Qualitative Model-Based Multi-sensor Data Fusion: The Qualitative Kalman Filter

Steven Reece

Department of Engineering Science
Robotics Research Group
University of Oxford

Trinity Term, 1998
This thesis is submitted to the Department of Engineering Science, University of Oxford, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The thesis is entirely my own work, and except where otherwise stated, describes my own research.

Steven Reece, Oriel College, Oxford

Copyright © 1998, Steven Reece
All Rights Reserved
Qualitative Model-Based Multi-sensor Data Fusion: The Qualitative Kalman Filter

Abstract

Sensor based navigation is a fundamental competence for any mobile robot. However, no single sensor is omnipotent and systems employ networks of sensors, each sensor designed to overcome the inadequacies of the others. Often, complementary information from these sensors can be used to determine environment characteristics which would otherwise be indeterminable using a single sensor. In conventional systems, the relationships between sensory information is often necessarily in the form of exact, quantitative, algebraic models (Abidi and Gonzalez, 1992) and in practice, fusion systems incorporating these models are extremely fragile:

- complex models can be unwieldy or the details of the mathematics too complex to offer precise models. For example, the interpretation of sensor output might require detailed knowledge of the physics of the sensing process (Bozma and Kuc, 1994).
- extensive sensor calibration is required (Manyika and Durrant-Whyte, 1994).
- a sufficient representation of systematic error must be included for unmodelled or non-determinable components such as temperature changes or hysteresis effects in sensor components (Barshan and Durrant-Whyte, 1995).
- numerical inaccuracies arise for noise filtering techniques which require linear model approximations (Bar-Shalom and Fortmann, 1988).

Often, numerical measures are not required and less detailed, qualitative information is perfectly adequate. A qualitative model is a (small) finite description of an infinitely complex reality, constructed for the purpose of answering particular questions (Kuipers, 1994). Such models allow us to use imprecise (i.e. partial or incomplete) knowledge of the system to reason about the interrelationships between sensor cues. Reasoning with incomplete knowledge and logical abstractions of the underlying geometry and physics of systems has been studied in the areas of Qualitative Reasoning and Qualitative Physics within Artificial Intelligence (Bobrow, 1985; Weld, 1990; Kuipers, 1994).

I investigate qualitative data fusion in the context of robot navigation using qualitative models of physical sensor observations. These aim to describe the world in terms of local sensor-centric representations of the observed environment. Each representation exploits those landmarks most natural to the physical sensor involved. No global description of the environment is maintained and
no explicit geometric representation of the world is assumed. This leads naturally to a navigation process defined in terms of relationships between different sensor observables. I investigate qualitative physical and spatial models for various sensors and describe a general method for filtering noisy information and stochastic processes. The Kalman filter is a popular, non-parametric data fusion method which maintains a point estimate of state using the first two moments (i.e. mean and standard error) of the underlying noise probability density function (Bar-Shalom and Fortmann, 1988). The Qualitative filter, presented in this thesis combines the Kalman filter and the Dempster-Shafer Theory of Evidential Reasoning for non-parametric estimation when point-estimates are not available. The Qualitative filter uses information about the mean and standard deviation of the underlying mass distribution to propagate and fuse evidence with process and observation models. The Qualitative filter is an intrinsically more robust mechanism than the Kalman filter which represents imprecision as systematic error covariance matrices. The Qualitative filter is also more robust than the commonly used interval intersection method found in the Qualitative Reasoning literature (Gao and Durrant-Whyte, 1994).

In summary, this thesis offers a novel framework for model-based, multi-sensor data fusion when process, observation and noise models are imprecisely defined. The framework is realised for both dependent and independent noise sources, mean and median unbiased estimation, conservative estimation and stochastic processes. Both hypothesis testing and parameter estimation techniques are developed. Various realisations of the Qualitative filter framework are applied to the Robot Localisation Problem. Qualitative models for wide-beam, time-of-flight sensor reflections from specular and scattering material (namely, the Acoustic Flow model and the QEndura method) are developed. Localisation is performed using qualitative spatial reasoning and qualitative data fusion for noisy and imprecisely modelled sensor modalities (namely, sonar, gyrometric and odometric information). We extend the capabilities of an existing robot from the "OxNav" project, specifically regarding its ability to navigate in non-trivial environments.
Acknowledgements

I thank first of all Dr. Hugh Durrant-Whyte for proposing an interesting area of research and for obtaining the necessary funding. I am especially grateful to Prof. Mike Brady who heroically took over the role of supervisor after Dr. Durrant-Whyte left Britain for a slightly sunnier climate and chair at Sydney University. Mike's encouragement during the darker phases of the research was of tremendous importance. I am extremely grateful to him for this and for considerable non-academic related support also.

During the course of this research, a number of people have made valuable comments and, in this regard, I thank Dr. Nic Wilson of Oxford Brookes University and Dr. John Leonard of MIT. I would also like to thank past and present members of the Robotics Research Group: Annette, Simukai, Rob, Zaff, Simon Julier, Jeff, Ben and the Stevens brothers. Amongst many friends at Oriel I thank Mike, Nick, Ralph, Liv and the Boaties.

I have been fortunate to receive simultaneously substantial funding and yet virtually total freedom as to the direction the research took. I must thank EPSRC for substantial support via grants GR/J/46067 and GR/J/57773 and Oriel College for support via a Graduate Scholarship.

Most importantly, I wish to acknowledge my mother and father for their love, support and encouragement at every stage of my career. I am deeply grateful, above all, to them.
# Contents

Table of Contents

1 The Sensor Data Fusion Problem ........................................ 2
   1.1 Introduction ...................................................... 2
   1.2 Autonomous Intelligent Artificial Systems ..................... 2
   1.3 The Multi-Sensor Data Fusion Problem .......................... 4
   1.4 Qualitative Reasoning ............................................ 5
       1.4.1 Qualitative Representations ................................. 9
   1.5 Reasoning Under Imprecision and Uncertainty .................. 11
       1.5.1 The Interval Intersection Filter (IIF) .................... 11
       1.5.2 Classical Bayesian Estimation ............................... 12
       1.5.3 Upper and Lower Probabilities .............................. 14
       1.5.4 The Point Estimation Paradigm ............................... 15
       1.5.5 Hybrid Methods ............................................. 23
       1.5.6 Summary .................................................... 23
   1.6 The Robotics Test Domain ......................................... 24
   1.7 Conclusions and Thesis Overview ................................ 30

