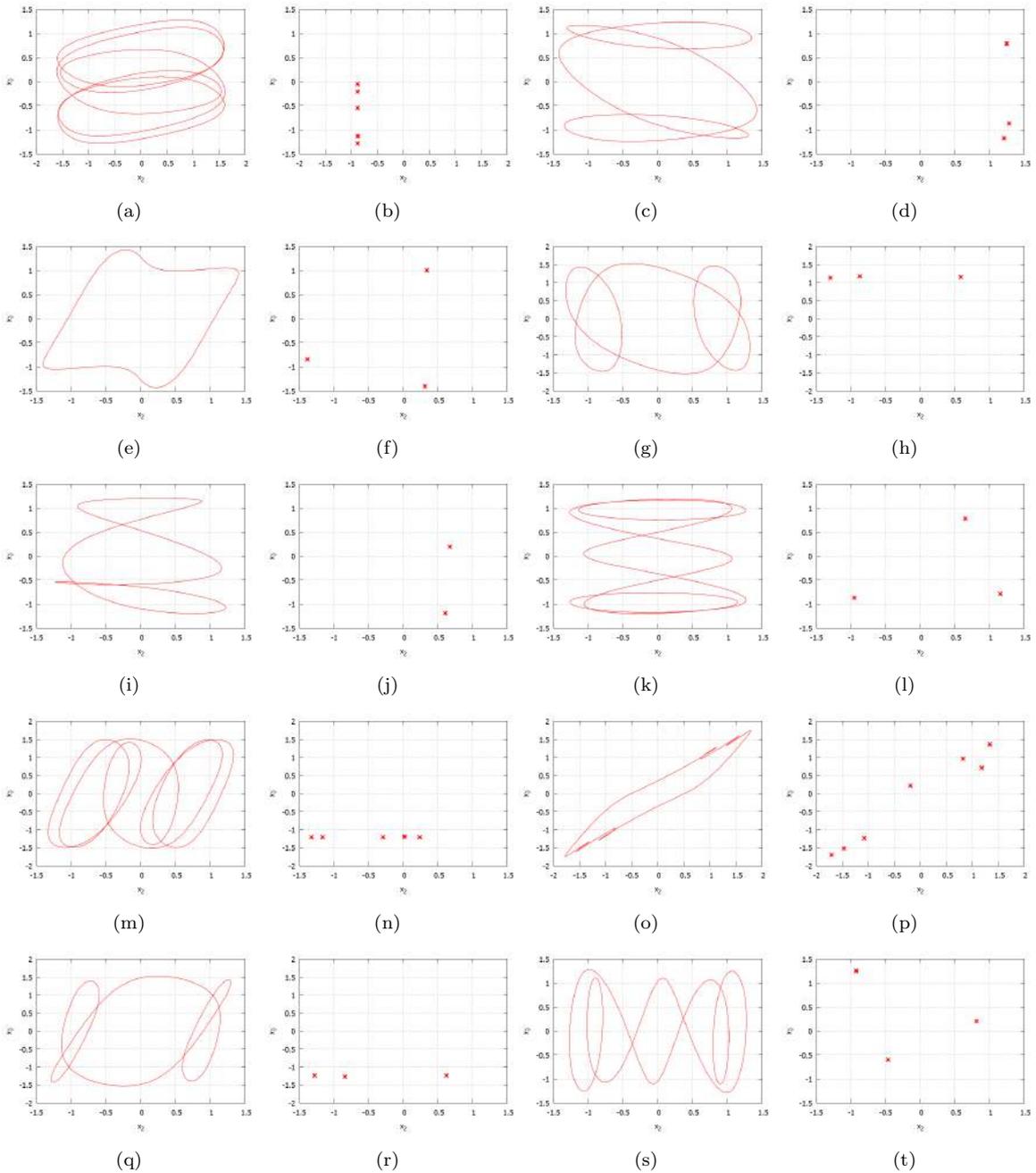


Figure 5.13: By selecting various relations for frequencies ω_2 and ω_3 , whilst keeping the parameters $B_D = 0.2$, $\varepsilon = 0.6$ and $\omega_1 = 1$ constant, the various dynamical regimes of three coupled van der Pol oscillators with dissipative coupling (equation(5.15) when $N = 3$) can be observed. The first and third columns demonstrate the systems trajectories in the phase space x_2, x_3 . The second and fourth columns demonstrate the intersection of these trajectories with a Poincaré plane set at $x_1 = 0$. Here we observe a number of parameter values that correspond to limit cycles. (a),(b): $\omega_2 = 0.89$ and $\omega_3 = 0.165$, (c),(d): $\omega_2 = 1.17$ and $\omega_3 = 0.33$,(e),(f): $\omega_2 = 0.16$ and $\omega_3 = 0.4$, (g),(h): $\omega_2 = 0.34$ and $\omega_3 = 1.14$, (i),(j): $\omega_2 = 1.52$ and $\omega_3 = 0.485$, (k),(l): $\omega_2 = 1.67$ and $\omega_3 = 0.32$, (m),(n): $\omega_2 = 0.16$ and $\omega_3 = 0.83$, (o),(p): $\omega_2 = 0.18$ and $\omega_3 = 0.15$, (q),(r): $\omega_2 = 0.26$ and $\omega_3 = 0.82$, (s),(t): $\omega_2 = 0.31$ and $\omega_3 = 1.68$. Limit cycle are indicated by points in the Poincaré plane. Higher order cycles will correspond to a greater number of points.

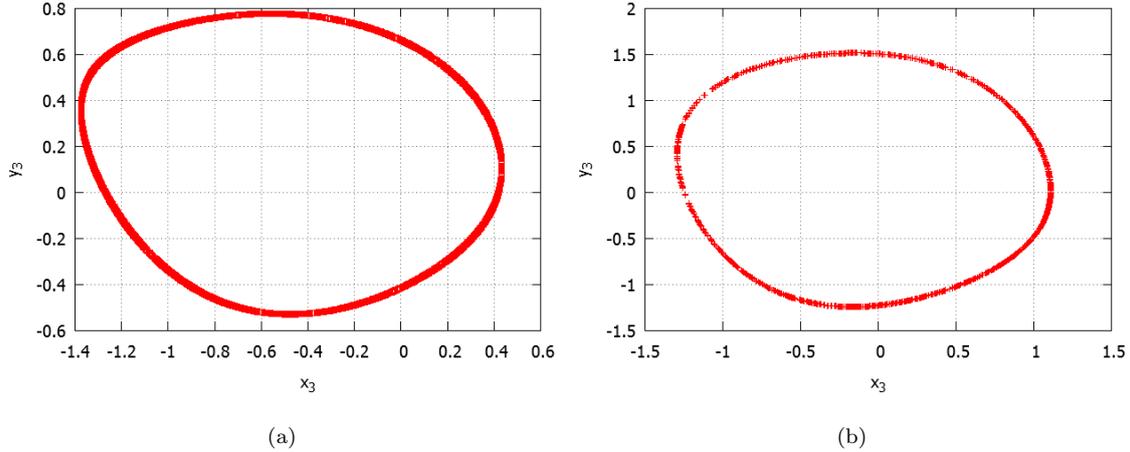


Figure 5.14: Poincaré sections with respect to the planes $x_1 = 0$ and $x_2 = 0$ for a set of 3 coupled oscillators described by equation (5.15) ($N = 3$). In both figures $\omega_1 = 1$, $\varepsilon = 0.6$ and $B_D = 0.2$. In **(a)** $\omega_2 = 1.7$ and $\omega_3 = 0.7$. **(b)** corresponds to the case where $\omega_2 = 1.7$ and $\omega_3 = 1.4$. In both examples the systems dynamics are classified as third order quasi-periodic, this is identified by the closed loop show by the Poincaré section. The quasi-periodic trajectories for these systems are illustrated in figures 5.11m and 5.11o respectively.

with similar regions to those demonstrated in figure 5.10 when $\omega_1 = 1$. However, the relative size and separation between regions is somewhat different as a result of the change in ω_1 .

It is evident from figure 5.15 that the values of $\omega_{1,2,3}$ have implications on the bifurcation structure of the system in addition to the implications of their differences. In each sub-plot the region surrounding the 1:1:1 synchronization tongue within the parameter space $((\omega_2 - \omega_1), (\omega_3 - \omega_1))$ is demonstrated for different values of ω_1 . Quasi-periodic (or chaotic) trajectories are associated with yellow regions, whilst limit cycles are indicated by black areas. Making a qualitative comparison, the shape and approximate size of the 1:1:1 synchronization region appears to remain relatively constant, however, comparing figures 5.10 and 5.16 highlights the wider differences that appear as a result of changing ω_1 .

Thus far we have only numerically considered the synchronization regions of the system for the parameter values $\varepsilon = 0.6$ and $B_D = 0.2$. It is evident that this parameter choice coincides with either quasi-periodic or limit cycle orbits. A basic investigation of the system demonstrates that, as was the case with a pair of coupled oscillators, the fixed point of the system may become stable as a result of oscillation death. Unlike for the two oscillator composition, the quasi-periodic and oscillation death regimes are not separated by the line $B_D = \varepsilon$. We may identify a set of conditions that permit oscillation death and again study the regimes of the system as the relations between $\omega_{1,2,3}$ are manipulated. Figure 5.17 illustrates the possible dynamics of the system when $\varepsilon = 0.3$, $B_D = 0.2$ and $\omega_1 = 1$. The black region of figure 5.17 corresponds to 1:1:1 synchronization, whilst the white regions highlights parameter values corresponding to oscillation death. Comparing this figure to figure 5.10 it is apparent that the selection of ε and B_D plays an important role in defining both the range of dynamics that the system may demonstrate and also the size and shape of the 1:1:1 synchronization region. In contrast to the previous demonstrations, figure 5.17 does not contain any higher order n:m:j synchronized regions.

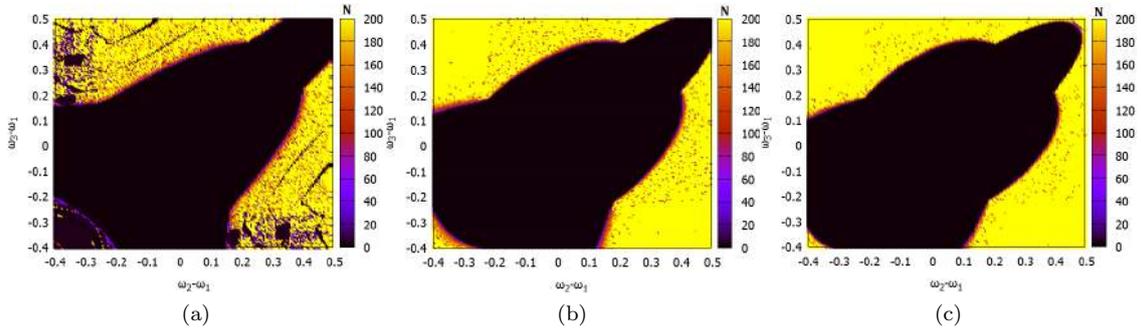


Figure 5.15: Comparison between the bifurcation space when (a): $\omega_1 = 0.5$, (b): $\omega_1 = 1.0$ and (c): $\omega_1 = 1.5$. Note that in each of these cases the same ranges of ω_2 and ω_3 are utilized, also $B_D = 0.2$ and $\varepsilon = 0.6$. The colour corresponds to the number of intersections of the system trajectory with a Poincaré plane set at $x_1 = 0$. No more than 200 intersections are registered, yellow regions therefore correspond to quasi-periodic orbits whilst black regions highlight 1:1 synchronization.

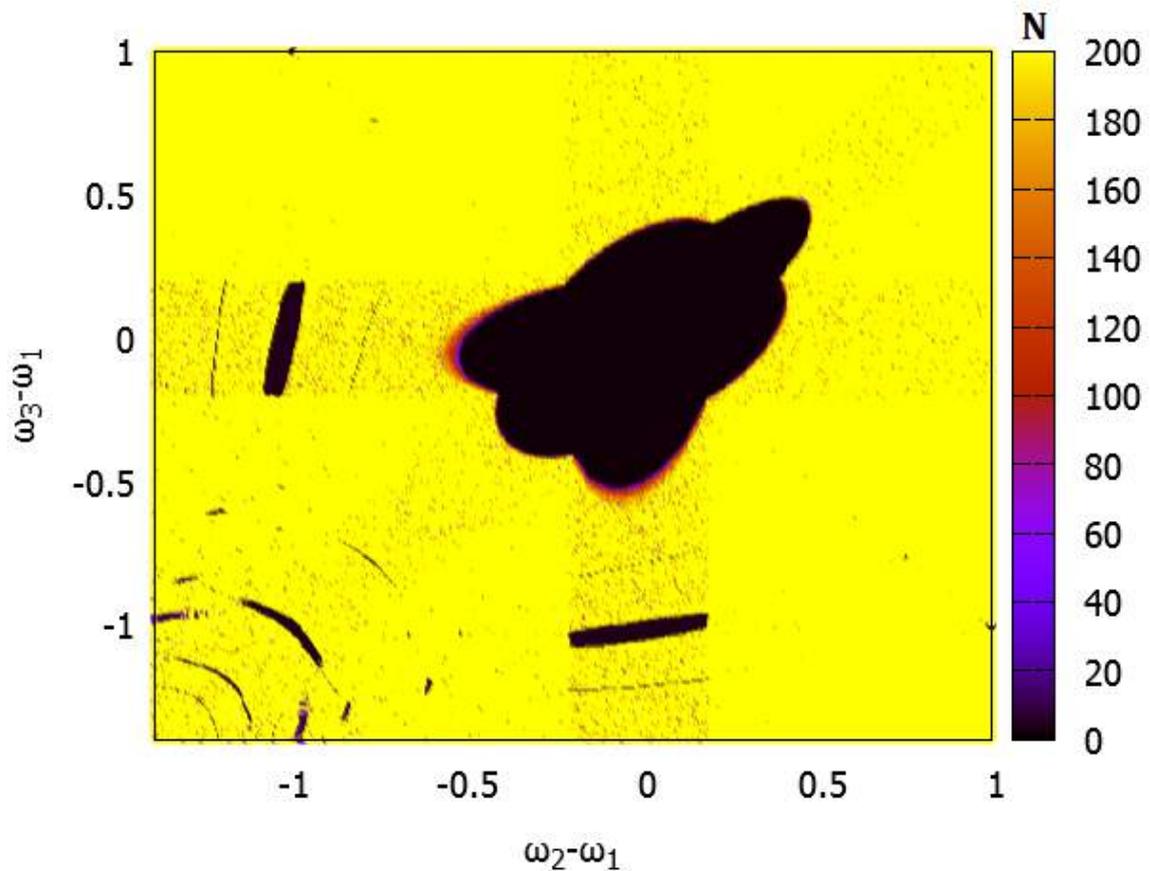


Figure 5.16: Bifurcation diagram indicating the behaviour of the system as a function of ω_2 and ω_3 . Colour indicates the number of intersections of a Poincaré plane located at $x_1 = 0$. The maximum number of permitted intersections is 200. Yellow regions therefore correspond to quasi-periodicity. For this illustration $\varepsilon = 0.6$, $B_D = 0.2$ and $\omega_1 = 1.5$.

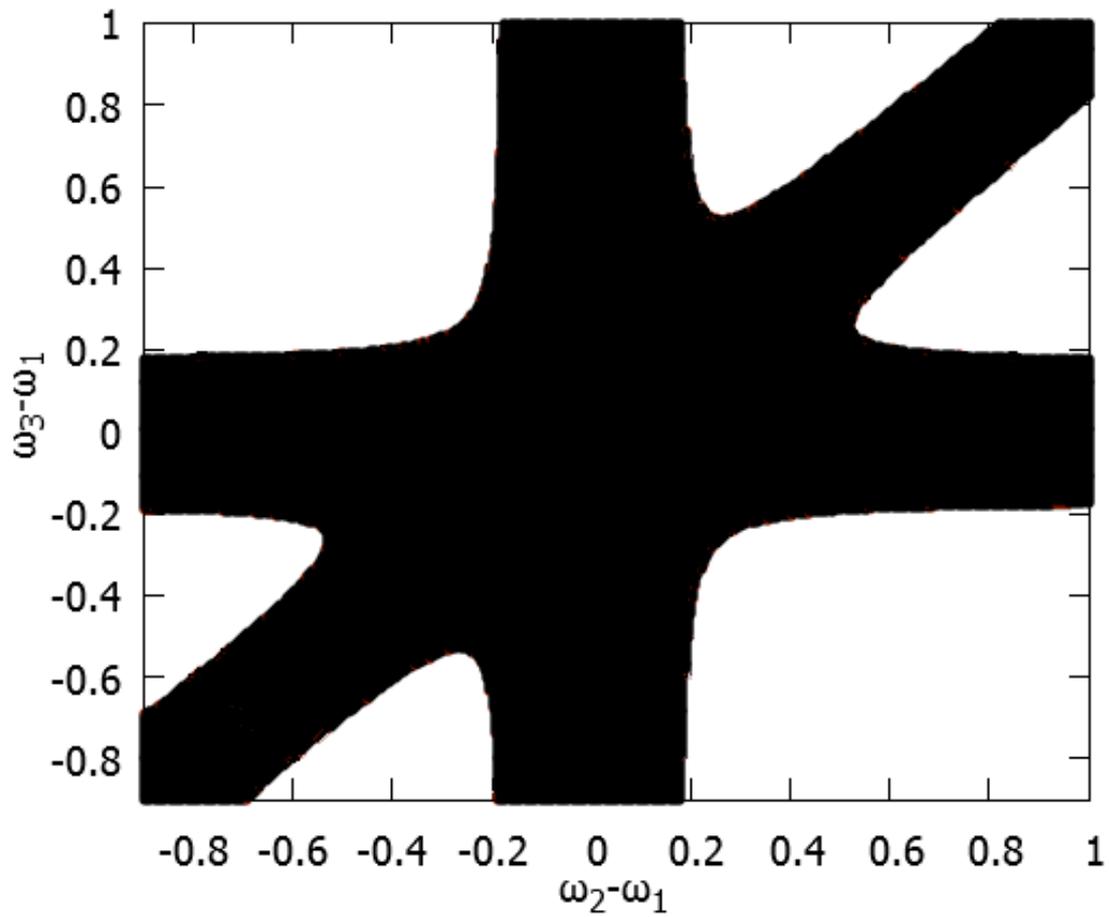


Figure 5.17: The dynamics of the system are represented via the colour of the plot, the area of white corresponds to oscillation death. Within this region the trajectories of the system converge to the fixed point located at the origin. Black infers the existence of a limit cycle. Parameters are set to $\varepsilon = 0.3$, $B_D = 0.2$ and $\omega_1 = 1$.