2 The Qualitative Sensor Model ........................................... 33
   2.1 Introduction ...................................................... 33
   2.2 Sensor Configuration and Experimental Technique ............... 33
       2.2.1 Noise Characteristics ......................................... 34
   2.3 Towards a Qualitative Model: The Acoustic Flow Equations .... 35
       2.3.1 1st Order Reflections from Arbitrary Curved Surfaces ...... 35
       2.3.2 Multiple Reflections from Planar Surfaces ................. 39
   2.4 The Qualitative Sensor Model ..................................... 41
       2.4.1 Differential Modelling ......................................... 50
   2.5 The Physical Reflector Model ..................................... 52
       2.5.1 Qualitative Response Models ................................ 54
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.2</td>
<td>FM Sonar Sensor Qualitative Rules</td>
<td>56</td>
</tr>
<tr>
<td>2.6</td>
<td>Sensor Management</td>
<td>59</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Qualitative Reasoning Towards Action</td>
<td>60</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Action Refinement</td>
<td>61</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Control Verification</td>
<td>62</td>
</tr>
<tr>
<td>2.7</td>
<td>Conclusions</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>The Qualitative Filter</td>
<td>66</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>66</td>
</tr>
<tr>
<td>3.2</td>
<td>Rational Decision Making</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>A Framework for Qualitative Model-Based Estimation</td>
<td>67</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The Qualitative Filter Philosophy</td>
<td>69</td>
</tr>
<tr>
<td>3.4</td>
<td>P-Norm Dempster-Shafer Theories of Evidential Reasoning</td>
<td>72</td>
</tr>
<tr>
<td>3.4.1</td>
<td>The (\infty)-Norm Theory</td>
<td>75</td>
</tr>
<tr>
<td>3.5</td>
<td>Statistical Estimation</td>
<td>77</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Aleatory Data and Statistical Unbiasedness</td>
<td>77</td>
</tr>
<tr>
<td>3.6</td>
<td>The Lagrangian (\infty)-Norm Interpretation</td>
<td>80</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Path Centred (Lagrangian) BPA</td>
<td>81</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Nary State BPAs</td>
<td>83</td>
</tr>
<tr>
<td>3.7</td>
<td>An Information Centred Approach</td>
<td>85</td>
</tr>
<tr>
<td>3.8</td>
<td>Stochastic Processes: The Stochastic QF (SQF)</td>
<td>88</td>
</tr>
<tr>
<td>3.9</td>
<td>Comparative Study of the QF and Other Filter Philosophies</td>
<td>92</td>
</tr>
<tr>
<td>3.9.1</td>
<td>How closely is the QF related to the KF?</td>
<td>93</td>
</tr>
<tr>
<td>3.10</td>
<td>Related Work</td>
<td>95</td>
</tr>
<tr>
<td>3.11</td>
<td>Conclusions</td>
<td>103</td>
</tr>
<tr>
<td>4</td>
<td>Efficiency, Sensitivity and Robustness of the Qualitative Filter</td>
<td>106</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>106</td>
</tr>
<tr>
<td>4.2</td>
<td>Efficient Estimators</td>
<td>106</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Symmetric Biscay Distributions: An Improved Filter</td>
<td>112</td>
</tr>
<tr>
<td>4.3</td>
<td>Conservative Estimators</td>
<td>113</td>
</tr>
<tr>
<td>4.4</td>
<td>QF Problems</td>
<td>114</td>
</tr>
<tr>
<td>4.5</td>
<td>Lagrangian Decision Preserving Eulerian Estimators (SOPAP)</td>
<td>116</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusions</td>
<td>121</td>
</tr>
<tr>
<td>5</td>
<td>Qualitative Filtering and Semi-Quantitative Representations</td>
<td>122</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>122</td>
</tr>
<tr>
<td>5.2</td>
<td>The Dynamic Landmark Filter (DLQF)</td>
<td>122</td>
</tr>
<tr>
<td>5.2.1</td>
<td>The Binary Biscay Distribution</td>
<td>123</td>
</tr>
</tbody>
</table>
5.2.2 Optimal Fusion .................................. 125
5.2.3 Inference ......................................... 128
5.3 Conservative Estimation .............................. 129
5.4 Some Properties of the DLQF ......................... 130
5.5 Example Application: The Ballistic Father Christmas Problem ......................... 133
5.6 Conclusions ......................................... 136

6 Experiments ........................................... 141
6.1 Introduction ....................................... 141
6.2 Experiment 1: Surface Discrimination (Sparse Representation) ...................... 141
6.3 Experiment 2: Multi-hypothesis Scale-space Reasoning .......................... 144
6.4 Experiment 3: DLQF and EKF Comparison under Zero Imprecision ............. 148
6.5 Conclusions ......................................... 149

7 Conclusions and Directions for Future Research ........................................... 151
7.1 Conclusions and Summary of Contributions ........................................ 151
7.2 Future Work ......................................... 152
  7.2.1 Domain Model Learning ................................ 153
  7.2.2 Adaptive Representations ................................ 155

A Higher Order Sonar Reflection Model ......................................................... 169
A.1 Superposition of Point Sources ..................................................... 171
A.2 Experimental Results ............................................................... 172
A.3 Limitations of Model: General Polygonal Surfaces ................................ 173

B The QSim Acoustic Flow Model ............................................................. 175

C Scale-Space Abstraction ................................................................. 177
C.1 Introduction .................................................. 177
C.2 Proof of COMOC Unbiased Property around Points of Symmetry .............. 178
C.3 COMOC as an Unbiased Mean Filter ........................................... 180
C.4 COMOC as an Unbiased Median Filter ....................................... 181
C.5 Limitations of the COMOC filter ................................................ 181

D A Critique of R. T. Cox's Paper "Probability, Frequency and Reasonable Expectation" ................................................................. 184
D.1 Introduction .................................................. 184
D.2 The Critique ................................................... 184
D.3 Upper and Lower Evidence ....................................................... 189
E  The Generalised Dempster-Shafer Theory  190

F  The Non-existence of a Generic Eulerian Approach To Qualitative Filtering  194
Chapter 1
The Sensor Data Fusion Problem

Ignorance is preferable to error; and he is less remote from truth who believes nothing, than he who believes what is wrong.

Thomas Jefferson
Notes on the State of Virginia
[1781 - 1785] Query VI

1.1 Introduction

In this chapter, we introduce the three component parts of this thesis: multi-sensor data fusion, qualitative reasoning and the mobile robots domain. We argue that qualitative reasoning is both necessary and sufficient for multi-sensor systems (such as a robot) which use highly imprecise models of the domain and sensing processes. This then motivates the thesis, which is a non-parametric filtering paradigm for qualitative representations.

1.2 Autonomous Intelligent Artificial Systems

Sensing is an essential element of most real-world systems. Sensors are used to monitor the operation of a system and to provide information through which it may be controlled. A robot interacts rationally with a system only when it is able to understand what it senses through an ability to relate its sensory cues to a model of the system. As much as the ability to construct languages with inter-personal structural commonalities is innate in man (Pinker, 1995), man is also able to reason about an ever-changing environment using an innate human-environment grammar grounded in the underlying logical structure of the physical laws governing the universe. In order to survive, not only have his sensing modalities evolved to recognise salient sensor cues (such as verticality which could indicate an attacking animal (Sacks, 1995; Pinker, 1995)) but his reasoning abilities have evolved towards efficient and coherent strategies for human-environment interaction.

Treating the environment as if it were an agent, it would have a belief state (and sub-states such as “the cup of coffee is on the table” or “the thermostat is switched on”) and, as we have all experienced, a goal to achieve other belief states (e.g., the “gravity” law generates the goal for the cup to fall to the floor when the table has been accidently kicked from under the cup; similarly, when a complex interaction of physical laws within the thermostat invokes a fan). Viewing physical systems as if they were rational agents allows us to view each system as a potential communicator.
We communicate with the environment continuously via action (e.g., by moving objects) and sensing (i.e., receiving messages, as such, from the environment via our sensors) and just as we assimilate context dependent interpretations of qualitative human verbal statements with our existing beliefs, we are able to assimilate context dependent interpretations of sensor data about the environment. If necessary, this may cause us to revise our beliefs in accordance with their degrees of evidential support (Logan et al., 1994). We plan our observations strategically as much as we plan our communication utterances and, to decide when to act on a specific plan, we must be able to predict the outcomes of our actions which inevitably requires some model or, equivalently, some understanding of the interlocutor.

Since qualitative representations pervade human-human communication, it is not infeasible that robot sensing should also be able to utilise qualitative representations. It is often not possible nor even necessary for the robot to understand every detail of the system or sensor behaviour; the robot's model need describe the system only at some scale (or at a limited discrete set of scales). When making a cup of coffee, very few of us are concerned with the atomic behaviour of the boiling water molecules or even the exact time it would take for the kettle to boil. Such details are difficult to obtain and are often irrelevant to our needs. A representation which captures relevant distinctions between system objects is crucial for an efficient understanding of the system. The physics literature itself distinguishes macro, micro and nano models of the same phenomena whether, for example, it be a component view (macro), a charge carrier view (micro) or an atomic (nano) view of an electric circuit's behaviour. The macro description of a physical behaviour captures a large class of distinct micro behaviours. The scale at which the model operates is governed by two factors: the precision of information that is required by the robot's enquiry and the precision of information that is available to the model builder (whether the system engineer or the robot itself). Further, the model parameters and inter-parameter relationships (i.e., models) may operate at different scales and, in order to obtain a complete (but not necessarily precise) description of the system by information exchange (inference) or fusion, it may be necessary to reconcile various levels of resolution.

The physical component that operates at fundamentally different scales than that described by the model is often called noise. The qualitatively distinct processes (deterministic physics and random physics) that underlie the observation sequence are often identified by the different scales at which they operate. For example, random noise due to statistical fluctuations often operates at a higher frequency than the physical laws we find in text books. The scale at which the model operates is usually at the level of the physical law description which necessitates reasoning away the random nature of the observations.

In summary, an intelligent robot must be able to communicate coherently with the environment and thus, it must be able to reconcile a model of its system with observations it makes of the system. This requires reasoning about the multiple processes (random and deterministic) that make-up the observed signal, and reconciling this data with abstract representations of the physical laws that describe some aspect of the observation sequence. Consequently, an intelligent robot must be able to represent and process two types of partial information (Ruspini, 1987; Walley, 1991):

- **Imprecise.** A system is imprecisely known when its long run behaviour cannot be predicted exactly. The outcome is deterministic but is highly ambiguous.

- **Uncertain.** A system is uncertain if it is impossible to predict a single outcome but a precise statement about its long run behaviour can be obtained. For example, it is impossible to predict that the next tossed coin will land heads or tails; but it is possible to predict that half
of a pocket full of coins will land heads when thrown.

1.3 The Multi-Sensor Data Fusion Problem

Many advanced systems now make use of large numbers of sensors in applications ranging from aerospace and defence, through robotics and automation systems, to the monitoring and control of process and power generation plants. The primary purpose of using multiple sensors is to provide a more accurate and more comprehensive understanding of the system and its operation (Abidi and Gonzalez, 1992). However, to provide such an understanding, it is essential that the information provided by these sensors is interpreted and combined in such a way that a reliable, complete and coherent description of the system is obtained. This, in essence, is the multi-sensor data fusion problem.

The ability to understand many different situations requires thorough, flexible and general purpose models. However, these models must encode sufficient knowledge to explain the observation behaviours informatively. Two extreme modelling paradigms dominate the sensor data fusion literature. Grid-based methods such as the occupancy grid (Eliffes, 1987; Matties and Eliffes, 1988; Courtney and Jain, 1994) absorb model imprecision into probability distributions over regular grids in parameter space. This representation requires few assumptions about the environment and hence each application has a wide domain of operation. However, this representation utilises little of the information that is available. Feature-based methods use a precise geometric description of the environment and can elicit all the information available in the observation. Grid-based and feature-based approaches occupy two extremes of a representation spectrum concerning the amount of information provided by the approach and the amount information required by the approach. Both approaches are very inflexible as they cannot represent the information at the level of resolution (or scale) appropriate to the information or to the task. A flexible representation is crucial (Leonard and Durrant-Whyte, 1992):^1

\[\text{The type of representation we use determines what information is made explicit in the model; the purposes for which a model can be used and the efficiency with which those purposes can be accomplished follow directly from the choice of representation.}\]

The exploitation of representation for flexible data fusion is the central motivation of this thesis. The choice of representation depends on the kind of information that is being combined. In data fusion, information plays one of two roles. It can be either competitive or complementary (Maheshkumar et al., 1996):

- **Competitive.** Different readings arise out of random noise. Integration increases the reliability of the information. Competitive information reduces uncertainty.