5.5 Oscillatory Recognition of Colour

Although gaps still remain in our knowledge of the behaviour of three coupled van der Pol oscillators, we may demonstrate the recognition of stimuli utilizing this approach. Here we shall revisit the recognition of colour as previously studied in Chapter 3. Our model is again built around the concept of a discrete R, G, B space where every point in this space refers to a different colour. The potential of the system is therefore a function of three spacial coordinates and time. The recognition of a stimulus is described by:

$$\begin{aligned} \frac{dx_{*R,G,B}}{dt} &= -\gamma \frac{\partial V}{\partial x_R} \\ \frac{dx_{R,G,B}}{dt} &= y_{R,G,B} \\ \frac{dy_{R,G,B}}{dt} &= \varepsilon \left(1 - (x_{R,G,B} - x_{*R,G,B})^2 \right) y_{R,G,B} - \alpha \frac{\partial U}{\partial x_{R,G,B}} \\ &\quad - B_D ((y_{R,G,B} - y_{G,B,R}) + (y_{R,G,B} - y_{B,R,G})) \end{aligned} \quad (5.17)$$

where

$$U(V, \mathbf{x}) = V + \frac{1}{2}g \left(\frac{\partial^2 V}{\partial x_{R,G,B}^2} \right) (x_{R,G,B} - x_{*R,G,B})^2$$

Here the subscripts R, G, B refer to the spacial dimensions of the system, each described by an oscillator. Each oscillator is influenced by the curvature of the potential in the corresponding dimension. This consideration is captured by taking partial derivatives with respect to the function $U(V(R, G, B, t))$, this determines the natural frequency of each oscillator. The function $g()$ is described by equation (4.6) and ensures that the function $U()$ takes an appropriate estimate of the local topology of the potential $V()$ surrounding \mathbf{x}_* . This function remains smooth, continuous and appropriately orientated. The three oscillators are mutually coupled via a dissipative coupling term with coupling strength B_D , the combined dynamics are therefore a function of the three natural frequencies. The location of the attractor in phase space is influenced by the local minima of the potential $V(R, G, B, t)$ due to the inclusion of the terms $(x_{R,G,B} - x_{*R,G,B})$. Here $x_{*R,G,B}$ is located via gradient descent. This formulation of recognition can be seen to closely resemble the description of recognition provided by equation (4.10) for a one-dimensional potential. The principal alteration is the inclusion of the coupling term, as such equation (5.17) may be seen as a generalization of equation (4.10).

As an initial test of this approach we may consider the trajectories of the system for a static potential. Figures 5.18b and 5.18d illustrate the oscillatory dynamics of recognition for the potentials shown in figures 3.8q, 3.8r, 3.8s and 3.11q, 3.11r, 3.11s respectively. These potentials are both the result of two million stimuli, corresponding to the cartoons Southpark and Futurama. 125 initial recognition conditions form a mesh of coordinates, as such, the realizations compare the recognition of colours from across the entire R, G, B space. The gradient aspects of the systems are demonstrated in figures 5.18a and 5.18c. Gradient descent locates the minima of the potentials. Oscillatory dynamics can be seen to correspond to the locations of these minima in the phase space. It is important that parameter values are chosen carefully if both quasi-periodicity and limit cycles are to be observed. As we previously investigated a set of three coupled van der Pol oscillators with the parameter values $\varepsilon = 0.6$ and $B_D = 0.2$, we shall persevere with these assignments. In this investigation ω_1 is now a function of the potential, as such we cannot directly compare the

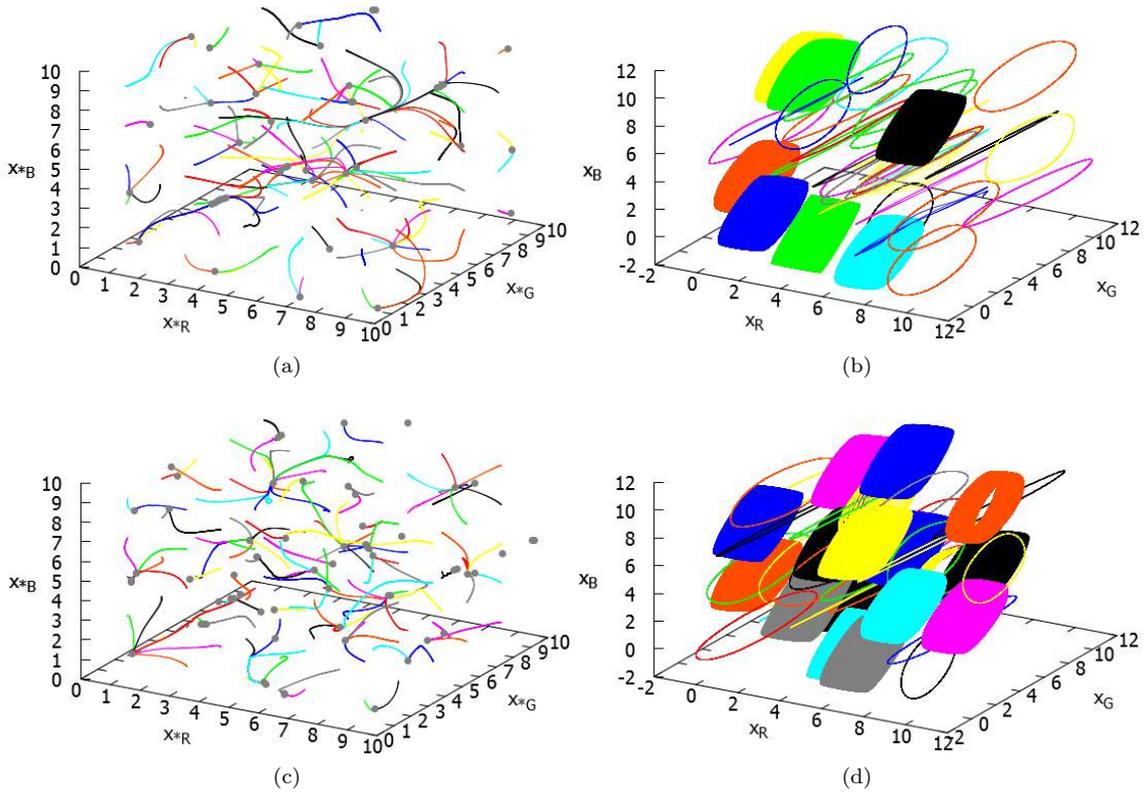


Figure 5.18: **(a)**: Gradient descent for an ensemble of 125 initial conditions that locate the local minima of a potential shown in figures 3.8q, 3.8r and 3.8s. This potential is created by the repeated application of stimuli corresponding to the cartoon Southpark. **(b)**: Recognition of these 125 stimuli in accordance with equation (5.17). **(c)(d)**: Repetition of the same procedures shown in (a)(b) but for the stationary potential shown in figures 3.11q, 3.11r, 3.11s. This corresponds to the cartoon Futurama. In each instance $\gamma = 50$, $\alpha = 400000$, $B_D = 0.2$ and $\varepsilon = 0.6$.

behaviour observed in this model with the bifurcation diagrams shown in figure 5.15. In both diagrams 5.18b and 5.18d the parameter $\alpha = 4 \times 10^5$, this ensures that the system is appropriately scaled to permit the existence of both limit cycles and quasi-periodic orbits. Previously we discussed methods to control ε so as to allow quasi-periodicity and oscillation death. Although these behaviours are show in section 5.4, it remains unclear what parameter values correspond to the bifurcation between quasi-periodicity and oscillation death.

In both figures 5.18b and 5.18d there are observable limit cycles and quasi-periodic trajectories. As ε remains constant the shape of each orbit is related to the frequency of oscillation. It is important to stress that these attractors reflect the recognition of stimuli for a static potential. Such a profile is only attainable when the learning system has received sufficient stimuli, in this circumstance we may consider the model to have finished learning and become an expert on its environment. We may compare the realizations of recognition demonstrated as oscillatory trajectories in figures 5.18b and 5.18d with recognition presented via gradient descent in figures 5.18a and 5.18c. In both cases the stimuli chosen correspond to the same locations in the RGB space. It is apparent that the oscillatory attractors yield much more information about the classes of colour than represented by fixed points. The gradient approach only indicates the dominant colour of each class, however, the oscillatory model indicates both the dominant colour and the size of the class. The size of the class is indicated by the frequency of oscillation, which relates

to the curvature of $V(R, G, B)$ in the vicinity of a local minimum, created by the clustering of stimuli. The behaviour of a trajectory may infer several important features of a class:

- If the attractor of the system is a two-dimensional torus then we can discern that the curvature and hence, the relative size of the class in each characteristic direction is substantially different. The attractor has three distinct frequencies that indicate the size of the class in each direction.
- If the attractor of the system is a one-dimensional torus then we should be aware that the class has a similar span in two directions, which are sufficiently different to the span of the class in the third direction. The one-dimensional torus has two characteristic frequencies, these frequencies highlight again highlight the size of the class.
- When the attractor is a limit cycle we should be aware that the curvature is similar in each characteristic direction of $V(R, G, B)$. This informs us that the class has a similar span in each characteristic direction. This span, and hence the size of the class is associated with the frequency of oscillation.

Information regarding the size of a class is useful as it may be used to specify how exact the systems representation of colour is. It is useful to know whether a class representing red, represents only one shade of red or a whole plethora of related colours. Each class of the model is represented by a unique phase space attractor that is specific to the topology of the class. By comparing attractors with similar dynamics, occupying different regions of the phase space, we can ascertain a relationship between the different classes. Although we have not yet been able to piece these different dynamic attractors together into a structure that represents the relation between different classes in a hierarchical framework, it would seem clear that characterising classes with different types of attractor is a pivotal ingredient towards achieving this goal.

The next phase of the models development is to consider the learning and recognition of unfamiliar stimuli whilst the model is learning. This procedure would typically follow the framework introduced in section 3.4, but this new representation of recognition would take the place of gradient descent. The phase space corresponding to the potential $V(R, G, B, t)$ contains six dimensions, we should be aware that this may present a problem when trying to illustrate the systems attractors. Although we may continue to neglect a number of these dimensions and still observe the corresponding behaviour, we should be careful as this may be to the detriment of the intuitiveness of the system. We should also be cautious about utilizing the van der Pol model when many oscillators are coupled. Although the coupling of two oscillators is well understood, we have highlighted that this knowledge is not directly transferable to systems comprising three or more oscillators. Due to non-linearity we may not be able to surmise the behaviour of many coupled oscillators without studying each composition in turn. This issue requires resolution as it is important that the model retains generality and is not restricted to a given number of dimensions.

5.6 Summary and Conclusion

A representation of recognition via oscillatory dynamics is again discussed in this Chapter. This is necessary as we require a method to characterize the different classes of the system when the stimuli, and hence the potential, is formed of several components. These components may refer

to variables such as red, green and blue as was considered for a gradient style recognition method back in Chapter 3. As with the approach for a one-dimensional stimuli, detailed in Chapter 4, we may associate the classes of the model with oscillators. This has the advantage that we can convey additional qualities of a class beyond its location in the space $\mathbf{x} = (x_1, x_2, \dots, x_N)$. Representing classes with higher order attractors also yields the possibility that hierarchy may be conveyed by the relation between different attractors.

We continue to consider the van der Pol oscillator as a paradigmatic system, but must make some important adjustments to reflect the dimensionality of the system. It is important to point out that there may be numerous methods to incorporate this increase in dimension, here we have considered coupling. The principal problem that we are faced with when considering this more general system is how to create a dynamical construct that encapsulates the notion of recognition. We require that the dynamical system identifies a class corresponding to the stimulus. In keeping with the discussion of Chapter 4 we also require that the scheme provides more useful information than is accessible with a simple point attractor. Permitting higher order dynamics such as limit cycles and quasi-periodicity should enable the development of a model where different stimuli may be considered as parts of classes and sub-classes. Here we have utilized coupling to create a model of recognition that indicates the location of a class in phase space and defines the class's topology. This progression is significant as the phase space and potential may plausibly be of any dimension. This assertion requires further investigation as it is apparent that the regimes of the van der Pol oscillator become more complex as the number of coupled oscillators is increased. Although we have not yet been able to capture the concept of class and sub-class by this dynamical representation, it is important to appreciate that by representing the recognition of stimuli via limit cycles and quasi-periodic orbits we make an important step towards this goal.

The premise behind utilizing coupling is that each dimension of the potential may be associated with a van der Pol oscillator. The natural frequency of each oscillator remains dependent on the curvature of the potential in the corresponding dimension. For a system where $\mathbf{x} = (x_1, x_2)$ the potential will be two-dimensional, two van der Pol oscillators will therefore interact to represent the recognition of a stimulus. One oscillator is related to the dimension characterized by x_1 and the other is related to x_2 . Where the system is one-dimensional, it is clear that the system is no different to the model discussed in Chapter 4. This new method should be applicable regardless of the dimensionality of the system. A problem does arise however, as the theory concerning many coupled oscillators is not well defined.

The theory of coupled van der Pol oscillators is well understood for some restricted cases, in other circumstances the theory is less well developed. Certainly the dynamics of pairs of oscillators with various approaches to coupling may be appreciated from the existing body of literature [10, 27, 124, 146, 132, 86, 122]. Conversely, less is known regarding the dynamics of larger sets of mutually coupled oscillators. In this domain studies have typically focused on oscillators arranged in rings or chains [42, 121, 40, 41], considerations of mutual coupling appear to have been neglected. From the perspective of our investigation it would appear illogical not to couple all oscillators mutually.

In this chapter we began our investigation by considering a potential $V(x_1, x_2)$ and discussed how coupling could be applied to create a model of recognition that utilized limit cycles and quasi-periodic orbits. Again we resolved to utilize a van der Pol system, however, in this instance coupling was utilized to relate the different oscillators describing the different dimensions of the model. Coupling of the system is shown to be important as it permits synchronization, without

this all trajectories would appear quasi-periodic in phase space. Before coupling could be applied to the learning model it was important to consider the different methods of coupling. Dissipative coupling appears to have the advantage over reactive coupling in this situation as the possibility of oscillation death is permitted and multi-stability is avoided. Having made this assessment, a clear understanding of the bifurcation structure was derived for a pair of coupled oscillators by considering the truncated equations representing the system. These analytic inferences were then compared to numerical simulations for validation. Having defined the regimes of the two-oscillator system, the same concepts could then be applied to the task of recognition. Initially a model was considered where ε remained fixed, here it was shown how trajectories may be quasi-periodic or form limit cycles. Where a position on the potential has a similar curvature in both dimensions the frequencies will be close to each other, the dynamics will therefore synchronize, and limit cycles will be observed. If the curvatures are sufficiently different, there will be a substantial difference between the two natural frequencies. In this case the oscillators will not synchronize, the phase space dynamics will indicate a torus if $\varepsilon > B_D$, or a fixed point if $\varepsilon < B_D$. In each eventuality the system will have a frequency that is dependent on the topology of the potential. As such, this method of recognition will identify a class and also define its relative structure. As the system permits the possibility of oscillation death, it would appear advantageous to exploit this property. This may be achieved by manipulating the parameter ε . By continuing to consider a two oscillator model we discussed two conceivable motivations for adjusting ε , firstly as a function of the potential's depth, and secondly as a function of the similarity between the minimum of the potential and the stimulus. Realizations of both of these approaches are graphically provided in section 5.3.1.