- **Complementary.** Different sensors return information about the same system and sensors themselves are not identical. They have different capabilities and limitations. Reliable information is obtained by overcoming the individual limitations of each sensor. Complementary information reduces imprecision.

Both types of information can take many forms. For example, competitive information can be statistical, probabilistic or linguistic whereas complementary information can be nominal, ordinal, interval or point-valued.

How then to capture these different knowledge types and use them profitably in data fusion tasks? An immediately obvious place to look for a methodology is in the field of Artificial Intelligence (AI). Broadly, AI encompasses methods based on a myriad of representation ontologies which include symbolic knowledge representation (i.e. nominal, ordinal and interval) rather than just a simple numeric information representation. Indeed, ideas of knowledge, in the sense of being able to reason and generalise about behaviour, cause and effect, are central themes in AI. A methodology is required that is capable of describing and reasoning about the physics of a system. Such a methodology must be able to incorporate descriptions of the underlying physics, it must be able to reason causally about different possible system behaviours, and it must be able to exploit real observed numerical information about the system when available. Out of all the methodologies developed to date, Qualitative Reasoning (QR) shows the most promise in satisfying these requirements simultaneously. QR methods are aimed specifically at describing the physics underlying physical situations. QR methods employ representations that allow the qualitative description of continuous systems that are more usually thought of in terms of differential equations relating physical parameters to the underlying process physics. Qualitative representations are most often defined in terms of a discrete quantity space that encapsulates important and relevant distinctions such as “increasing”, “decreasing”, “zero", or “positive”. QR begins with a qualitative description of a basic mathematical calculus. To this physical constraints are applied delineating relationships between parameters. QR represents imprecision by the range of possible predicted parameter values. Of key importance, QR guarantees that the true parameter value is contained within the set of predicted parameter values. In addition, work has shown how numerical information may be exploited in QR through the use of landmarks (Kuipers and Berleant, 1988). However, for our purposes the most important aspect of QR methods is the fundamentally different modelling philosophy employed when compared to conventional quantitative or expert system techniques. Conventional techniques make the implicit assumption that the world can only evolve according to a stated model, that is, an event is not deemed possible unless it is specifically modelled. In contrast, QR methods entertain the possibility of any event occurring unless it is specifically excluded by a constraint. Broadly, conventional methods apply a “model inclusion” principle, whereas QR methods apply a “model exclusion” principle.

1.4 Qualitative Reasoning

A system is some subset of the entire world whose behaviour and interaction with the rest of the world can be described or modelled. Conventional models are based on sets of differential equations from which the dynamic behaviour of important variables or parameters of the system can be inferred. Model builders are often faced with incomplete and uncertain knowledge of the system under investigation and often it is not possible to fully specify the functional relationships between the system parameters in which case (Davis, 1997):

Quantitative analysis should not even be attempted. In many systems and domains, collecting the necessary data is prohibitively expensive or constructing and validating reliable models is very difficult.
Figure 1.1: Qualitative models are abstractions of differential equations (reproduced from (Kuipers, 1994)).

Qualitative Reasoning is about reasoning with abstract representations of a system. Many interpretive methodologies reason with precise quantitative information before abstracting the results into a qualitative representation. Qualitative Reasoning exists for when such precise models are unavailable and the QR approach involves the early abstraction of information prior to reasoning (see Figure 1.1).

Quantitative model-based systems can be made easier to build and more reliable by supplementing them with abstract (i.e. qualitative) information (Davis, 1997). Qualitative information can be used to complete a model of the system so that the behaviour of the system can be predicted at some level of abstraction. Alternatively, qualitative information can be used to complement the quantitative model; to verify that the quantitative model parameters generate expected qualitative behaviour (e.g. (Parsons and Saffiotti, 1996)), to verify that simplifications introduced to the quantitative model do not invalidate the results, and to manage quantitative inference by identifying problem areas that might warrant more detailed analysis. Flynn (Flynn, 1988) demonstrates an early example of qualitative reasoning to manage quantitative fusion of information from Polaroid ultrasonic ranging systems and an infrared programmable proximity detector. These “expert-system” rules inform when data from either sensor is valid and which sensor to rely upon when conflicting information is obtained:

1. If the sonar reading is greater than the maximum range for the near-infrared sensor, then ignore the near-infrared output.

2. If the sonar reading is at its maximum value, the real distance is greater.

3. Whenever the infrared sensor detects a ... discontinuity (a change from no detection to detection), and the associated sonar reading is less than 10 ft, then a valid depth discontinuity has been detected.

These, simple rules qualify the operation regions of the sensors and direct the quantitative fusion algorithms accordingly.
1.4 Qualitative Reasoning

Qualitative and semi-quantitative representations and reasoning has been the focus of Qualitative Reasoning over the past 10 years. Qualitative Reasoning (QR) is a development of Expert System Theory and is involved both in reasoning about natural phenomena and everyday physics towards systems which exhibit common-sense in the performance of scientific, economic, ecological or engineering problem solving (Struss, 1997). The more conventional Expert System encodes surface knowledge of the behaviour of the system only based on empirical associations. The rule-based representation "if-then-else" rules they employ often contain implicit and redundant information which is specialised and inflexible. Consequently, expert system databases are often extremely large and cumbersome. Although these basic expert systems are able to trace the rules they used to assert a result, they are unable to understand the mechanisms underlying the object or process under investigation (Struss, 1997). Understanding is the ability to assert the causal relationships between observed events (Iwasaki, 1997) (i.e. "reasoning from first principles" (Cohn, 1989)) which contrasts with the expert system approach of merely chronicling the sequence of events. Understanding the behaviour of a system is crucial for an intelligent agent to adapt and respond appropriately to novel situations, for reducing the complexity of vast search spaces and for interpreting vast amounts of data according to some user-friendly ontology. These tasks require reasoning about the structure and function of the system (conceptualisation) using comprehensive (but not necessarily precise) descriptions of the system physics. QR contributes to each task as follows:

- **Novelty.** Deep knowledge of the system physics is more general than surface knowledge encoded by ordinary expert systems. QR is able to build new model instances for novel environment contexts.

- **Search.** A single qualitative behaviour captures an infinite number of numerically simulated behaviours. Thus, qualitative approaches can generate efficiently a broad view of all possible behaviours of the system: "a broad-brush picture of a large space of possibilities to give us a quick overall shape of the space" (Iwasaki, 1997). This property is useful for sensor management for guiding optimal sensor placement when a numerical search of the complete space of possible utility/action pairs would otherwise be required (in contrast with brute force search at a single scale of representation (Manyika and Durrant-Whyte, 1994)).

- **Interpretation.** Numerical analysis may produce vast amounts of numerical information which must be aggregated if it is to be interpreted. Qualitative reasoning supplies a high-level summary of this information (e.g. (Yip and Zhao, 1996)).

- **Accuracy.** Precise predictions require precise models which are often inaccurate. Thus, qualitative approaches offer an accurate, if not precise description of the system behaviour. Sometimes, when all information is complete, data transformations can render only a qualitative interpretation possible (e.g. scale-space reasoning using gauss smoothing filters (Witkin, 1983)).

Thus, QR is dedicated to finding mechanisms for drawing inferences about physical phenomena using incomplete information in order to increase the effectiveness, efficiency and competence of computer systems.

Three approaches to reasoning about physical phenomena dominate the QR literature and they differ only in the ontological primitives they use for describing physical systems. Their mechanisms for inferring behaviour based on reasoning about quantities share a common abstract basis (Struss, 1997). Qualitative Process Theory (QPT) (Forbus, 1984) describes the processes acting within a
system and how these processes generate other processes. Thus, the process which converts the process of current exchange to the process of heat flow is the primitive which describes the heating of water in a kettle. Component orientated ontologies (for example, the ENVISION system (deKleer and Brown, 1984)) use components as basic primitives; it is the turning of a switch which eventually causes the kettle to boil. The basis for QSim (Kuipers, 1994) is the underlying mathematical description of the system and is very much a parameter-centred approach which comprises constraints based on the abstract descriptions of differential calculus and algebraic functions. QSim essentially makes explicit the basic system representation for dealing with quantities underlying all three approaches to qualitative reasoning.

Two themes link ENVISION, QPT and QSim: the first theme involves determining the behaviours of a system by simulating the quantities of system variables from initial conditions. The second theme is that all systems deal with interval abstractions of the real numbers which represent the conceptual values of the system parameters and the differential equations linking these parameters. The interval representation provides a convenient method for conceptualising primitives of the system (both ordinal and value measures) which can incorporate incomplete knowledge required to generate a comprehensive, if not precise description of the system. Thus, the interval representation is particularly useful in engineering systems for which a detailed algebraic description of the underlying physics is difficult to obtain. The basic building blocks of ENVISION, QPT and QSim are summarised by (Struss, 1997):

- **Constituents** (i.e. processes, components) are described by local variables and their values which are inferred from constraints on their derivatives.

- **Qualitative values** capture ordinal, interval or point relationships between quantities. A *qualitative value* is an open interval of the reals or a single real value. The *quantity space* of a parameter is the set of possible qualitative values the parameter can take and is generally an exhaustive and mutually exclusive set defined over the extended real line (i.e. \( \mathbb{R} \cup \{-\infty, \infty\} \)).

- **Qualitative constraints** (i.e. confluences, or the Qualitative Differential Equation (QDE)) describe consistent value assignments for sets of parameters. QDE’s are abstract representations of the system ordinary differential equations (ODE).

- **States** are consistent assignments of qualitative values to variables according to the constraints. A *simple state* \( Q \) of a set of system parameters \( \{W\} \) is a vector of qualitative values \( Q = \langle q_0, \ldots, q_n \rangle \), in which the parameter \( W_i \) takes the qualitative value \( q_i \). A *compound state* is a set of simple states. A *state* \( S \) is either a simple state or a compound state.

- **System Behaviours** are described by a sequence of states.

*Qualitative Simulation* is the method of generating dynamic behaviour from qualitative differential equation (QDE) models of the physical system. Locative information is encoded as **landmarks** which denote specific values on the real number line (known or unknown) which are associated with the stationary points of the solutions to the QDE. For example, in the robot sensing domain, radar bearing angle \( \theta \) may have landmarks corresponding to the numeric values \( 0, \frac{\pi}{2}, \pi, -\frac{\pi}{2} \) denoting *front*, *left*, *behind*, and *right*. Zero is a special landmark which allows ordinal relationships to be encoded using the sign algebra +, 0 and −. Thus, \( A \) is greater than \( B \) may be encoded as \( A - B \) is positive.

In qualitative algebras the qualitative value and time derivative of variables are encoded explicitly. \( Qmag(A) \) represents the magnitude of variable \( A \) and \( Qdir(A) \) represents the value of its
first derivative. The set of landmarks arranged in increasing numeric order for each variable is the variable’s quantity space and Qmag can take any landmark or any interval bounded by landmarks as values. Thus, Qmag(θ) = behind and Qmag(θ) = (front left) are valid assignments. Qdir, representing the first derivative, can take any ordinal sign value. Qdir(θ) = 0 states that θ is constant and Qdir(θ) = + states that θ is increasing with time.

These ideas have been implemented in the QSim (i.e. Qualitative Simulator) algorithm (Kuipers, 1994). QSim takes a set of QDEs and an initial instantiation of the qualitative parameters and generates the qualitative behaviour of the parameters over time. At each time step the QDEs are used to instantiate the qualitative values of the parameter derivatives. The Qdir values indicate whether there is subsequently a change in the qualitative value of the parameter. For example, if Qmag(θ) = (left frontal) (i.e. θ lies between 0 and $\frac{\pi}{2}$) and Qdir(θ) = – (i.e. θ is decreasing) then at the next qualitative time step Qmag(θ) = frontal. Ambiguities in the assignments of values to variables lead to branching during behaviour generation. The tree structure of successor states generated using QSim is called the environment.