We may also consider that oscillation death may be utilized as part of a hierarchical scheme. Previously we have speculated about the interaction of limit cycles, this description may actually begin from a point attractor, the result of oscillation death. This may represent the lowest subclass, before larger classes are described by limit cycles and tori. As yet we have not been able to define an appropriate method to illicit this hierarchy, we must stress that permitting an ensemble of different phase space attractors should be an appropriate place to begin.

Progressing from a potential with two spatial dimensions to a profile with three, the learning model is associated with three coupled oscillators. This presents a greater challenge as the dynamics of such systems are not easily understood. We begin our investigation of three coupled oscillators from the perspective of the truncated equations, however, this line of investigation appears to be intractable when trying to derive the conditions for synchronization. Because of this issue we resolved to consider a numerical appreciation of the system. Because three coupled oscillators have five feasible bifurcation parameters ($\varepsilon, B_D, \omega_1, \omega_2, \omega_3$), it is not possible to plot bifurcation diagrams that describe the entire range of the system's dynamics. Here we focused on the system's bifurcation structure as the parameters ω_1, ω_2 and ω_3 were adjusted. As we can only adjust two of these parameters on a single plot, we considered various values of ω_1 and varied $\omega_{2,3}$ accordingly. It is apparent from these realizations that various regions of synchronization exist. The largest region corresponds to 1:1:1 synchronization. It is difficult to speculate how other n:m:j regions should be perceived in the context of learning. As the size and location of each is a function of so many parameters, it is unlikely that we can restrict the frequency of our system to a space where only 1:1:1 oscillations and surrounding quasi-periodicity remains. The large cross-section of different limit cycles and quasi-periodic orbits (second and third order) disappear if the parameters ε and

B_D are selected from within the range corresponding to oscillation death. In this region it appears that only fixed points and 1:1:1 synchronization remains. All other $n:m:j$ orbits die out.

Despite a lack of complete knowledge regarding the higher order synchronization regions, we may still apply the van der Pol structure to our learning system. We should stress that a much deeper investigation of coupled oscillators is required before we can satisfactorily utilize and appreciate the regimes of the model in a wider context. Nevertheless, we consider the recognition of stimuli related to two different static potentials in section 5.5. These potentials correspond to the learning of the different colours composing the cartoons Southpark and Futurama. It is clear that with an appropriate choice of α , we may scale the frequency of the system so as to permit both quasi-periodic and limit cycle trajectories. Although we previously considered the effect of ε as a control of oscillation death/quasi-periodicity, we do not incorporate this in the model at this stage due to a lack of knowledge regarding three coupled van der Pol systems.

It is clear that further work is required before we can adequately model recognition when the potential is N -dimensional. The basis of this further investigation centres around deriving the conditions for certain dynamical behaviours for sets of coupled oscillators and also appreciating the effect of increasing the number of coupled oscillators. As the system under study is non-linear, an increase in the number of oscillators may have a significant effect on the system's dynamics.

Chapter 6

Summary and Conclusion

Understanding the mechanics of the human brain is indisputably one of the most complex tasks faced by modern science. Despite the vast sums of money driving multidisciplinary research, a whole plethora of questions remain unanswered. Large research bodies such as the Human Brain Project [1] and the Brain initiative [155] seek to discern the intricate processes that occur in the brain and hope to define how these infer thought, behaviour and memory. New knowledge may not only instigate a revolution in brain research, but may also have far reaching ramifications across a diverse range of related fields. It is thought that a better understanding of biology will inspire not only a new frontier of medicine and healthcare, but also instruct new approaches in areas such as AI and computer intelligence.

Unfortunately, this AI revolution has not yet come to fruition. Considering the classical models of AI, it is clear that a substantial paradigm shift may be required if a new generation of AI systems are to be developed. As discussed in section 1.2, the current models of AI take exceptionally varied approaches. The neural network draws its motivation from the networks formed by neurons in the biological brain. This is certainly a solid starting point, however, the model meets with some clear difficulties. These limitations are typically centred around the algorithms used to model the interactions of neural units. Although the structure of the system is similar to the brain, the way in which such models learn typically bear little resemblance to the mechanisms thought to be occurring in humans. The human brain continuously learns from the myriad of stimuli that it receives about its environment. It is able to simultaneously interpret stimuli perceived from each of the body's receptors, autonomously selecting the important features and assigning essential information to memory. All the while it is able to compare these stimuli to retained knowledge, in order to reach logical conclusions and decide actions. Learning is a continuous process that may be achieved via a number of strategies, these include, reinforcement learning, supervised learning, goal-based learning, and perhaps most commonly, unsupervised learning. Modelling unsupervised learning appear to be a particular challenge for AI systems, however, the brain appears to be most reliant on this process. Consider how we interact with our surroundings during our everyday life, an example may be learning where objects are in a room. This is not something we are taught, we perform such learning intrinsically. Unsupervised learning typically requires little focussed attention, but is the most essential, basic, core mechanism that allows us to successfully interact with our surroundings. An AI model that only learns via reinforcement or supervised techniques is clearly a poor substitute for biology. Although no artificial system is able to perform computations that remotely resemble those performed by the brain, it is clear that some models have more in

common with the way the brain learns than others. In Chapter 1 we considered a number of prominent AI systems. The majority operate in two discrete phases. The first phase involving learning, before the system is then applied to some task. During this task the system is incapable of further learning. These systems are clearly highly inflexible and do not resemble natural learning. Models that operate in this fashion include neural networks, decision tree learners and genetic algorithms. In fact, the majority of current AI approaches appear to share this fundamental flaw. It would seem clear that any plausible model of AI should be able to learn about its surrounding autonomously (unsupervised learning) in order to create a flexible internal representation of its experiences. It should also be able to learn continuously, and should quickly recognize and respond to stimuli. We postulate that these features should form the basic requirements for a new generation of AI machine.

The need for a new approach to the problem of learning in artificial systems is apparent. The current methods utilized within the field of AI suffer from a range of issues. In contrast, the self shaping dynamical system proposed by Janson and Marsden [74, 99] demonstrates a number of advantages over conventional methods. The greatest success of this model is that the recognition and classification of inputs occurs simultaneously via a mechanism referred to as on-line learning. The dynamical system is known to adjust its architecture to external stimuli in a non-algorithmic, unsupervised manner. The potential function ($V(t, \mathbf{x})$) of the dynamical system is perturbed by a special function that represents a stimuli applied to the system. In Chapter 2 this perturbing function is described by a Gaussian curve. With the repeated application of a stimuli function the shape of this potential is transformed to demonstrate different classes of retained knowledge. The potential may therefore be conceptualized as a form of memory, where each class corresponding to a local minimum refers to information retained by the system. This in itself is significant, as it presents the possibility for the system to be used as an auto-classifier, grouping together similar inputs into distinct classes. Because the system groups similar data autonomously, the model possesses an inherent tolerance to noise.

Despite the major promise shown by this prototype, it is clear that further refinements may be applied to improve performance. In this investigation we have postulated several pertinent changes that may further the development of this model and improve its ability to characterize the processes of learning, retention, classification and recognition. We have focused on relating these refinements to the mechanisms apparent in real learning systems. We show the ability of the system to learn from, and recognize stimuli, in an arbitrary environment, without the restriction of supervision.

As this new class of system remains in its infancy, it was important to test its performance for a range of stimuli so as to verify the assertions made about its self-shaping characteristics. In Chapter 2 we studied the dynamics of the system for a range of different simulated inputs by comparing the shape of the system's potential $V(t, x)$ to the probability density distributions of defined stimuli. It was shown that the system's potential autonomously shapes to reflect the PDD of inputs. This behaviour has already been reported in [74, 99], but here we have considered stimuli that possess a much higher level of complexity. Instead of selecting arbitrary values for the Gaussian width parameter σ , which relates to the perturbing function and controls the system's ability to differentiate between classes of input (and can therefore be considered as an indicator of certainty), we considered the impact that this parameter has on the number of classes that emerge. We show that selecting smaller values of σ ensures that the potential function $V(t, x)$

closely approximates the PDD of a stimuli, this relation is not observed if σ is too large. It should be highlighted that selecting a value too small increases the time taken for convergence and may also result in a disproportionate number of minima forming early in the model's evolution. We postulate that the system should be able to identify finer differences between inputs, based on how knowledgeable it is about an environment. This is intended to parallel a human's attention. Here we may draw an analogy to how a person may appreciate a piece of artwork. At first glance they may only discern the principal subject of the picture, spending larger amounts of time on the task (receiving more stimuli) the person begins to appreciate the finer details. These finer details can only be represented by the system if σ is reduced. The system would first identify the major themes (whilst σ is large), before focussing on the finer details (when σ is small). We proposed a pair of functions that may meet this criteria. The first relates σ to time, the longer the system learns, the smaller the value of σ (down to a small positive constant). Alternatively, we may describe σ as a function of $V(t, x)$, relating lower values of $V(t, x)$, indicative of more focussed learning, to smaller values of σ . We should be aware that both of these methods may only be applicable when subsequent stimuli are uncorrelated.

Towards the latter parts of Chapter 2 we considered the shaping of the system when the stimuli is a multi-dimensional vector. Although the simplest stimuli are associated with a single string of input values, it is likely that a stimuli will normally consist of several components. As an example we may conceive of a circumstance where the model is utilized to classify different sounds. In this case each stimuli may be characterized by two components, one relating to volume and the other to pitch. We show that by applying a collection of simulated two-dimensional stimuli to a potential $V(t, x_1, x_2)$, the plane (x_1, x_2) is partitioned into basins of attraction relating to each fixed point of the system. Each fixed point is associated with a class of information stored by the model. A human learner may simultaneously interpret stimuli regarding sight, smell, sound and touch. From the perspective of our model, several dimensions may be required to characterize each of these faculties. The dimensionality of the system may be particularly high if temporal patterns of input are to be learnt and retained. It is clear that the dimensionality of the potential $V(t, \mathbf{x})$ may always be increased to accommodate the complexity of the stimuli.

In Chapter 3 we investigated the logical extension from a contrived, simulated stimuli, to learning in a more complex real world environment. This testing of the model is necessary to highlight the scope of the system's applicability. Here we presented an experiment that demonstrated how the system automatically forms classes when presented with three-dimensional RGB inputs. In this investigation RGB inputs define colours prescribed by a web-camera observing cartoons displayed on a LCD monitor. The system was shown to autonomously cluster and recognize similar colours. The classes formed by the system were again represented by attracting fixed points with basins partitioning an RGB space. We showed how unfamiliar stimuli, selected at regular time intervals, could be identified based on the information previously learnt by the system. Because the model learns from the application of subsequent inputs, the locations of fixed points are shown to migrate with time, new fixed points may emerge as new local minima of $V(t, R, G, B)$ are formed. Recognition is therefore a flexible process that changes depending on the current state of the potential. Similar stimuli that are encountered at different time intervals are not guaranteed to be recognized in the same way. This is not dissimilar to a human learner, as more knowledge about an environment is obtained, we are able to draw a greater distinction between inputs and highlight the finer details separating classes. It is clear that the model is only able to accurately recognize

stimuli that are closely related to the environment that it has observed. If the system experiences an environment bereft of a certain colour class, then subsequently tries to recognize a colour resembling this class, it is likely that the relationship between stimuli and recognized colour will be poor. Conversely, when colours all reside within a narrow bandwidth the system may become an expert in differentiating between subtle differences in shade and tone. This concept may be replicated in human populations. It would appear that different people perceive colours differently, or at least derive a different lexigraph with which to define observed colours as a function of their experiences [93]. Certainly some populations have been shown to demonstrate the ability to discern between shades more rapidly than other populations [164], suggesting that the way colour is perceived is different. This may be a result of continual exposure to an environment containing a narrow range of similar colours [32].

The main drawback of the three-dimensional representation of colour is the extensive amount of time taken to integrate the variables of the system. This prohibits the model from learning on a time-scale comparable to a human observer. In the second half of Chapter 3 we explored a method to map three-dimensional colours onto a one-dimensional spectrum. This meant that the inputs to the learning system became single values. The advantage of this approach is a substantial speed up in computation, however, it is shown that such mappings are incapable of representing a diverse range of colours. Such simplifications of the input may be applicable to certain special domains, however, for real world learning situations, the representation of colour by a one-dimensional spectrum appears to be overly simplistic.

In Chapters 4 and 5 we considered how the dynamics of the model may be manipulated to demonstrate higher order, non-linear attractors such as limit cycles and quasi-periodic orbits. As advocated by [99], the gradient type system is not the most biologically relevant or informative way to visualize recognition. An amendment to incorporate higher order attractors is instigated by describing recognition via a system derived from the van der Pol oscillator. The decision to model recognition via oscillatory dynamics is motivated by more than just the fact that oscillations are synonymous in nature. The basic fixed point attractor is able to purvey very little information about the topology of a class beyond the location of its centre. By representing a class as a limit cycle we are able to convey its features via both the position in the phase space and the frequency of oscillation. The prototype of the plastic system recognizes a stimulus by assigning it to a known class represented by a fixed point. It draws no relation between the information stored in different classes. It would be advantageous to illustrate the relation between different classes of the system within a hierarchical structure. As an analogy for this we may refer to our investigation of colour. We may consider the model to represent three principal classes; red, green and blue. As the system learns, it may start to appreciate other colours like pink; which it may consider as a sub-class of red, or azure; that may be considered a subclass of blue. If the system were to receive a stimulus that closely resembled the colour azure, we would like recognition to be characterized by an attractor that not only identifies the stimulus as azure, but also highlights that this is a subclass of blue. The attractors of the system should illustrate the relationships between different classes. This information cannot be conveyed by a fixed point. We postulate that higher order attractors such as limit cycles, quasi-periodic orbits and chaos may be able to characterize such relations. We speculate that a hierarchical framework may be represented by a series of different attractors. A fixed point may represent the lowest sub-class of the system, whilst a larger class may be defined by a limit cycle, the class above this may be characterized by a two-dimensional tori and a yet

larger class may be demonstrated by a three-dimensional tori. We should then consider a pyramid structure where each step of the pyramid is represented by a higher order tori. Eventually the increasing of tori dimensions may lead to a chaotic type attractor. The search for higher order attractors related to the plastic learning system is detailed in Chapters 4 and 5.

In Chapter 4 we studied methods to relate the potential $V(t, x)$ of the learning system to the van der Pol oscillator in order to represent the classes of system with limit cycle trajectories. We first considered replacing the potential term of the van der Pol system with $V(t, x)$, defined by the learning system. The single fixed point of the van der Pol system would then be replaced by an ensemble of fixed points determined by the distribution of stimuli. It was hypothesised that a Hopf bifurcation about any fixed point would lead to a unique limit cycle encapsulating that point. This notion was proven inaccurate by considering a VDP system with three fixed points. The phase space was shown to only contain a single limit cycle. This offered little scope for distinguishing between the different classes of the system. To overcome this we contemplated the parabolic shape indicative of the van der Pol system's potential. It is possible to represent the local topology of $V(t, x)$ about any point with a Taylor expansion and so model $V(t, x)$ as a series of parabolas. Each parabola may then define the potential of the van der Pol system in the region of a local minimum. Oscillatory trajectories are shown to cycle around a fixed point that migrates with time towards a minimum of $V(t, x)$. The frequency associated with this trajectory is dependent on the curvature of $V(t, x)$, frequency may therefore be considered to indicate the relative width of a class. We may explain the appeal of this utilizing an analogy to colours: The location of the attractor may indicate the colour, but the frequency of oscillation will highlight how wide this class is. A class may encapsulate the colours pink, red and orange, whilst the centre of the class corresponds to red. It is useful to know that the class also contains these other colours. More stimuli may cause the formation of three distinct classes, one relating to each colour. The frequency related to each cycle will indicate the reduction in class size.

Throughout Chapter 4 we focussed on the basic case where stimuli are one-dimensional. Increasing their dimensionality adds complexity as we have to consider how the different dimensions of the system should interact to describe recognition. In Chapter 5 we considered the case where the stimuli are two-dimensional, before starting to extend our analysis to a three-dimensional stimulus. For the two-dimensional case, stimuli are associated with the characteristic directions x_1 and x_2 . We made the logical extension to associate each dimension with its own oscillator. A stimulus $\eta = (\eta_1, \eta_2)$ is therefore associated with a pair of oscillators, that demonstrate properties dependent on $V(t, \mathbf{x})$. The frequency of each oscillator is then dependent on the curvature at a minimum of $V(t, \mathbf{x})$ in the characteristic directions x_1 and x_2 . Referring to our requirement that different classes within the phase space are associated with different attractors, we showed that limit cycles, quasi-periodicity and oscillation death may all be present when oscillators are dissipatively coupled. The behaviour of the system was shown to depend on the curvature of the potential and on the relationship between the coupling strength B_D and non-linearity parameter ε . The existence of different attractors is an important ingredient to create a hierarchical learning framework. We note that the bifurcation separating the oscillation death region from the quasi-periodic region is controlled by ε and B_D . As we would like to represent different classes by either quasi-periodicity or oscillation death, we proposed two functions that relate ε to relevant aspects of a classes topology.