The QSim algorithm is sound but incomplete. Soundness guarantees that the actual system behaviour is included in the environment and this is proved by the Guaranteed Coverage Theorem (Kuipers, 1994). Incompleteness is a natural consequence of abstraction for which the QSim state description includes behaviours which are inconsistent with all ODE instances of the QDE. Soundness guarantees that at least one state behaviour in the environment is necessarily true if our knowledge of the physical processes is necessarily true. Thus, if a state of affairs exists in all state behaviours then that state of affairs is necessarily true. Similarly, if a state of affairs is not observed in any behaviour then this state of affairs is necessarily not true. If such a state of affairs appears in only a proper subset of the behaviours then it is deemed to be possible. Soundness and incompleteness are properties of the representation itself. We will define a context (model) specific qualifier which refers to the information content in the qualitative description. Suppose we want to know whether some state of affairs C is a necessary consequence of all behaviours consistent with a qualitative description Q, whether a single state or a subset of the environment. If only one state of affairs is possible then the qualitative description Q is said to be maximally informative. If some, but not all states of affairs are possible then Q is said to be informative. However, if all states of affairs are possible then Q is said to be uninformative.

1.4.1 Qualitative Representations

The two key issues to choosing an appropriate qualitative representation are composition and resolution. Composition is concerned with the ability to combine representations for different aspects of a phenomenon or system to create a representation of the phenomenon or system as a whole. Compositionality is an issue because one goal of qualitative physics is to formalise the modelling process itself. This thesis is not concerned with this issue.

Resolution concerns the level of information detail in a representation. Resolution is an issue because one goal of qualitative reasoning is to understand how little information suffices to draw useful conclusions. Low resolution information is available more often than precise information but conclusions drawn from low-resolution information are often ambiguous. Higher resolution information is often needed to draw subtle distinctions. The lowest resolution representation is the status abstraction, which represents a quantity by whether or not it is “normal” and can distinguish between something working or not. The next step in resolution is the sign algebra, which represents
continuous parameters as either +, 0 or – according to whether the sign of the underlying continuous parameter is positive, zero or negative. The sign representation can be extended by introducing landmarks, which are the critical values of the QDE or specific observed events, to provide additional resolution. This thesis is concerned with reasoning and combining information described at different levels of resolution.

The quantity-space is an exhaustive disjoint partition of the real-line. Partition boundaries are landmarks and the intervals bounded by landmarks are termed regions. The qualitative algebra refers to a specific partitioning of a quantity-space. Many different but compatible algebras may be in use at any one time and it may be necessary to convert between them.

The ordinal relationships between landmarks can be supplemented with (possibly incomplete) knowledge of their quantities in the form of bounding intervals (Kuipers, 1994). Q2 (Qualitative and Quantitative), Q3 and NSIM (Numerical Simulation) are semi-quantitative reasoners implemented as extensions to QSim (Kuipers and Berleant, 1988). They use interval arithmetic (Moore, 1979) to infer and to tighten bounds on value-denoting terms. Q2 (Kuipers and Berleant, 1988) uses the mean value theorem and known landmark quantities to bound the values of a parameter between landmarks. User specified static envelopes provide further bounds on the values of possible monotonic functions. Q3 (Berleant, 1991) allows the insertion of extra landmarks into qualitative intervals to decrease the effective step-size and strengthen inference.

To illustrate the difference between Q2 and Q3, consider a robot which accelerates with an unknown but bounded value $A \in (0,4) ms^{-2}$ from rest. After $T = 2$ secs the robot’s velocity is $V = \int_{0}^{2} A dt \in (0,8)$ and therefore, Q2 would estimate that after 2 secs the robot would have moved $D = \int_{0}^{2} v(t)dt \in [0,8] \times 2 = (0,16)m$. By inserting a landmark at time 1 sec, Q3 infers that after 1 sec the robot attains a velocity $V_{[0,1]} = \int_{0}^{1} A dt \in (0,4)$ and has moved a distance $D = \int_{0}^{1} V dt \in (0,4)m$. In the following 1 sec the robots velocity remains within $V_{[1,2]} = V_{(0,1)} + \int_{1}^{2} A dt = (0,8)$. The robot therefore moves a further distance $D \in (4,8)m$. Thus, Q3 infers a total distance travelled $D_T \in (0,4) + (0,8) = (0,12)$ which is tighter than that predicted by Q2. NSIM (Kay and Kuipers, 1993) uses an extended Runge-Kutta method to simulate the numerical solutions to an extremal system of ODEs (the so called dynamic envelopes). NSIM generally produces tighter bounds than either Q2 or Q3 but can be numerically less stable. These semi-quantitative methods have also been extended to fuzzy inference (i.e. FUSSIM (Shen and Letch, 1993)).

Ward (Ward et al., 1989) offers two interpretations for an interval meaning in relation to some predicate $F$:

- **Universal Quantification (\forall)** asserts that the predicate is true for all members of the interval: $F(I) \iff (\forall i \in I) F(\{i\})$.

- **Existential Quantification (\exists)** asserts that the predicate is true for at least one member of the interval: $F(I) \iff (\exists i \in I) F(\{i\})$.

Although universal quantification entails existential quantification different tasks can demand that the interpretation be made explicit. Existential and universal quantifications are used for estimation and inference respectively. An estimate of the value of some parameter is an interval $I$ which contains the actual value $x_T$ of the parameter. An estimator $F_E(I)$ is true if and only if $x_T \in I$ and since $x_T$ is single-valued then $F_E(I) \iff (\exists i \in I) i = x_T$. The aggregate of a set of interval estimates $\{I_i\}$
is their intersection (Moore, 1979; Gao and Durrant-Whyte, 1994; Knippers, 1994):

\[ I_{0} \cap I_{1} = \bigcap_{i} I_{i}. \quad (1.1) \]

Interval-based inference rules are understood to be universally quantified over the antecedent intervals and the quantitative mapping set and existentially quantified over the consequent intervals. Thus, \( F(I_{0}, G, I_{1}) \equiv G(I_{0}) = I_{1} \) which describes the multi-valued mapping of the interval \( I_{0} \) onto the interval \( I_{1} \) should be understood to mean:

\[ G(I_{0}) = I_{1} \iff (\forall i_{0} \in I_{0}, g \in G)(\exists i_{1} \in I_{1}) g(i_{0}) = i_{1}. \]

In summary, Qualitative Reasoning holds the key to model-based reasoning at various scales and to reconciling (aggregating) multiple sensor views when operating at different scales. System observations are noisy and the next section considers how noisy estimates can be fused within the QR framework.

### 1.5 Reasoning Under Imprecision and Uncertainty

Estimates of stochastic systems inferred from noisy observations indicate an experimentally determined propensity for the system to be in some specified state (the true state). Uncertainty has been accommodated in a number of QR systems. Doyle (Doyle and Sadeh, 1989) interprets the state dynamics transition graph as a Markov chain and assigns probabilities to each of the transitions in order to capture the plausibility of a path. Similarly, Decoste’s DATMI (i.e. Dynamic Across Time Measurement Interpretation) (DeCoste, 1990) uses path probability estimates to determine the most likely operation of a system. New observations are compared with this estimate and a large deviation indicates a system fault. More recently Brajnik (Brajnik, 1997) describes a Monte-Carlo method for deriving statistical state transition values by analysing the space of trajectories by qualitative differential equations. Another important system which integrates qualitative reasoning under uncertainty is Hellerstein’s NIMF system (Hellerstein, 1992). NIMF uses point estimation methods to obtain, to within a prespecified confidence level, the monotonic trend for the median of a symmetrically distributed noise parameter. None of these methods address the problem of data fusion.

We examine a number of approaches to reasoning under uncertainty and conclude that none of these offer an adequate framework for reasoning when domain and uncertainty models are incomplete.

#### 1.5.1 The Interval Intersection Filter (IIF)

The Interval Intersection filter fusion rule for noise-free estimation exploits the noise-free intersection rule of Equation 1.1 in noisy situations (Gao and Durrant-Whyte, 1994). This method accommodates the uncertainty of a random observation within the interval estimate. The interval is centred on the observation and the true state is assumed to lie within the interval with a certain probability. However, the IIF does not exploit information about the degree of belief of an estimate and therefore, outliers are problematic. Also the convergence rate is slow as IIF effectively ignores an observation if its estimation interval is a superset of the current estimate. In which case, the new observation does not contribute to the degree of belief of the predicted estimate. These problems are discussed in more detail in Chapter 5.
1.5.2 Classical Bayesian Estimation

The usefulness of probabilistic semantics is a motivation for considering extending Bayesian model-based estimation to qualitative representations. In probability theory, the probability density function (pdf) not only indicates the preference ordering of states but also quantifies false-positive decision frequencies for decisions based on this preference ordering. Quantitative model-based Bayesian estimation is well established. However, maintaining consistent probabilistic semantics when the process model is incomplete in propositional systems is problematic. In this section, we investigate these problems by attempting to construct an analogy to point Bayesian estimation.

Bayesian estimation is a straightforward recursive application of Bayes’ rule. Let \( z_{t}^{b} \) denote a finite sample sequence of \( t + 1 \) observations indexed \( a \) to \( b = a + t \). The sequence \( z_{t}^{b} \) may include observations from different sensors. If \( x \) is the true state of the system at time \( t \), and \( f_{t} \) and \( g_{t} \) are the conditional pdfs for \( x \) and for the observation \( z_{t} \) respectively at time \( t \), then by Bayes’ rule:

\[
f_{t}(x \mid z_{1}^{t}) = K f_{t}(x \mid z_{1}^{t-1}) g_{t}(z_{t} \mid x, z_{1}^{t-1})
\] (1.2)

When the observations are conditionally independent and the (possibly stochastic) process is Markovian then:

\[
g_{t}(z \mid x, z_{1}^{t-1}) = g_{t}(z \mid x)
\]

\[
f_{t}(x \mid z_{1}^{t-1}) = \sum_{y \in Y} f_{t-1}(y \mid z_{1}^{t-1}) h_{t-1}(x \mid y)
\]

where the apriori conditional distribution \( h \) is the process model. Thus:

\[
f_{t}(x \mid z_{1}^{t}) = K g_{t}(z \mid x) \sum_{y \in Y} f_{t-1}(y \mid z_{1}^{t-1}) h_{t-1}(x \mid y).
\] (1.3)

Analogously, a qualitative Bayesian estimator can be obtained by Bayes’ rule. The probability that the true state \( x \) belongs to the qualitative state \( Q \) at time \( t \) is:

\[
P_{t}(Q \mid z_{1}^{t}) = K P_{t}(Q \mid z_{t}) G_{t}(z_{t} \mid Q, z_{1}^{t-1}).
\] (1.4)

where \( P_{t}(Q \mid z_{1}^{t-1}) \) is the probability that the true state is in \( Q \) according to observations \( z_{1}^{t-1} \) and \( G_{t}(z_{t} \mid Q, z_{1}^{t-1}) \) is the likelihood of observation \( z_{t} \) at time \( t \). A strong structural analogy exists between this equation and Equation 1.2. However, whereas the conditional terms \( g \) and \( h \) in Equation 1.3 make use of the Markov property, that the observations are conditionally independent given the true state of the system, a simplifying assumption of this nature does not exist for Equation 1.4. It is not possible to assert the following:

\[
P_{t}(Q \mid z_{1}^{t}) = K G_{t}(z \mid Q) \sum_{R \in Y-Q} P_{t-1}(Q \mid z_{1}^{t-1}) H_{t-1}(Q \mid R).
\]

To illustrate why this is the case, suppose we do assert this identity then we obtain the following likelihood ratio for two states \( Q \) and \( Q' \) when \( R \rightarrow Q \) is a unitary mapping for all \( t \):

\[
\frac{P_{t}(Q \mid z_{1}^{t})}{P_{t}(Q' \mid z_{1}^{t})} = \frac{P_{t}(Q)}{P_{t}(Q')} \prod_{i=1}^{t} \frac{H_{t-1}(Q_{i} \mid Q_{i}^{t-1})}{H_{t-1}(Q'_{i} \mid Q'_{i}^{t-1})} \prod_{i=1}^{t} \frac{G_{t}(z_{i} \mid Q_{i})}{G_{t}(z_{i} \mid Q'_{i})}.
\] (1.5)

The first term on the right hand side of Equation 1.5 is the apriori distribution of initial conditions and, unless at its extremes, this value becomes insignificant after many observations have been
1.5 Reasoning Under Imprecision and Uncertainty

assimilated. The second term is the transition probability; the probability that a specific path is followed independent of the observations. Unlike the initial condition term, the transition probabilities recur and contribute to the final probability estimate for all time \( t \). A probabilistic interpretation of \( \frac{P_t(Q|z_1^t)}{P_t(Q|z_2^t)} \) is meaningful only when the transition and the observation probabilities are meaningful. Exact values for these probabilities are not required provided that the apriori specified false-positive decision error rate is not exceeded.