- ε may be related to the depth of $V(t, \mathbf{x})$. We may consider that a deeper local minimum of $V(t, x_1, x_2)$ is associated with a narrower range of stimuli. When the minimum is deep we associate this with oscillation death. A deep well may infer that the class is very well defined and so the fixed point attractor may represent the certainty that the system has about identifying the stimulus with this class.
- ε may be related to the level of similarity between a stimulus and the class that it is assigned to. The parameter ε is scaled to the distance between the centre of the class and the initial condition of the stimulus. Where this is large, the system is considered to vaguely recognize this stimulus, this is represented by oscillation death, where the distance between stimulus and centre of class is small, the stimulus is considered to be a good match for the class and this relation is identified by a quasi-periodic orbit.

It should be noted that neither of these functions exclude the emergence of a limit cycle. In each case this remains dependent on the similarity between the curvatures in each characteristic dimension. The effect of manipulating ε via either of these approaches is still observed for a limit cycle trajectory, as the adjustment of the parameter has implications on the topology of the cycle. The primary achievement that should be emphasised is that an attractor representing a class is always dependent on the distribution of stimuli. We can therefore appreciate the knowledge retained by the system based on the recognition of a stimulus.

The implications of increasing the dimensionality of the stimulus to three or more dimensions are not yet fully understood. In the latter half of Chapter 5 we began to develop an appreciation of the dynamics of three coupled oscillators as this presents the logical next step from the two-dimensional stimulus consideration. Three oscillators would be required to represent the recognition of classes built using a three-dimensional stimuli. Before we may relate the learning system to a set of three oscillators it is important to understand the dynamics of such systems. The theory relating to three mutually coupled oscillators is much less developed than for a pair of oscillators. We therefore began by considering the bifurcation structure of the system. This is necessary so that we may select appropriate system parameters that permit informative bifurcations. It is also important to know the range of attractors that may be permitted. The existence of three-dimensional tori, two-dimensional tori, limit cycles and oscillation death are all demonstrated. As our analysis remains incomplete, we cannot exclude the possibility of other attractors appearing for certain parameter ranges. Here we focussed on the bifurcations that occurred as ω_1 , ω_2 and ω_3 were varied. These are the three natural frequencies describing the curvature in the three characteristic directions x_1 , x_2 and x_3 . Despite our understanding of such systems being incomplete, we demonstrated the emergence of different attractors corresponding to the coloured RGB examples discussed in chapter 3.

It is clear that further investigation is required to adequately demonstrate the effect that increasing the dimensionality has on this new model of recognition. The extension of this approach to consider an N-dimensional stimuli would require the coupling of N oscillators and at this stage it is not clear how these will behave. Based on our analysis of 1-3 coupled oscillators we may speculate that the system's parameter space will permit the existence of fixed points, limit cycles and an array of higher-order tori whose dimension will be up to N-1. Future developments of this system should focus on exploring these higher-order attractors and should consider how they may be related to a theory of hierarchy. It is clear that the computational demands for high-dimensional

systems will be expensive. It would appear that dedicated hardware may be required to assimilate these stimuli in real-time. This presents a difficult engineering challenge as the system is very different from the architectures currently utilized within AI. Nevertheless, an engineering solution may be required if this new model is to gain traction within the field.

In summary, we have explored a new type of self-shaping system that completely adjusts its vector field in response to an applied stimuli. Such classes of dynamical system have thus far received little attention. This dissertation is one of the first attempts to systematically study a system of this type. The new model is shown to simultaneously classify and recognize inputs, without the influence of external supervision. We began our investigation by contrasting this method to the mainstream approaches found within the AI field. We indicated the advantages of this new paradigm, and tested the prototype system for a range of complex stimuli including RGB inputs obtained experimentally utilizing a web-camera. After highlighting the deficiencies of gradient descent, we studied the van der Pol oscillator as a basis to develop an alternative representation of recognition, reliant on the concepts of non-linear dynamics. This approach was motivated by a desire to create a hierarchy which would highlight the associations between different classes of the system. Finally we considered the extension of the model when the potential, characterising the memory of the system, is multi-dimensional. Utilizing the concepts of synchronization and coupling it was shown that the classes of the system can be characterized by a number of unique phase-space attractors. Representing a class of the system by a higher-order attractor is shown to permit a more complete description of the properties of a class.

Appendix A

Appendix

A.1 Runge-Kutta Numerical Integration

The Runge-Kutta method has been shown to numerically provide more accurate solutions to differential equations than many other approaches. This is particularly apparent if we consider it against the commonly utilized Euler Method [150]. Differential equations are solved by computing the next iterative step based on 4 estimates of slope between subsequent time units. These slope estimates are then weighted and used to determine the average slope [150]. The Runge-kutta method for the differential equation:

$$\dot{x} = F(t, x(t)) \quad x(t_0) = x_0 \quad (\text{A.1})$$

Takes the form:

$$\begin{aligned} K_1 &= h.F(t_n, x_n) \\ K_2 &= h.F(t_n + (1/2)h, x_n + (1/2)K_1) \\ K_3 &= h.F(t_n + (1/2)h, x_n + (1/2)K_2) \\ K_4 &= h.F(t_n + h, x_n + K_3) \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} t_{n+1} &= t_n + h \\ x_{n+1} &= x_n + 1/6(K_1 + 2K_2 + 2K_3 + K_4) \end{aligned}$$

Each of the K values is a measure of the slope between x_n and x_{n+1} at different points on this interval. The variable h is referred to as the integration step size, to aid accuracy a small value should be chosen. Note that the differential equation has been transformed into a difference equations. The errors in the Runge-Kutta values are of the order h^5 per step and h^4 cumulatively.

A.2 The Box-Muller Transformation

There are several ways to transform numerically generated, uniformly distributed, random numbers into Gaussian distributed random numbers. These include the inverse transform sampling method [108] and the Ziggurat algorithm [98]. Possibly the most well used method is the Box-Muller Transform [16]. This appears in two forms. The first is perhaps the most basic:

Given uniform deviates U_1 and U_2 on the interval $(0, 1)$:

$$\begin{aligned} z_0 &= \sqrt{-2\ln U_1} \cos(2\pi U_2) \\ z_1 &= \sqrt{-2\ln U_1} \sin(2\pi U_2) \end{aligned} \tag{A.3}$$

Where z_0 and z_1 are Gaussian distributed random numbers.

The second method uses polar co-ordinates:

Given that U_1 and U_2 are uniformly distributed random number on the closed interval $(-1, 1)$:

$$\begin{aligned} z_0 &= \sqrt{-2\ln R^2} \left(\frac{U_1}{R} \right) = U_1 \sqrt{\frac{-2\ln R^2}{R^2}} \\ z_1 &= \sqrt{-2\ln R^2} \left(\frac{U_2}{R} \right) = U_2 \sqrt{\frac{-2\ln R^2}{R^2}} \end{aligned} \tag{A.4}$$

Here $R = \sqrt{-2\ln U_1}$. Again, z_0 and z_1 are Gaussian distributed random numbers.

Both of these methods require the input of pairs of uniformly distributed random numbers and return comparable Gaussian distributed values. The difference in uniform input ranges for the two methods should be noted. For computer simulation the polar form of the Box-Muller transform is less computationally demanding as it does not require the use of trigonometric functions. The polar form differs from the basic method in that it is a type of rejection sampling, typically ‘throwing away’ $1 - \pi/4 \approx 21.46\%$ of the total input uniformly distributed random number pairs [128].

A.3 Realizations of Colour Learning

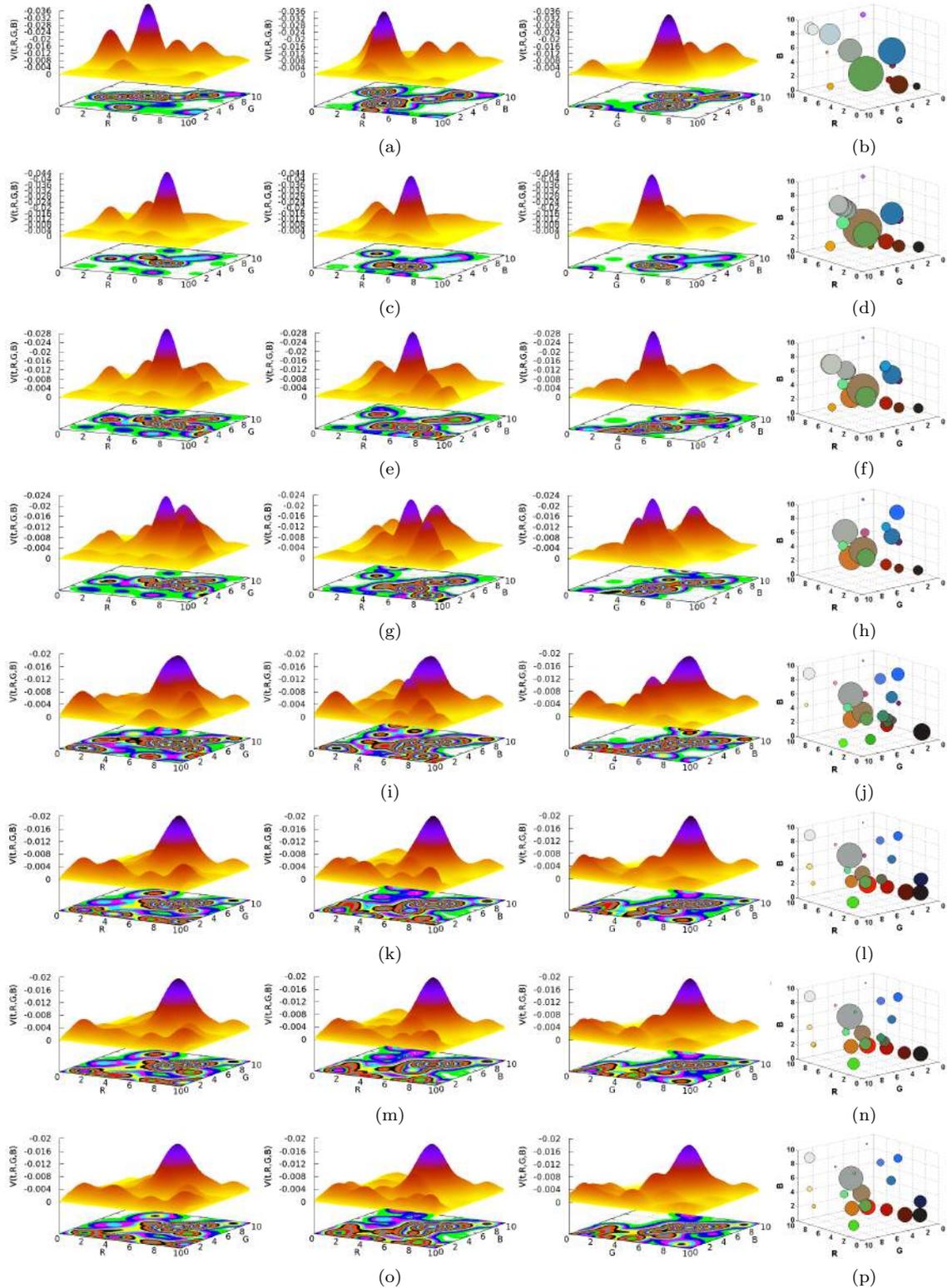


Figure A.1: Each row corresponds to a different time moment. The stimulus is equivalent to that utilized in figure 3.8 for the cartoon Southpark. Here we demonstrate the profile at much earlier time instances. Each row depicts three projections of the potential and highlights the locations of the local minima. The colour of a minima relates to its location within the RGB space whilst the size of the point relates to the depth of the well. The rows correspond to the time moments $t = 5, 10, 15, 20, 40, 80, 160, 200$. As with the previous demonstration for this stimulus $\sigma = \sqrt{0.5}$.

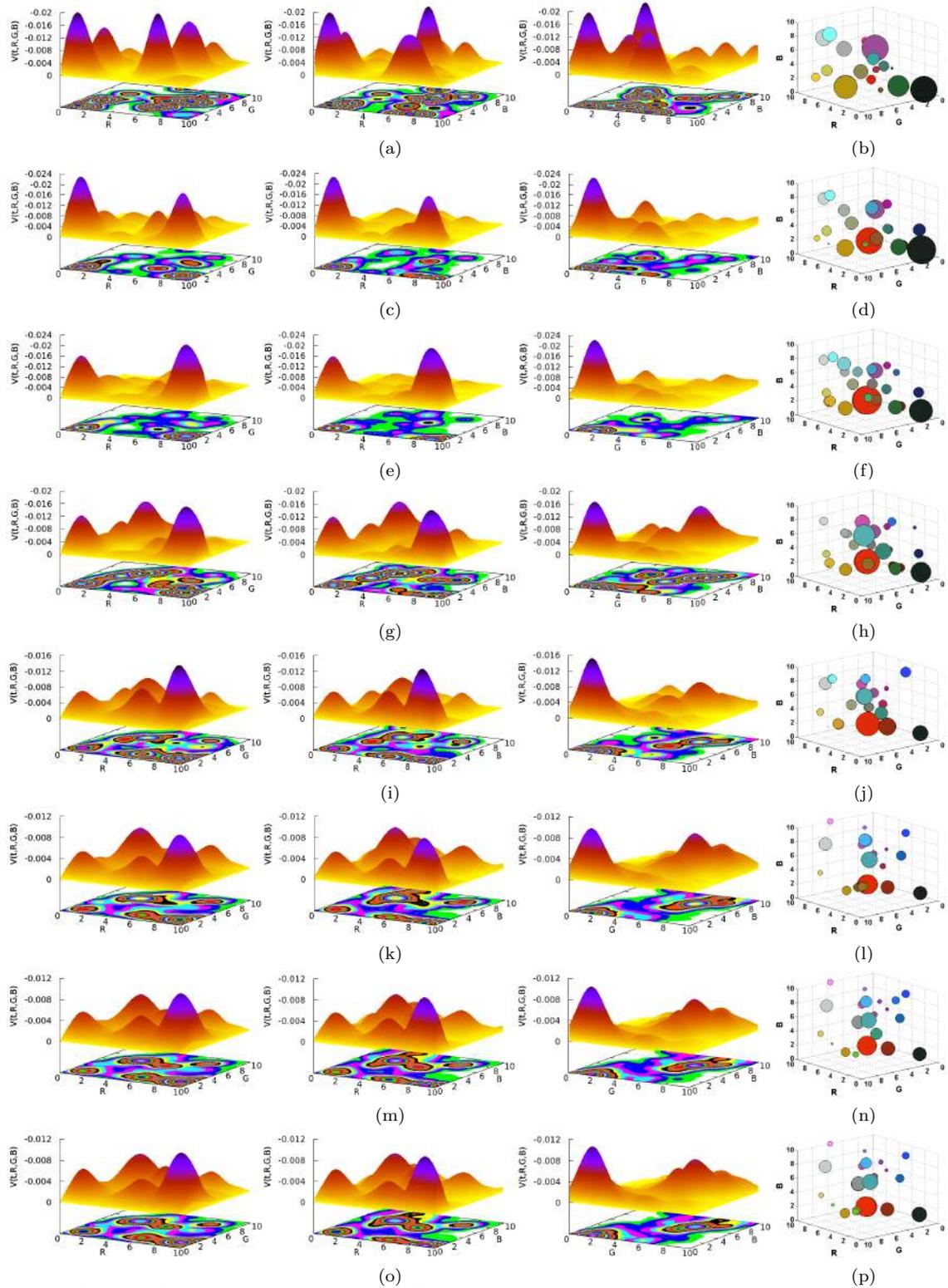


Figure A.2: Each row corresponds to a different time moment. The stimulus is equivalent to that utilized in figure 3.11 for the cartoon Futurama. Here we demonstrate the profile at much earlier time instances. Each row depicts three projections of the potential and highlights the locations of the local minima. The colour of a minimum relates to its location within the RGB space whilst the size of the point relates to the depth of the well. The rows correspond to the time moments $t = 5, 10, 15, 20, 40, 80, 160, 200$. As with the previous demonstration for this stimulus $\sigma = \sqrt{0.5}$.

A.4 The Truncated Equations for a Pair of Coupled van der Pol Oscillators

Starting from the differential equations:

$$\begin{aligned}\ddot{x}_1 - \varepsilon_1 (1 - x_1^2) \dot{x}_1 + \omega_1^2 x_1 + B_R (x_1 - x_2) + B_D (\dot{x}_1 - \dot{x}_2) \\ \ddot{x}_2 - \varepsilon_2 (1 - x_2^2) \dot{x}_2 + \omega_2^2 x_2 + B_R (x_2 - x_1) + B_D (\dot{x}_2 - \dot{x}_1)\end{aligned}\quad (\text{A.5})$$

We begin by looking for solutions of the form:

$$x_{1,2} = A_{1,2}(t) \cos(\omega t + \varphi_{1,2}(t)) = \frac{1}{2} (a_{1,2}(t) e^{i\omega t} + a_{1,2}^*(t) e^{-i\omega t}) \quad (\text{A.6})$$

Here $A_{1,2}$ is the amplitude, ω is the frequency and $\varphi_{1,2}$ are the relative phases of oscillation. a_1 and a_2 are complex amplitudes, these form complex conjugate pairs with $a_{1,2}^*$. The complex amplitudes are related to the real amplitudes by: $a_{1,2} = A_{1,2} e^{i\varphi_{1,2}}$ and $a_{1,2}^* = A_{1,2} e^{-i\varphi_{1,2}}$. Ambiguity is introduced by the phase variables $A_{1,2}(t)$ and $\varphi_{1,2}(t)$. This can be overcome if we introduce the condition that $\dot{x}_{1,2} = -A_{1,2}\omega \sin(\omega t + \varphi_{1,2})$. Here we have dropped the notation of time dependence, however, we should stress that is purely for succinctness. Taking the derivative of equation (A.6) with respect to time:

$$\begin{aligned}\dot{x}_{1,2} &= \dot{A}_{1,2} \cos(\omega t + \varphi_{1,2}) - A_{1,2} \omega \sin(\omega t + \varphi_{1,2}) - A_{1,2} \dot{\varphi}_{1,2} \sin(\omega t + \varphi_{1,2}) \\ &= \frac{1}{2} (\dot{a}_{1,2} e^{i\omega t} + a_{1,2} i\omega e^{i\omega t} + \dot{a}_{1,2}^* e^{-i\omega t} - a_{1,2}^* i\omega e^{-i\omega t})\end{aligned}\quad (\text{A.7})$$

and including the condition $\dot{x}_{1,2} = -A_{1,2}\omega \sin(\omega t + \varphi_{1,2})$:

$$\dot{A}_{1,2} \cos(\omega t + \varphi_{1,2}) - A_{1,2} \dot{\varphi}_{1,2} \sin(\omega t + \varphi_{1,2}) = \dot{a}_{1,2} e^{i\omega t} + \dot{a}_{1,2}^* e^{-i\omega t} = 0 \quad (\text{A.8})$$

and so:

$$\begin{aligned}\dot{x}_{1,2} &= \frac{i\omega}{2} (a_{1,2} e^{i\omega t} - a_{1,2}^* e^{-i\omega t}) \\ \ddot{x}_{1,2} &= \frac{i\omega}{2} (\dot{a}_{1,2} e^{i\omega t} + a_{1,2} i\omega e^{i\omega t} - \dot{a}_{1,2}^* e^{-i\omega t} + a_{1,2}^* i\omega e^{-i\omega t}) \\ &= \frac{i\omega}{2} (\dot{a}_{1,2} e^{i\omega t} - \dot{a}_{1,2}^* e^{-i\omega t}) - \frac{\omega^2}{2} (a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t}) \\ &= i\omega \dot{a}_{1,2} e^{i\omega t} - \frac{\omega^2}{2} (a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t})\end{aligned}\quad (\text{A.9})$$

Substituting these derivative into equation (A.5) the van der Pol system can be described by:

$$\begin{aligned}i\omega \dot{a}_{1,2} - \frac{\omega^2}{2} (a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t}) - \varepsilon_{1,2} \frac{i\omega}{2} (a_{1,2} e^{i\omega t} - a_{1,2}^* e^{-i\omega t}) + \\ \varepsilon \frac{i\omega}{2} (a_{1,2} e^{i\omega t} - a_{1,2}^* e^{-i\omega t}) \left(\frac{1}{2} (a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t}) \right)^2 + \frac{\omega_{1,2}^2}{2} (a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t}) \\ = \frac{B_R}{2} (a_{2,1} e^{i\omega t} + a_{2,1}^* e^{-i\omega t} - a_{1,2} e^{i\omega t} - a_{1,2}^* e^{-i\omega t}) \\ + \frac{B_D i\omega}{2} (a_{2,1} e^{i\omega t} - a_{2,1}^* e^{-i\omega t} - a_{1,2} e^{i\omega t} + a_{1,2}^* e^{-i\omega t})\end{aligned}\quad (\text{A.10})$$

Subjected to some manipulation and multiplying through by $\frac{e^{i\omega t}}{i\omega}$ this becomes:

$$\begin{aligned}
& \dot{a}_{1,2} + \left(\frac{\omega_{1,2}^2 - \omega^2}{2i\omega} \right) (a_{1,2} + a_{1,2}^* e^{-2i\omega t}) \\
& - \frac{\varepsilon_{1,2}}{2} \left(a_{1,2} - a_{1,2}^* e^{-2i\omega t} - \frac{a_{1,2}^3}{4} e^{2i\omega t} - \frac{a_{1,2}^* a_{1,2}^2}{4} + \frac{a_{1,2}}{4} (a_{1,2}^*)^2 e^{-2i\omega t} + \frac{(a_{1,2}^*)^3}{4} e^{-4i\omega t} \right) \\
& = \frac{B_R}{2i\omega} (a_{2,1} + a_{2,1}^* e^{-2i\omega t} - a_{1,2} - a_{1,2}^* e^{-2i\omega t}) + \frac{B_D}{2} (a_{2,1} - a_{2,1} e^{-2i\omega t} - a_{1,2} + a_{1,2}^* e^{-2i\omega t})
\end{aligned} \tag{A.11}$$

We may assume that $a_{1,2}$ are slow functions of time. Taking this into consideration we may apply the Krylov-Bogolubov method [85] and take an average over the period T . This ensures that the average value of terms containing $e^{-i2\omega t}$, $e^{i2\omega t}$ and $e^{-i4\omega t}$ equals zero. Hence

$$\dot{a}_{1,2} - \left(\frac{\omega_{1,2}^2 - \omega^2}{2\omega} \right) i a_{1,2} - \frac{\varepsilon_{1,2}}{2} \left(a_{1,2} - \frac{a_{1,2}^*}{4} a_{1,2}^2 \right) = \left(\frac{B_D}{2} - \frac{B_R i}{2\omega} \right) (a_{2,1} - a_{1,2}) \tag{A.12}$$

If we then represent this in terms of real amplitudes $A_{1,2}$:

$$\begin{aligned}
& \dot{A}_{1,2} e^{i\varphi_{1,2}} + A_{1,2} i \dot{\varphi}_{1,2} e^{i\varphi_{1,2}} - \left(\frac{\omega_{1,2}^2 - \omega^2}{2\omega} \right) i A_{1,2} e^{i\varphi_{1,2}} - \frac{\varepsilon_{1,2}}{2} \left(A_{1,2} e^{i\varphi_{1,2}} - \frac{A_{1,2}^3 e^{i\varphi_{1,2}}}{4} \right) \\
& = \left(\frac{B_D}{2} - \frac{B_R i}{\omega} \right) (A_{2,1} e^{i\varphi_{2,1}} - A_{1,2} e^{i\varphi_{1,2}})
\end{aligned} \tag{A.13}$$

Dividing through by $e^{i\varphi_{1,2}}$:

$$\dot{A}_{1,2} + A_{1,2} i \dot{\varphi}_{1,2} - \left(\frac{\omega_{1,2}^2 - \omega^2}{2\omega} \right) i A_{1,2} - \frac{\varepsilon_{1,2}}{2} \left(A_{1,2} - \frac{A_{1,2}^3}{4} \right) = \left(\frac{B_D}{2} - \frac{B_R i}{2\omega} \right) (A_{2,1} e^{i(\varphi_{2,1} - \varphi_{1,2})} - A_{1,2}) \tag{A.14}$$

This can then be re-written in terms of the functions *cos* and *sin* utilizing Euler formula:

$$\begin{aligned}
\dot{A}_{1,2} + A_{1,2} i \dot{\varphi}_{1,2} & = \left(\frac{\omega_{1,2}^2 - \omega^2}{2\omega} \right) i A_{1,2} + \frac{\varepsilon_{1,2}}{2} \left(A_{1,2} - \frac{A_{1,2}^3}{4} \right) \\
& + \left(\frac{B_D}{2} - \frac{B_R i}{2\omega} \right) (A_{2,1} (\cos(\varphi_{2,1} - \varphi_{1,2}) + i \sin(\varphi_{2,1} - \varphi_{1,2})) - A_{1,2})
\end{aligned} \tag{A.15}$$

Separating real and imaginary parts:

$$\begin{aligned}
\dot{A}_1 & = \frac{\varepsilon_1}{2} \left(A_1 - \frac{A_1^3}{4} \right) + \frac{B_D}{2} (A_2 \cos(\varphi_2 - \varphi_1) - A_1) + \frac{B_R}{2\omega} A_2 \sin(\varphi_2 - \varphi_1) \\
\dot{\varphi}_1 & = \frac{\omega_1^2 - \omega^2}{2\omega} + \frac{B_D}{2} \frac{A_2}{A_1} \sin(\varphi_2 - \varphi_1) - \frac{B_R}{2\omega} \frac{A_2}{A_1} \cos(\varphi_2 - \varphi_1) + \frac{B_R}{2\omega} \\
\dot{A}_2 & = \frac{\varepsilon_2}{2} \left(A_2 - \frac{A_2^3}{4} \right) + \frac{B_D}{2} (A_2 \cos(\varphi_1 - \varphi_2) - A_2) + \frac{B_R}{2\omega} A_1 \sin(\varphi_1 - \varphi_2) \\
\dot{\varphi}_2 & = \frac{\omega_2^2 - \omega^2}{2\omega} + \frac{B_D}{2} \frac{A_1}{A_2} \sin(\varphi_1 - \varphi_2) - \frac{B_R}{2\omega} \frac{A_1}{A_2} \cos(\varphi_1 - \varphi_2) + \frac{B_R}{2\omega}
\end{aligned} \tag{A.16}$$

Finally introducing $\theta = \varphi_2 - \varphi_1$ and $\Delta = \frac{\omega_2^2 - \omega_1^2}{2\omega}$ (where $\omega \approx \omega_{1,2}$) we may define the truncated equations for the van der Pol system as:

$$\begin{aligned}\dot{A}_1 &= \frac{\varepsilon_1}{2} \left(A_1 - \frac{A_1^3}{4} \right) + \frac{B_D}{2} (A_2 \cos(\theta) - A_1) + \frac{B_R}{2\omega} A_2 \sin(\theta) \\ \dot{A}_2 &= \frac{\varepsilon_2}{2} \left(A_2 - \frac{A_2^3}{4} \right) + \frac{B_D}{2} (A_1 \cos(\theta) - A_2) - \frac{B_R}{2\omega} A_1 \sin(\theta) \\ \dot{\theta} &= \Delta - \frac{B_D}{2} \sin(\theta) \left(\frac{A_2}{A_1} + \frac{A_1}{A_2} \right) + \frac{B_R}{2\omega} \cos(\theta) \left(\frac{A_2}{A_1} - \frac{A_1}{A_2} \right)\end{aligned}\tag{A.17}$$

A.5 The Truncated Equations for Three Coupled van der Pol Oscillators

For a set of three dissipatively coupled oscillators the system may be described by the differential equation:

$$\ddot{x}_{1,2,3} - \varepsilon_{1,2,3} (1 - x_{1,2,3}^2) \dot{x}_{1,2,3} + \omega_{1,2,3}^2 x_{1,2,3} + B_D ((\dot{x}_{1,2,3} - \dot{x}_{2,3,1}) + (\dot{x}_{1,2,3} - \dot{x}_{3,1,2})) = 0\tag{A.18}$$

As presented in section A.4 we may introduce the complex amplitude and define:

$$\begin{aligned}x_{1,2,3} &= \frac{1}{2} (a_{1,2,3} e^{i\omega t} + a_{1,2,3}^* e^{-i\omega t}) \\ \dot{x}_{1,2,3} &= \frac{i\omega}{2} (a_{1,2,3} e^{i\omega t} - a_{1,2,3}^* e^{-i\omega t}) \\ \ddot{x}_{1,2,3} &= i\omega \dot{a}_{1,2,3} e^{i\omega t} - \frac{\omega^2}{2} (a_{1,2,3} e^{i\omega t} + a_{1,2,3}^* e^{-i\omega t})\end{aligned}\tag{A.19}$$

Inserting these functions into equation (A.18) and manipulating terms the differential equation may take the form:

$$\begin{aligned}\dot{a}_{1,2,3} + \left(\frac{\omega_{1,2,3}^2 - \omega^2}{2i\omega} \right) (a_{1,2,3} + a_{1,2,3}^* e^{-2i\omega t}) \\ + \frac{\varepsilon_{1,2,3}}{2} \left(\frac{a_{1,2,3}^3}{4} e^{2i\omega t} + \frac{a_{1,2,3}^2 a_{1,2,3}^*}{4} - \frac{a_{1,2,3} a_{1,2,3}^{*2}}{4} e^{-2i\omega t} - \frac{a_{1,2,3}^{*3}}{4} e^{-4i\omega t} + a_{1,2,3}^* e^{-2i\omega t} - a_{1,2,3} \right) \\ + \frac{B_D}{2} (2a_{1,2,3} - 2a_{1,2,3}^* e^{-2i\omega t} - a_{2,3,1} + a_{2,3,1}^* e^{-2i\omega t} - a_{3,1,2} + a_{3,1,2}^* e^{-2i\omega t}) = 0\end{aligned}\tag{A.20}$$

Then applying the Krylov-Bogoliubov method of averaging:

$$\begin{aligned}\dot{a}_{1,2,3} - \left(\frac{\omega_{1,2,3}^2 - \omega^2}{2\omega} \right) i a_{1,2,3} + \frac{\varepsilon_{1,2,3}}{2} \left(\frac{a_{1,2,3}^2 a_{1,2,3}^*}{4} - a_{1,2,3} \right) \\ + \frac{B_D}{2} (2a_{1,2,3} - a_{2,3,1} - a_{3,1,2}) = 0\end{aligned}\tag{A.21}$$

Expressing this in terms of real amplitude $A_{1,2,3}$ and phase $\varphi_{1,2,3}$ (note that $a_{1,2,3} = A_{1,2,3}e^{i\varphi_{1,2,3}}$ and $a_{1,2,3}^* = A_{1,2,3}e^{-i\varphi_{1,2,3}}$) we arrive at:

$$\begin{aligned} & \dot{A}_{1,2,3}e^{i\varphi_{1,2,3}} + A_{1,2,3}i\dot{\varphi}_{1,2,3}e^{i\varphi_{1,2,3}} - \left(\frac{\omega_{1,2,3}^2 - \omega^2}{2\omega}\right) iA_{1,2,3}e^{i\varphi_{1,2,3}} + \frac{\varepsilon_{1,2,3}}{2} \left(\frac{A_{1,2,3}^3}{4}e^{i\varphi_{1,2,3}} - A_{1,2,3}e^{i\varphi_{1,2,3}}\right) \\ & + \frac{B_D}{2} (2A_{1,2,3}e^{i\varphi_{1,2,3}} - A_{2,3,1}e^{i\varphi_{2,3,1}} - A_{3,1,2}e^{i\varphi_{3,1,2}}) = 0 \end{aligned} \quad (\text{A.22})$$

Dividing equation (A.22) through by $e^{i\varphi_{1,2,3}}$ and rewriting in terms of sine and cosine:

$$\begin{aligned} & \dot{A}_{1,2,3} + A_{1,2,3}i\dot{\varphi}_{1,2,3} - \left(\frac{\omega_{1,2,3}^2 - \omega^2}{2\omega}\right) iA_{1,2,3} + \frac{\varepsilon_{1,2,3}}{2} \left(\frac{A_{1,2,3}^3}{4} - A_{1,2,3}\right) \\ & + \frac{B_D}{2} (2A_{1,2,3} - A_{2,3,1}(\cos(\varphi_{2,3,1} - \varphi_{1,2,3}) + i\sin(\varphi_{2,3,1} - \varphi_{1,2,3}))) \\ & - \frac{B_D}{2} (A_{3,1,2}(\cos(\varphi_{3,1,2} - \varphi_{1,2,3}) + i\sin(\varphi_{3,1,2} - \varphi_{1,2,3}))) = 0 \end{aligned} \quad (\text{A.23})$$

If we then separate real and imaginary parts:

$$\begin{aligned} \dot{A}_1 &= \frac{\varepsilon_1}{2} \left(A_1 - \frac{A_1^3}{4}\right) - \frac{B_D}{2} (2A_1 - A_2\cos(\varphi_2 - \varphi_1) - A_3(\varphi_3 - \varphi_1)) \\ \dot{\varphi}_1 &= \left(\frac{\omega_1^2 - \omega^2}{2\omega}\right) + \frac{B_D}{2} \left(\frac{A_2}{A_1}\sin(\varphi_2 - \varphi_1) + \frac{A_3}{A_1}\sin(\varphi_3 - \varphi_1)\right) \\ \dot{A}_2 &= \frac{\varepsilon_2}{2} \left(A_2 - \frac{A_2^3}{4}\right) - \frac{B_D}{2} (2A_2 - A_3\cos(\varphi_3 - \varphi_2) - A_1(\varphi_1 - \varphi_2)) \\ \dot{\varphi}_2 &= \left(\frac{\omega_2^2 - \omega^2}{2\omega}\right) + \frac{B_D}{2} \left(\frac{A_3}{A_2}\sin(\varphi_3 - \varphi_2) + \frac{A_1}{A_2}\sin(\varphi_1 - \varphi_2)\right) \\ \dot{A}_3 &= \frac{\varepsilon_3}{2} \left(A_3 - \frac{A_3^3}{4}\right) - \frac{B_D}{2} (2A_3 - A_1\cos(\varphi_1 - \varphi_3) - A_2(\varphi_2 - \varphi_3)) \\ \dot{\varphi}_3 &= \left(\frac{\omega_3^2 - \omega^2}{2\omega}\right) + \frac{B_D}{2} \left(\frac{A_1}{A_3}\sin(\varphi_1 - \varphi_3) + \frac{A_2}{A_3}\sin(\varphi_2 - \varphi_3)\right) \end{aligned} \quad (\text{A.24})$$

Introducing the phase differences $\theta_A = \varphi_1 - \varphi_2$, $\theta_B = \varphi_1 - \varphi_3$ and $\theta_C = \varphi_2 - \varphi_3$ these equations may be written in the more amenable form:

$$\begin{aligned} \dot{A}_1 &= \frac{\varepsilon_1}{2} \left(A_1 - \frac{A_1^3}{4}\right) - \frac{B_D}{2} (2A_1 - A_2\cos(\theta_A) - A_3\cos(\theta_B)) \\ \dot{A}_2 &= \frac{\varepsilon_2}{2} \left(A_2 - \frac{A_2^3}{4}\right) - \frac{B_D}{2} (2A_2 - A_3\cos(\theta_C) - A_1\cos(\theta_A)) \\ \dot{A}_3 &= \frac{\varepsilon_3}{2} \left(A_3 - \frac{A_3^3}{4}\right) - \frac{B_D}{2} (2A_3 - A_1\cos(\theta_B) - A_2\cos(\theta_C)) \\ \dot{\theta}_A &= \frac{\omega_1^2 - \omega_2^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_2}{A_1} + \frac{A_1}{A_2}\right)\sin(\theta_A) + \frac{A_3}{A_1}\sin(\theta_B) - \frac{A_3}{A_2}\sin(\theta_C)\right) \\ \dot{\theta}_B &= \frac{\omega_1^2 - \omega_3^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_3}{A_1} + \frac{A_1}{A_3}\right)\sin(\theta_B) + \frac{A_2}{A_1}\sin(\theta_A) + \frac{A_2}{A_3}\sin(\theta_C)\right) \\ \dot{\theta}_C &= \frac{\omega_2^2 - \omega_3^2}{2\omega} - \frac{B_D}{2} \left(\left(\frac{A_3}{A_2} + \frac{A_2}{A_3}\right)\sin(\theta_C) - \frac{A_1}{A_2}\sin(\theta_A) + \frac{A_1}{A_3}\sin(\theta_B)\right) \end{aligned} \quad (\text{A.25})$$

Colour Recognition and Classification by a Dynamical System with Self-adapting Vector Field

Department of Physics, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom

February 2, 2015

Abstract

As a contrasting approach to the Neural Network description of learning we present a dynamical system modelled by differential equations capable of performing the primary tasks of artificial intelligence (classification, retention, recognition). This work draws upon the concepts introduced by Janson [arXiv:1107.0674 (2011)]. The system shapes its phase velocity vector field in response to the received inputs, classifying stimuli in an unsupervised manner. Recognition of stimuli occurs in parallel to class formation. A 3-dimensional phase velocity vector field is utilized to demonstrate the system's response to stimuli received from a web-camera. The model is shown to be able to classify and recognize colours using the RGB colour classification system.

1 Introduction

Existing approaches to AI are diverse, broadly classified into systems seeking to emulate human behaviour or thinking patterns and those that attempt to prescribe optimal solutions [18]. The justification of models based on brain structure need little validation as nature has proven that such an approach is viable. Artificial Neural Networks (NN) have become a cornerstone of AI research but have not yet yielded the breakthroughs that many researchers postulated at their inception. Other learning constructs such as Decision Tree Learning [18], Association Rule Learning [2], Reinforcement Learning [3], Support Vector Machines [21] and Genetic Algorithms [13] have met with limited success however each approach is restricted to certain well-defined classes of problems. A machine capable of emulating human levels of intelligence appears to be far from realizable.

Due to a lack of biological knowledge NN research takes a somewhat trial and error approach. Interactions between neural computational units and electrochemical signals are poorly understood. Furthermore how memory and learning is inferred is a point of debate. New insights may yet be uncovered by the ongoing research of the European 'Human Brain Project' [1] and the American "BRAIN" Initiative' [14].

The current research focus on Recurrent NN's and deep feed-forward NN's attempts to alleviate some of the difficulties of classical NN models [7]. High levels of memory redundancy required to avoid spurious attractors [6] and the related 'curse of dimensionality' remain principal problems. Although silicon microchip manufacturing continues to push the frontiers of scale and performance; capable of modelling 1 million neurons and 256 million

synapses [12], the production of neural chips with 'brain-like' capacity; 120 billion neurons and 100 trillion connections [5] is somewhat beyond the scope of modern technologies. Designed for specific tasks most NN's cannot be re-purposed while retaining previous knowledge, more resilient models such as Bayesian HTP [20] display better retention however rely on reinforcement [3] or supervised learning.

With existing AI models failing to meet expectations new machine learning methods should be pursued. In [8] Janson describes a new non-algorithmic, unsupervised system that retains knowledge via a potential energy function and identifies stimuli utilizing its vector field. Unlike existing machine learning routines, training phase is not distinct, instead learning and recognition occur 'on-line'. Devoid of rigid network architecture this autonomous system should be exempt from the scaling problems faced by other systems, offering an alternative paradigm to model learning. Here we consider a problem of machine vision and demonstrate how this new learning paradigm can be applied.

The ability to distinguish between different colours within a machine vision framework is limited, firstly by the hardware used and then by the algorithms applied. Colour recognition has been attempted with NN's in the context of minimizing colour matching error in a supervised scheme [9]. In addition NN have been used to smooth fuzzy colour data [10].

We shall show how Janson's model can be applied to stimuli in the form of RGB colour triplets taken from a real-world environment. The system uses these stimuli to form classes of retained information and identifies new stimuli based on what it has learnt. A complete explanation of Janson's model is presented in [8] and [11], here

we shall only state the critical features.

Model

A time-dependant evolving profile $\mathbf{V}(t, \mathbf{x})$ is continuously shaped by a stimuli $\boldsymbol{\eta} \in \mathbf{x}$. This shaping utilizes the stacking properties of the Gaussian function to ensure smoothness and continuity. $\mathbf{V}(t, \mathbf{x})$ contains all the information pertaining to the classification of stimuli, its evolution is described by differential equation 1a. Biological systems tend to forget information that is not regularly re-enforced. This idea is introduced by the term $-k\mathbf{V}(t, \mathbf{x})$. By manipulating the constant k we control the rate at which local minima disappear. Local minima of the profile represent retained knowledge. In parallel to the formation of the landscape $\mathbf{V}(t, \mathbf{x})$ recognition can be characterized by the gradient of the energy profile function. Any stimuli will be assigned to a class as prescribed by equation 1c, γ reflects the rate at which a stimuli is recognized and $\xi(t)$ is a source of noise.

$$\frac{\partial \mathbf{V}(t, \mathbf{x})}{\partial t} = -\frac{1}{t}(\mathbf{V}(t, \mathbf{x}) + g(\mathbf{x} - \boldsymbol{\eta})) - k\mathbf{V}(t, \mathbf{x}) \quad (1a)$$

$$g(z) = \frac{1}{\sqrt{2\pi\sigma^2}^N} \exp\left(-\frac{z^2}{\sigma^2}\right) \quad (1b)$$

$$\frac{d\mathbf{x}}{dt} = -\gamma \frac{\partial \mathbf{V}(\mathbf{x}, t)}{\partial \mathbf{x}} + \xi(t) \quad (1c)$$

The parameter σ controls the width of the normalized Gaussian function. If the Gaussian is excessively wide then the number of classes that present will be restricted. In contrast an excessively narrow Gaussian will impair the functionality of the system as and longer time-scales will be required if the system is to converge to a suitable time-averaged density of inputs. In the context of the following results we consider a 3-dimensional model, as such $N = 3$ and $\mathbf{x} = (x_R, x_G, x_B)$.

Machine Representation of Colour

It is evident that within an animal population the ability of the eye to differentiate between colours can be affected by factors such as: age, fatigue and visual impairments. Consider for instance that approximately 8% of US males with Northern European ancestry had Red-Green colour blindness as of April 2011 [15]. We should also mention that 2 people may perceive the same colour completely differently however assign the same 'label' or 'name' to it [22]. It is also possible that colour-class boundaries are not consistent between different populations [4]. Determining if this is the case presents a considerable problem as we cannot simply swap input device and compare results. The average trichromatic human is able to perceive 1 million colours, far below the number of unique colours represented by the RGB colour palette. A machine vision model is not limited to the realms of human vision and may hope to appreciate a far greater number of colours. Furthermore the incorporation of stimuli from the entire range of the electromagnetic spectrum may be possible.

Consider what advantages there could be for a learning system capable of understanding stimuli emitted within the infra-red range or ultra-violet regions. The world we see with our eyes is limited by the receptors we possess, increasing the range of sensitivity would completely transform the world we perceive. Equally a reduction in sensitivity would have a similar effect. In the case of our machine learning model we are again limited by receptors, the colour range of our web-camera is significantly inhibited by low light intensity levels. In contrast the human eye contains as many as 6 million cones cells, sensitivity is therefore much greater than for the standard camera. This hardware concern should be considered when comparing the performance between biological and artificial.

2 Methods

Stimuli to the system are provided in the form of RGB triplets defining the colour of the central pixel of a web camera. The RGB Colour scheme is an additive method of representing colours, taking 3 principal light beams (Red, Green, Blue) and adding their light spectra [16]. This process was initially intended to mirror the action of the 3 principal types of cone cells found in the human eye [19],[24],[23]. Within a digital framework each component takes a value between 0-256, this infers a finite range of colours rather than a continuous spectrum. Thankfully the discretized 16581375 distinct colours provides a more than adequate approximation of a continuous spectrum. Unfortunately this level of acuity is unlikely to be recognized due to insufficient hardware/software sensitivity.

Stimuli $\boldsymbol{\eta} = (\eta_R, \eta_G, \eta_B)$ are applied to a 3-dimensional version of Jansons model. As such a graphical demonstration of the evolution of $\mathbf{V}(t, x_R, x_G, x_B)$ cannot be realized. We can however observe projections of the profile against 2 of the 3 principal dimensions at finite time intervals. We can also consider the positions of the local minima within the 3-dimensional (x_R, x_G, x_B) space. Combining these methods should enable us to gain a clear understanding of the systems attractors at any point in time. Of greater interest is the recognition of stimuli. This can be observed in the same (x_R, x_G, x_B) space as the local minimas and can also be demonstrated by the direct comparison of colour.

In order to ensure a diverse stimuli we position our web-camera (Logitech HD 720p C270) to observe an LCD monitor playing the popular cartoon 'South park'. Gaussian width parameter σ is set to $\sqrt{0.3}$ and the range 0-256 is constrained onto the scope 0-10. We also set $k = 0$, which infers the system does not forget. Recognition is characterised by the rate parameter $\gamma = 1$ and system noise is neglected. 2 million stimuli are input to the model (equations 1a,1c) which is numerically integrated using a 4th order Runge-Kutta method [17] with an integration step width of 0.001. Light intensity lev-

els are not regulated however remain relatively moderate throughout. Because of the low frame-rate of the camera, which is further reduced by our coding we choose to slow the video playback speed by 50% so that each presented colour can be appreciated. The video is allowed to run in a continuous loop until the model terminates at $t = 2000$.

3 Results

The world of cartoon possesses a wealth of vibrant colours making cartoons such as ‘South park’ ideal candidates for the exploration of this approach. A visualization of the Recognition of an ensemble of 1361 stimuli encountered at various time moments on the interval $0 : 2000$ is provided in figure 1. These stimuli are recognized ‘on-line’ while the system continues to learn. The energy landscape characterizing class formation cannot be entirely visualized graphically, however we can consider projections of this as shown in figures 3a,3b,3c. More accessible and complementary are the positions of local minima in the RGB space, these can be considered the most significant time evolving variables (local minima infer distinct classes). Figure 2 demonstrates the location of all local minima at several discrete time intervals. The colour of each point is related to location and size is scaled to the depth of the potential well. To demonstrate the basins of attraction of each class we can consider the same gradient descent routine used for stimuli recognition. Taking the stationary energy landscape for $t = 2000$ a mesh of initial points can be allowed to evolve. The evolution of these points is shown in figures 4b and 4c from different viewpoints. The corresponding colour classes are given in figure 4a. Basin size can be observed at any moment by this approach. In figures 4b and 4c line colour indicates the class to which the point will converge. During general operation this learning system will automatically perceive and classify huge arrays of stimuli while recognizing the class to which certain selected stimuli belong.

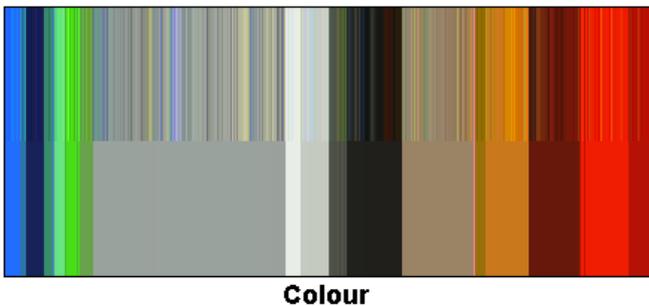


Figure 1: Recognition of an ensemble of stimuli during the active formation of classes (while the system continues to learn). Stimuli are shown on the top row with recognized colour classes below.

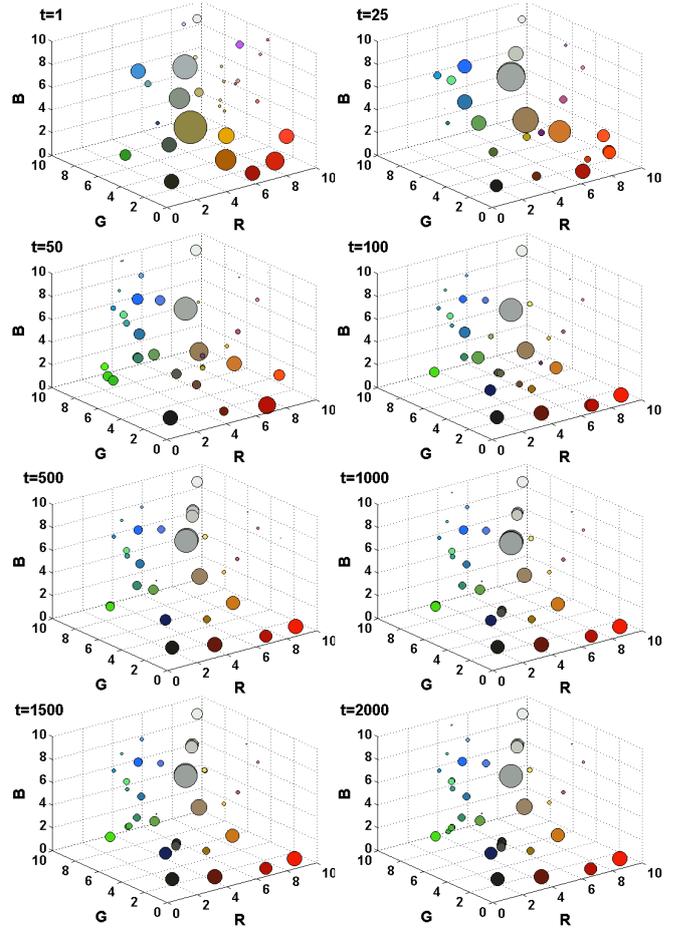


Figure 2: Evolution of the local minima of the energy landscape at discrete times. The colour of each point corresponds to the position in RGB space while the size of a point is scaled to the depth of the local minima.

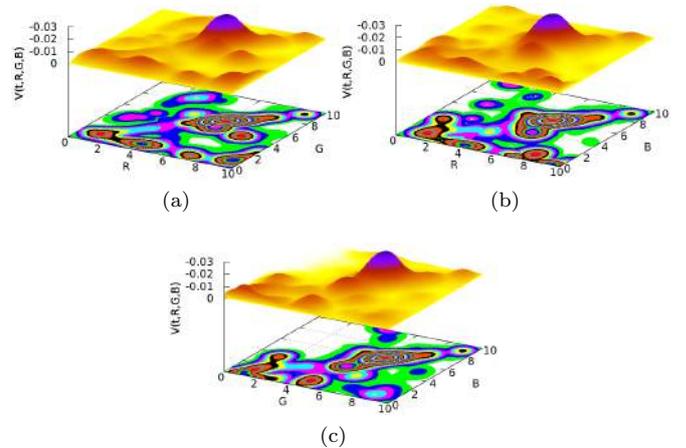
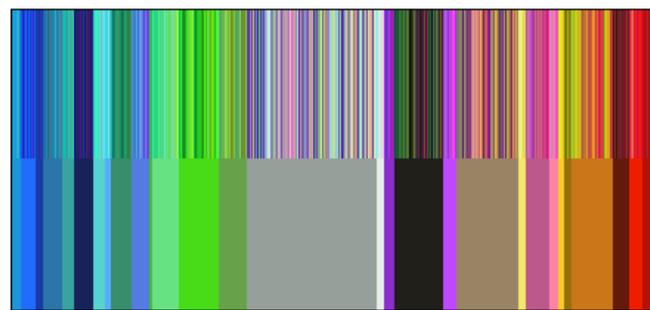
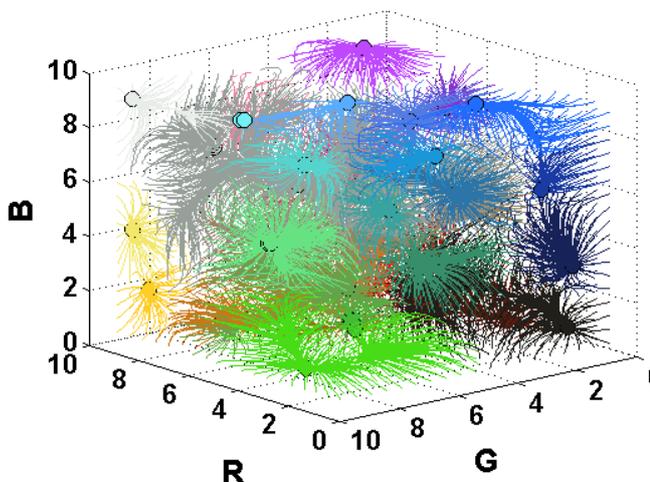


Figure 3: (a)-(c): Projections of the the energy profile $V(t, R, G, B)$ at $t = 2000$ on the 3 pairs of axis (R,G), (R,B) and (G,B).

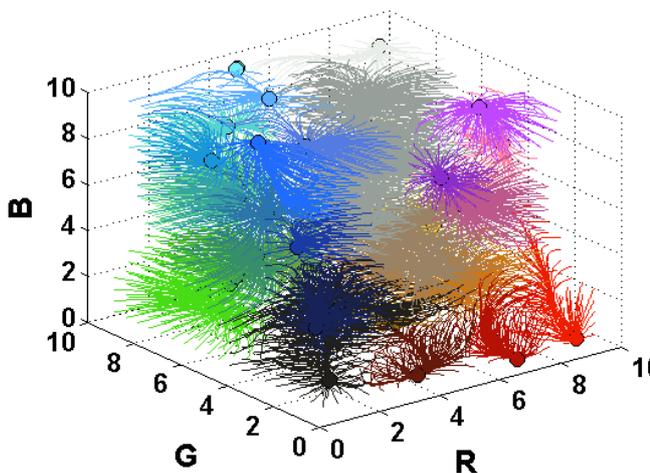
4 Discussion



(a)



(b)



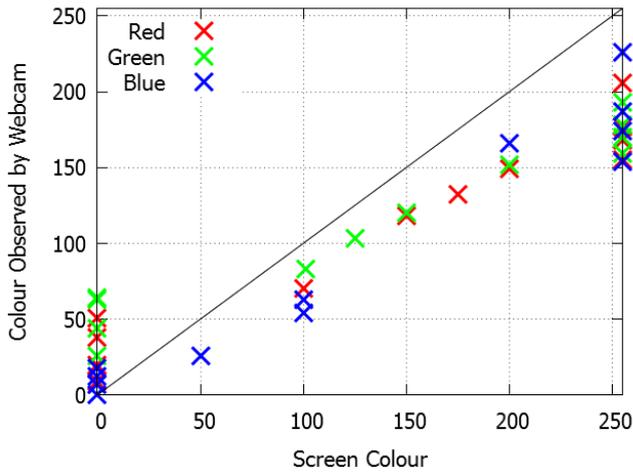
(c)

Figure 4: (a): Colour Classification of an array of RGB colours (top) and their associated classes (bottom). Energy Potential is fixed at the landscape for $t = 2000$. (b),(c): Gradient descent to local minima (class), initial and final states corresponding to those shown in (a). Colour infers the final class to which a stimuli is attributed. The sizes of each basin of attraction can be inferred.

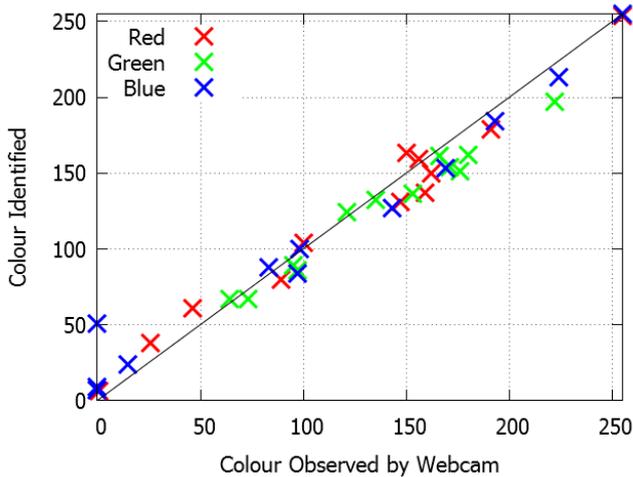
Distinct classes emerge that are represented by the local minima of the potential energy function. Classes are successfully realized by gradient descent demonstrating the recognition of a stimuli based on the knowledge retained by the system. The rate of convergence to the minima of a class is associated with the parameter γ . Control of this parameter may be related to the urgency/focus of the learning system. Since the energy profile continuously receives stimuli the basin of attraction for a given attractor in phase space is continuously adjusted, the recognition rate parameter may therefore directly influence the classification of a stimuli. From a human learning perspective this may be perceived as ‘thinking time’, decisions made in haste are less likely to be optimal compared to those where more time is exceeded. Selecting a rate of convergence that is too high may result in sub-optimal stimuli class recognition. From an evolving time perspective the system is not guaranteed to always recognized identical stimuli as belonging to the same class. A class recognized at one moment in time may no longer exist at a subsequent moment. New experiences since that stimuli was last encountered can have a dramatic effect on recognition. This feature is indicative of human learners where greater experience of a subject allows us to make more accurate decisions and better asses our environment. The depth of a local minima within the potential energy landscape (figure 3) infers the rate of convergence, a narrow deep well can be considered a well defined class, a stimuli within this class is recognized rapidly. Broader shallower potential well’s result in a longer convergence time and may also be considered to be less well defined, reflecting a vague knowledge. When considering the efficacy of the model it is important to consider the limitations of the hardware. The success of the routine is emphasized when one considers the home computing, budget equipment utilized. If we consider the Acer monitor to be correctly calibrated then we can compare the known RGB values to colours detected by the hardware/software. Figure 5a compares the known colour, the colour observed by the web camera and the colour identified by our code. By ‘observed colour’ we refer to the colour observed by commercial web camera software whilst the camera observes the screen. Standard indoor lighting conditions are maintained. The limitations of the hardware are made apparent by figure 5b where the observed colour is not linear with screen colour. Problems exist at the extremes of the RGB range and observed colour intensity is generally deficient. In contrast figure 5c presents an almost linear relationship between observed colour and the colour identified by our code. This suggests that any differences between colours perceived by a human learner and those learnt by our artificial system, whilst observing the same input, are a result of the hardware receptors used rather than the learning methodology.



(a)



(b)



(c)

Figure 5: **(a)**: Comparison between the screen colour, the observed colour (screenshot of webcam software) and output RGB colour. **(b)**: A graphical comparison of the screen colour and the observed colour. The optimal case is described by the line $x = y$. **(c)**: Observed Colour against Output colour. Note that the colour of the marker matches the corresponding component of the RGB value.

If we re-consider the earlier comments referring to the differing levels of visual acuity within biology then these hardware limitations can be perceived as having negligible effect on our overall goal. We effectively classify stimuli that are received by our artificial system, there is no requirement that machine vision should perceive colours equivalent to those inferred by humans. Repeating the routine with a superior web camera is expected to improve the correlation between what a human observes and the colours learnt. The functionality of the learning model would remain unchanged.

5 Conclusion

The system autonomously shapes its architecture in an unsupervised manner to a received stimuli. Where this input is colour data results may not exactly replicate those of a human learner however this is due to hardware restrictions rather than large learning errors. Colours observed by the system are attributed to classes, the characteristics of a class depending on the frequency and distribution of inputs. The recognition and classification of stimuli occur simultaneously without the need for an isolated training phase. Although we have utilized colour data the same paradigm can be applied to any learning goal. Previous work ([8]) has featured auditory stimuli. Increasing the dimensionality of the system should allow for the combined modelling of a vast numbers of sensory inputs. This would however increase computational demand. The development of dedicated hardware may lead to this model becoming a viable model of Artificial Intelligence that could out-perform current neural network models.

References

- [1] Human Brain Project. <https://www.humanbrainproject.eu/>, 2013. [Online; accessed 26-August-2014].
- [2] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207–216, June 1993.
- [3] Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 1998.
- [4] Jules Davidoff, Ian Davies, and Debi Roberson. Colour categories in a stone-age tribe. *Nature*, 398(6724):203–204, 1999.
- [5] Marc W. Halterman. Neuroscience, 3rd edition. *Neurology*, 64(4):769–769–a, 2005.
- [6] Jolm Hertz, Anders Krogh, and Richard G Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley Publishing Company, 1991.
- [7] John J Hopfield. Neural networks and physical systems with emergent collective computational abili-

- ties. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [8] Natalia B Janson and Christopher J Marsden. “memory foam” approach to unsupervised learning. *arXiv preprint arXiv:1107.0674*, 2011.
- [9] Paul J Jungwirth, Andrew H Rawicz, and Romuald Lakowski. Applications of neural networks to colour recognition. *Biological cybernetics*, 79(5):361–365, 1998.
- [10] Eun Ryung Lee, Pyeoung Kee Kim, and Hang Joon Kim. Automatic recognition of a car license plate using color image processing. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 2, pages 301–305. IEEE, 1994.
- [11] Christopher J Marsden. *Nonlinear dynamics of pattern recognition and optimization*. PhD thesis, Loughborough University, October 2012.
- [12] Paul A Merolla, John V Arthur, Rodrigo Alvarez-Icaza, Andrew S Cassidy, Jun Sawada, Filipp Akopyan, Bryan L Jackson, Nabil Imam, Chen Guo, Yutaka Nakamura, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.
- [13] Melanie Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- [14] U.S. Department of Health & Human Services (HHS). the BRAIN Initiative. <http://www.nih.gov/science/brain/>, 2013. [Online; accessed 26-August-2014].
- [15] National Library of Medicine. Genetics Home Reference: Color vision deficiency. <http://ghr.nlm.nih.gov/condition/color>, 2011. [Online; accessed 18-August-2014].
- [16] Charles Poynton. *Digital video and HD: Algorithms and Interfaces*. Elsevier, 2012.
- [17] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes in C*, volume 2. Citeseer, 1996.
- [18] Stuart Russell and Peter Norvig. *Artificial intelligence: A modern approach*. 2009.
- [19] Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73, 1931.
- [20] Andrea Soltoggio. Short-term plasticity as cause-effect hypothesis testing in distal reward learning. *arXiv preprint arXiv:1402.0710*, 2014.
- [21] Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer, 2008.
- [22] Michael A Webster. Human colour perception and its adaptation. *Network: Computation in Neural Systems*, 7(4):587–634, 1996.
- [23] W D Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.
- [24] W D Wright. A re-determination of the mixture curves of the spectrum. *Transactions of the Optical Society*, 31(4):201, 1930.

Bibliography

- [1] Human brain project. <https://www.humanbrainproject.eu/>, 2013. [Online; accessed 26-August-2014].
- [2] Wavelength to rgb in python. http://www.noah.org/wiki/Wavelength_to_RGB_in_Python, 2014. [Online; accessed 13-November-2014].
- [3] R. Agrawal, T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2):207–216, June 1993.
- [4] M. B. Ahrens, M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, 10(5):413–420, 2013.
- [5] A. P. Alivisatos, M. Chun, G. M. Church, R. J. Greenspan, M. L. Roukes, and R. Yuste. The brain activity map project and the challenge of functional connectomics. *Neuron*, 74(6):970–974, 2012.
- [6] J. Aron. Software tricks people into thinking it is human. *New Scientist*. [Online] September, 6, 2011.
- [7] D. Aronson, E. J. Doedel, and H. G. Othmer. An analytical and numerical study of the bifurcations in a system of linearly-coupled oscillators. *Physica D: Nonlinear Phenomena*, 25(1):20–104, 1987.
- [8] D. Aronson, G. B. Ermentrout, and N. Kopell. Amplitude response of coupled oscillators. *Physica D: Nonlinear Phenomena*, 41(3):403–449, 1990.
- [9] E. Athanassoula. Nonlinear dynamics in astronomy and physics. *Ann. NY Acad. of Sci*, 1045:168, 2005.
- [10] A. Balanov, N. Janson, D. Postnov, and O. Sosnovtseva. *Synchronization: from simple to complex*. Springer, 2008.
- [11] J. C. A. Barata and M. S. Hussein. The moore–penrose pseudoinverse: A tutorial review of the theory. *Brazilian Journal of Physics*, 42(1-2):146–165, 2012.
- [12] A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 1998.
- [13] C. S. Bertuglia and F. Vaio. *Nonlinearity, chaos, and complexity: the dynamics of natural and social systems*. Oxford University Press, 2005.
- [14] A. Beuter. *Nonlinear dynamics in physiology and medicine*, volume 25. Springer, 2003.

- [15] C. Blum and X. Li. *Swarm intelligence in optimization*. Springer, 2008.
- [16] G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *Ann. Math. Statist.*, 29(2):610–611, 06 1958.
- [17] R. Boyd. Do people only use 10 percent of their brains. *Scientific American*, 2008.
- [18] R. Bracewell. *The Fourier transform and its applications*. McGraw-Hill, 1986.
- [19] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.
- [20] S. Bringsjord. Is the connectionist-logicist clash one of ai’s wonderful red herrings? *Journal of Experimental & Theoretical Artificial Intelligence*, 3(4):319–349, 1994.
- [21] S. Bringsjord. The logicist manifesto: At long last let logic-based artificial intelligence become a field unto itself. *Journal of Applied Logic*, 6(4):502–525, 2008.
- [22] S. Bringsjord and M. Zenzen. In defense of hyper-logicist ai. In *IJCAI*, volume 91, pages 1066–1072, 1991.
- [23] A. D. Broadbent. A critical review of the development of the cie1931 rgb color-matching functions. *Color Research & Application*, 29(4):267–272, 2004.
- [24] D. Bruton. Approximate rgb values for visible wavelengths. <http://www.physics.sfasu.edu/astro/color/spectra.html>, 1996. [Online; accessed 13-November-2014].
- [25] A. E. Bryson, Y.-C. Ho, and G. M. Siouris. Applied optimal control: Optimization, estimation, and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(9):366–367, 1979.
- [26] R. Cellan-Jones. Stephen hawking warns artificial intelligence could end mankind. *BBC News Online*, 2014. [Online; accessed 09-January-2015].
- [27] T. Chakraborty and R. H. Rand. The transition from phase locking to drift in a system of two weakly coupled van der pol oscillators. *International Journal of Non-Linear Mechanics*, 23(5):369–376, 1988.
- [28] O. Chapelle, B. Schölkopf, A. Zien, et al. *Semi-supervised learning*, volume 2. MIT press Cambridge, 2006.
- [29] C. Chiarella, G. I. Bischi, and L. Gardini. *Nonlinear Dynamics in Economics, Finance and Social Sciences: Essays in Honour of John Barkley Rosser Jr.* Springer, 2009.
- [30] D. Crevier. *The Tumultuous History of the Search for Artificial Intelligence*. Basic Books, Inc., 1994.
- [31] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines*. Cambridge University Press Cambridge, United Kingdom:, 2000.
- [32] J. Davidoff, I. Davies, and D. Roberson. Colour categories in a stone-age tribe. *Nature*, 398(6724):203–204, 1999.

- [33] P. Dayan and C. J. C. H. Watkins. *Reinforcement learning, Encyclopedia of Cognitive Science*. MacMillan Press, 2001.
- [34] E. D. Dickmanns. The development of machine vision for road vehicles in the last decade. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 1, pages 268–281. IEEE, 2002.
- [35] H. A. Dijkstra. *Nonlinear Climate Dynamics*. Cambridge University Press, 2013.
- [36] N. Doidge. *The brain that changes itself: stories of personal triumph from the frontiers of brain science*. Viking Press, 2007.
- [37] W. Duch. What is computational intelligence and where is it going? In *Challenges for computational intelligence*, pages 1–13. Springer, 2007.
- [38] A. Eden, E. Steinhart, D. Pearce, and J. Moor. Singularity hypotheses: An overview. In *Singularity Hypotheses*, pages 1–12. Springer, 2012.
- [39] F. L. Eldrige. Phase resetting of respiratory rhythm experiments in animals and models. In *Rhythms in physiological systems*, pages 165–175. Springer, 1991.
- [40] Y. P. Emelianova, A. P. Kuznetsov, and L. V. Turukina. Quasi-periodic bifurcations and amplitude death in low-dimensional ensemble of van der pol oscillators. *Physics Letters A*, 378(3):153–157, 2014.
- [41] Y. P. Emelianova, A. P. Kuznetsov, L. V. Turukina, I. R. Sataev, and N. Y. Chernyshov. A structure of the oscillation frequencies parameter space for the system of dissipatively coupled oscillators. *Communications in Nonlinear Science and Numerical Simulation*, 19(4):1203–1212, 2014.
- [42] T. Endo and S. Mori. Mode analysis of a ring of a large number of mutually coupled van der pol oscillators. *Circuits and Systems, IEEE Transactions on*, 25(1):7–18, 1978.
- [43] G. B. Ermentrout. Oscillator death in populations of all to all coupled nonlinear oscillators. *Physica D: Nonlinear Phenomena*, 41(2):219–231, 1990.
- [44] G. B. Ermentrout. Xppaut. In *Computational Systems Neurobiology*, pages 519–531. Springer, 2012.
- [45] M. D. Fairchild. *Color appearance models*. John Wiley & Sons, 2013.
- [46] J. Fan and D. Li. An overview of data mining and knowledge discovery. *Journal of Computer Science and Technology*, 13(4):348–368, 1998.
- [47] D. Ferrucci. Introduction to this is watson. *IBM Journal of Research and Development*, 56(3.4):1–1, 2012.
- [48] D. Ferrucci, E. Brown, J. Chu-Carroll, J. Fan, D. Gondek, A. Kalyanpur, A. Lally, J. Murdock, E. Nyberg, J. Prager, et al. Building watson: An overview of the deepqa project. *AI magazine*, 31(3):59–79, 2010.
- [49] D. Ferrucci, A. Levas, S. Bagchi, D. Gondek, and E. Mueller. Watson: Beyond jeopardy! *Artificial Intelligence*, 199:93–105, 2013.

- [50] E. Fiesler and R. Beale. *Handbook of neural computation*. Oxford University Press, 1996.
- [51] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical journal*, 1(6):445, 1961.
- [52] L.-M. Fu. *Neural networks in computer intelligence*, 2003.
- [53] J. A. Gámez, S. Moral, and A. S. Cerdan. *Advances in Bayesian networks*, volume 146. Springer, 2013.
- [54] C. W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, 1994.
- [55] Z. Ghahramani. Unsupervised learning. In *Advanced Lectures on Machine Learning*, pages 72–112. Springer, 2004.
- [56] E. Guizzo. How googles self-driving car works. *IEEE Spectrum Online*, October, 18, 2011.
- [57] M. W. Halterman. Neuroscience, 3rd edition. *Neurology*, 64(4):769–769–a, 2005.
- [58] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.
- [59] S. Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, 3, 2009.
- [60] J. Hertz, A. Krogh, and R. G. Palmer. Introduction to the theory of. *Neural Computation*, 1, 1991.
- [61] R. C. Hilborn. *Chaos and nonlinear dynamics: an introduction for scientists and engineers*. oxford university press Oxford, 2000.
- [62] J. H. Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. U Michigan Press, 1975.
- [63] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [64] G. Huang, G.-B. Huang, S. Song, and K. You. Trends in extreme learning machines: A review. *Neural Networks*, 61(0):32 – 48, 2015.
- [65] G.-B. Huang, D. H. Wang, and Y. Lan. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2(2):107–122, 2011.
- [66] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 985–990 vol.2, July 2004.
- [67] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: a new learning scheme of feedforward neural networks. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 2, pages 985–990. IEEE, 2004.
- [68] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

- [69] T. E. Hull and A. R. Dobell. Random number generators. *SIAM review*, 4(3):230–254, 1962.
- [70] K. Itô. Stochastic integral. volume 20, pages 519–524. The Japan Academy, 1944.
- [71] R. Jain, R. Kasturi, and B. G. Schunck. *Machine vision*, volume 5. McGraw-Hill New York, 1995.
- [72] J. F. James. *A student’s guide to Fourier transforms: with applications in physics and engineering*. Cambridge university press, 2011.
- [73] W. James. The energies of men. *The Philosophical Review*, pages 1–20, 1907.
- [74] N. B. Janson and C. J. Marsden. “memory foam” approach to unsupervised learning. *arXiv preprint arXiv:1107.0674*, 2011.
- [75] B. Javidi. *Image recognition and classification: algorithms, systems, and applications*. CRC Press, 2002.
- [76] G. Jordan, S. S. Deeb, J. M. Bosten, and J. D. Mollon. The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of Vision*, 10(8):12, 2010.
- [77] D. Jurafsky and J. H. Martin. *Speech and language processing (prentice hall series in artificial intelligence)*. 2008.
- [78] D. Karakos, S. Khudanpur, J. Eisner, and C. E. Priebe. Unsupervised classification via decision trees: An information-theoretic perspective. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP’05). IEEE International Conference on*, volume 5, pages 5–1081. IEEE, 2005.
- [79] G. V. Kass. An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, pages 119–127, 1980.
- [80] C. King. Fractal and chaotic dynamics in nervous systems. *Progress in Neurobiology*, 36(4):279–308, 1991.
- [81] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice Hall New Jersey, 1995.
- [82] A. Konar. *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*. CRC press, 1999.
- [83] A. Konar. Competitive learning using neural nets. *Computational Intelligence: Principles, Techniques and Applications*, pages 267–293, 2005.
- [84] A. Konar. *Computational Intellingence*. Springer, 2005.
- [85] N. M. Krylov and N. N. Bogoliubov. *Introduction to non-linear mechanics*. Princeton University Press, 1943.
- [86] A. P. Kuznetsov and J. P. Roman. Properties of synchronization in the systems of non-identical coupled van der pol and van der pol–duffing oscillators. broadband synchronization. *Physica D: Nonlinear Phenomena*, 238(16):1499–1506, 2009.
- [87] M. Kyan, P. Muneesawang, K. Jarrah, and L. Guan. *Unsupervised Learning via Self-Organization: A Dynamic Approach*. Wiley-IEEE Press, 2014.

- [88] D. H. Lehmer. Mathematical methods in large-scale computing units. In *Proc. 2nd Symp. on Large-Scale Digital Calculating Machinery*, pages 141–146. Harvard Univ. Press Cambridge, MA, 1951.
- [89] D. S. Lemons and A. Gythiel. Paul langevins 1908 paper on the theory of brownian motion [sur la théorie du mouvement brownien, c. r. acad. sci. (paris) (1908)]. *American Journal of Physics*, 65(11):1079–1081, 1997.
- [90] V. Lertnattee, S. Chomya, et al. Applying clustering and association rule learning for finding patterns in herbal formulae. *ACEEE Interational Journal on Network Security*, 3(2), 2012.
- [91] P. A. W. Lewis, A. S. Goodman, and J. M. Miller. A pseudo-random number generator for the system/360. *IBM Systems Journal*, 8(2):136–146, 1969.
- [92] J. Lighthill. Artificial intelligence: A general survey. In *Science Research Council, Artificial Intelligence: A Paper Symposium, mimeo. London: Science Research Council*, 1973.
- [93] D. T. Lindsey and A. M. Brown. The color lexicon of american english. *Journal of Vision*, 14(2):17, 2014.
- [94] M. D. MacLaren and G. Marsaglia. Uniform random number generators. *Journal of the ACM (JACM)*, 12(1):83–89, 1965.
- [95] A. N. Malakhov. Time scales of overdamped nonlinear brownian motion in arbitrary potential profiles. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 7(3):488–504, 1997.
- [96] J. Mandziuk. Some thoughts on using computational intelligence methods in classical mind board games. In *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 4002–4008, June 2008.
- [97] H. Markram. *The Human Brain Project: A Report to the European Commission*. 2012.
- [98] G. Marsaglia, W. W. Tsang, et al. The ziggurat method for generating random variables. *Journal of statistical software*, 5(8):1–7, 2000.
- [99] C. J. Marsden. *Nonlinear dynamics of pattern recognition and optimization*. PhD thesis, Loughborough University, October 2012.
- [100] J. B. Martin, C. M. Pechura, et al. *Mapping the Brain and Its Functions:: Integrating Enabling Technologies into Neuroscience Research*, volume 91. National Academies Press, 1991.
- [101] M. Matsumoto and T. Nishimura. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8(1):3–30, 1998.
- [102] P. C. Matthews, R. E. Mirollo, and S. H. Strogatz. Dynamics of a large system of coupled nonlinear oscillators. *Physica D: Nonlinear Phenomena*, 52(2):293–331, 1991.
- [103] R. Mausfeld and D. Heyer. *Colour Perception: Mind and the Physical World*. Oxford University Press, 2003.

- [104] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.
- [105] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.
- [106] B. Midollini, E. Re, and P. Magnani. Space exploration: the automation and robotics impulse. In *SpaceOps 2006 Conference*, page 5509, 2006.
- [107] D. Mihai and E. Străjescu. From wavelength to rgb filter. 2007.
- [108] F. P. Miller, A. Vandome, and M. John. *Inverse Transform Sampling*. VDM Publishing, 2010.
- [109] M. Minsky and S. Papert. Perceptrons. *Cambridge, Ma*, pages 105–110, 1969.
- [110] M. Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.
- [111] T. Mitchell. *Machine learning*. McGraw Hill, 1997.
- [112] C. N. Moschopoulos, P. Tsiatsis, G. N. Beligiannis, D. Fotakis, and S. D. Likothanassis. Dealing with large datasets using an artificial intelligence clustering tool. In C. Koutsojannis and S. Sirmakessis, editors, *Tools and Applications with Artificial Intelligence*, volume 166 of *Studies in Computational Intelligence*, pages 105–120. Springer Berlin Heidelberg, 2009.
- [113] A. Mourikis, N. Trawny, S. Roumeliotis, A. E. Johnson, A. Ansar, L. Matthies, et al. Vision-aided inertial navigation for spacecraft entry, descent, and landing. *Robotics, IEEE Transactions*, 25(2):264–280, 2009.
- [114] S. Muggleton, R. Otero, and A. Tamaddoni-Nezhad. *Inductive logic programming*, volume 168. Springer, 1992.
- [115] D. G. Myers. *Psychology*. New York : Worth, 1995.
- [116] National Library of Medicine. Genetics home reference: Color vision deficiency. <http://ghr.nlm.nih.gov/condition/color>, 2011. [Online; accessed 18-August-2014].
- [117] R. Neapolitan. *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, San Francisco, California, United States, 1988.
- [118] J. Needham. *Science and Civilisation in China Volume 2 History of Scientific Thought*. Cambridge University Press, 1956.
- [119] A. Newell and J. C. Shaw. A variety of intelligent learning in a general problem solver. *RAND Report*, 6, 1959.
- [120] M. A. Nielsen. *Neural Networks an Deep Learning*. Determination Press, 2015.
- [121] T. Ookawara and T. Endo. Effects of the deviation of element values in a ring of three and four coupled van der pol oscillators. *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 46(7):827–840, 1999.

- [122] D. E. Panayotounakos, N. D. Panayotounakou, and A. F. Vakakis. On the lack of analytic solutions of the van der pol oscillator. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 83(9):611–615, 2003.
- [123] S. Park and K. Miller. Random number generators: good ones are hard to find. *Communications of the ACM*, 31(10):1192–1201, 1988.
- [124] I. Pastor-Díaz and A. López-Fraguas. Dynamics of two coupled van der pol oscillators. *Physical Review E*, 52(2):1480, 1995.
- [125] W. H. Payne, J. R. Rabung, and T. P. Bogyo. Coding the lehmer pseudo-random number generator. *Communications of the ACM*, 12(2):85–86, 1969.
- [126] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: a universal concept in nonlinear sciences*, volume 12. Cambridge university press, 2003.
- [127] E. J. Powers, D. Gray, R. C. Green, S. Levialdi, V. Cantoni, and V. Roberto. *Artificial Vision: Image Description, Recognition, and Communication*. Academic Press, 1996.
- [128] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992.
- [129] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*, volume 2. Citeseer, 1996.
- [130] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.
- [131] J. R. Quinlan. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [132] R. H. Rand and P. J. Holmes. Bifurcation of periodic motions in two weakly coupled van der pol oscillators. *International Journal of Non-Linear Mechanics*, 15(4):387–399, 1980.
- [133] I. Riedel, P. Gueguen, F. Dunand, and S. Cottaz. Macroscale vulnerability assessment of cities using association rule learning. *Seismological Society of America*, 85(2):295–305, 2014.
- [134] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Series in Synergetics, 1996.
- [135] R. Rojas. *Neural networks: a systematic introduction*. Springer, 1996.
- [136] K. Rompala, R. Rand, and H. Howland. Dynamics of three coupled van der pol oscillators with application to circadian rhythms. *Communications in Nonlinear Science and Numerical Simulation*, 12(5):794–803, 2007.
- [137] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Cognitive modelling*, 1988.
- [138] D. E. Rummelhart. Learning representations by back-propagating errors. *Nature*, 323(9):533–536, 1986.
- [139] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2009.

- [140] A. P. Saygin, I. Cicekli, and V. Akman. Turing test: 50 years later. In *The Turing Test*, pages 23–78. Springer, 2003.
- [141] P. Sibi, S. A. Jones, and P. Siddarth. Analysis of different activation functions using back propagation neural networks. *Journal of Theoretical and Applied Information Technology*, 47(3):1344–1348, 2013.
- [142] P. K. Simpson. *Artificial neural systems*. Pergamon, 1990.
- [143] S. N. Sivanandam and S. N. Deepa. *Introduction to Genetic Algorithms*. Springer, 2008.
- [144] T. Smith and J. Guild. The cie colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73, 1931.
- [145] A. Stevenson. *Oxford dictionary of English*. Oxford University Press, 2010.
- [146] D. W. Storti and R. H. Rand. Dynamics of two strongly coupled van der pol oscillators. *International Journal of Non-Linear Mechanics*, 17(3):143–152, 1982.
- [147] R. L. Stratonovich. A new representation for stochastic integrals and equations. *SIAM Journal on Control*, 4(2):362–371, 1966.
- [148] S. H. Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(14):1 – 20, 2000.
- [149] S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology and chemistry*. Perseus publishing, 2001.
- [150] K. A. Stroud and D. J. Booth. *Engineering mathematics*. palgrave macmillan, 2013.
- [151] C. Tenreiro and R. Elgueta. Modeling the sleep–wake cycle using coupled van der pol oscillators. *Biological Rhythm Research*, 41(2):149–157, 2010.
- [152] G. Teschl. *Ordinary differential equations and dynamical systems*, volume 140. American Mathematical Soc., 2012.
- [153] A. Tocino and R. Ardanuy. Rungekutta methods for numerical solution of stochastic differential equations. *Journal of Computational and Applied Mathematics*, 138(2):219 – 241, 2002.
- [154] A. M. Turing. Computing machinery and intelligence. *Mind*, pages 433–460, 1950.
- [155] U.S. Department of Health & Human Services (HHS). the BRAIN Initiative. <http://www.nih.gov/science/brain/>, 2013. [Online; accessed 26-August-2014].
- [156] B. Van der Pol and J. Van der Mark. Lxxii. the heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(38):763–775, 1928.
- [157] N. G. Van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2011.
- [158] M. Versichele, L. de Groote, M. Claeys Bouuaert, T. Neutens, I. Moerman, and N. van de Weghe. Pattern mining in tourist attraction visits through association rule learning on bluetooth tracking data: A case study of ghent, belgium. *Tourism Management*, 44:67–81, 2014.

- [159] J. Walker. Colour rendering of spectra. <http://www.fourmilab.ch/documents/specrend/specrend.c>, 2003. [Online; accessed 13-November-2014].
- [160] G. Watts. Neuroscientists condemn european brain simulation project. *BMJ*, 349, 2014.
- [161] M. A. Webster. Human colour perception and its adaptation. *Network: Computation in Neural Systems*, 7(4):587–634, 1996.
- [162] P. Werbos. *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, Jan 1974.
- [163] A. Wilson. Machine vision and image processing: Past, present, and future. *Vision Systems Design*, 2013.
- [164] J. Winawer, N. Witthoft, M. C. Frank, L. Wu, A. R. Wade, and L. Boroditsky. Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19):7780–7785, 2007.
- [165] W. D. Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.
- [166] W. D. Wright. A re-determination of the mixture curves of the spectrum. *Transactions of the Optical Society*, 31(4):201, 1930.
- [167] M. Zeidenberg. *Neural networks in artificial intelligence*. Ellis Horwood, 1990.
- [168] C. Zimmer. 100 trillion connections. *Scientific American*, 304(1):58–63, 2010.