Suppose a stationary system (i.e. \( H_t \) and \( G_t \) are time independent) can be in one of three states \( Q_0, Q_1 \) or \( Q_2 \) and \( Q_0 \rightarrow \{Q_0\}, Q_1 \rightarrow \{Q_0, Q_1\} \) and \( Q_2 \rightarrow Q_2 \) describe a single time step transition of the system, so that \( H(Q_0|Q_0) = 1, H(Q_2|Q_2) = 1 \) and \( H(Q_0|Q_1) + H(Q_1|Q_1) = 1 \). Using Bayes' rule we can estimate the probability of \( Q_0, Q_1 \) and \( Q_2 \) at time \( t \) given such values at time \( t-1 \) and estimates from the observation at time \( t \) (the unconditional prior are ignored for the reasons given above):

\[
\begin{align*}
P_t(Q_0|z_1^t) &= [H(Q_0|Q_1)P_{t-1}(Q_1|z_1^{t-1}) + P_{t-1}(Q_0|z_1^{t-1})]G(z|Q_0), \\
P_t(Q_1|z_1^t) &= H(Q_1|Q_1)P_{t-1}(Q_1|z_1^{t-1})G(z|Q_1), \\
P_t(Q_2|z_1^t) &= P_{t-1}(Q_2|z_1^{t-1})G(z|Q_2).
\end{align*}
\]

The values of \( H(Q_0|Q_1) \) and \( H(Q_1|Q_1) \) are constrained by \( H(Q_1|Q_1) + H(Q_0|Q_1) = 1 \). Thus:

\[
\begin{align*}
P_t(Q_0|z_1^t) &= [(1-H(Q_1|Q_1))P_{t-1}(Q_1|z_1^{t-1}) + P_{t-1}(Q_0|z_1^{t-1})]G(z|Q_0) \\
P_t(Q_1|z_1^t) &= H(Q_1|Q_1)P_{t-1}(Q_1|z_1^{t-1})G(z|Q_1) \\
P_t(Q_2|z_1^t) &= P_{t-1}(Q_2|z_1^{t-1})G(z|Q_2)
\end{align*}
\]

In the non-trivial case, for some constant \( \delta, 1 - H(Q_1|Q_1) = \delta > 0 \) and:

\[
H(Q_1|Q_1) = 1 - \delta < 1
\]

When the true state lies on the landmark separating \( Q_1 \) and \( Q_2 \) the estimate should not favour \( Q_1 \) nor \( Q_2 \). However, when Condition 1.8 holds then we can show by recursion that there is a persistent bias in the system such that, when many observations have been assimilated, the Bayesian estimator will prefer \( Q_2 \). Firstly, we obtain a recursive form for our static problem:

\[
\log \left( \frac{P_t(Q_1|z_1^t)}{P_t(Q_2|z_1^t)} \right) = \sum_{i=1}^{t} \log \left( \frac{(1-\delta)G(z_1|Q_1)}{G(z_1|Q_2)} \right)
\]

As \( \lim t \to \infty \) then, by the law of large numbers:

\[
\log \left( \frac{P_t(Q_1|z_1^t)}{P_t(Q_2|z_1^t)} \right) = \sum_{i=1}^{t} \log \left( \frac{(1-\delta)G(z_1|Q_1)}{G(z_1|Q_2)} \right) = t \left( E_z \left( \log \frac{G(z_1|Q_1)}{G(z_1|Q_2)} \right) + \log(1-\delta) \right)
\]

Ideally, we would like to assign observation probabilities independent of the model of the system since checking the probability consistency for each of a possibly large number of model transitions would be computationally expensive. We would like to assign unbiased probability estimates from observations and guarantee that, through conditional assignment and further observation assimilation, the estimate would remain unbiased. Unfortunately, this is not the case for Bayesian estimation. Suppose we insist that each observation is unbiased then, when the true state of an observable coincides
with the landmark, a single observation should, on average, favour neither \( Q_1 \) nor \( Q_2 \). Thus:

\[
E_z \left( \log \frac{G(z|Q_1)}{G(z|Q_2)} \right) = 0.
\]

However, Equation 1.9 then becomes:

\[
\lim_{t \to \infty} \log \left( \frac{P_t(Q_1|z^t)}{P_t(Q_2|z^t)} \right) = t \log(1 - \delta)
\]

and our transition burdened system is biased in favour of \( Q_2 \) since \( \log(1 - \delta) < 0 \). So, in general simple conservative transition probability estimates for unbiased estimation cannot be found.

Other problems with the probabilistic approach include the assignment of conservative pdfs when an exact noise model is not known, and the problem of correlated sources which potentially involves the need to obtain, store and process \( 2^N \) conditional probability values for systems comprising \( N \) propositions. All these problems make the creation of a probabilistic qualitative model reasoner extremely difficult.

### 1.5.3 Upper and Lower Probabilities

The famous Dempster-Shafer Theory of Evidential Reasoning was motivated by the need to represent unknown, but constrained, probability distributions (e.g., (Dempster, 1967)). Suppose that an unknown probability distribution on the disjoint partition \( \Theta = \{a, b, c\} \) is one of a set of probabilities with known parametric forms \( \mathcal{P} \). Suppose further that the range of \( \mathcal{P} \) constraints \( P \) thus:

\[
0.1 \leq P(a) \leq 0.7, \quad 0.1 \leq P(b) \leq 0.8, \quad 0.1 \leq P(c) \leq 0.2
\]

Dempster proposed representing the unknown probability distribution by a measure \( \mu \) which commits probability to subsets of \( \Theta \). When \( P^* \) and \( P_* \) are the upper and lower bounds on the unknown probability distribution then:

\[
P^*(A) = \sum_{B \in A} \mu(B) \quad \text{and} \quad P_*(A) = \sum_{B \in A} \mu(B)
\]

The example above becomes:

\[
\mu(\{A\}) = 0.1, \quad \mu(\{B\}) = 0.1, \quad \mu(\{C\}) = 0.1, \quad \mu(\{A, B\}) = 0.6, \quad \mu(\{B, C\}) = 0.1.
\]

The combination of an ensemble of (independent) observations should be consistent with Bayes rule for any true state sequence \( x \):

\[
(\forall x_i \in Q) \ P(x_i^t | z^t) \propto \prod_{i=1}^{t} P(z_i | x_i).
\]

Thus:

\[
P^*(Q) \geq \max_{x_i^t \in Q^t} P(x_i^t | z_i^t) \quad \text{and} \quad P_*(Q) \leq \min_{x_i^t \in Q^t} P(x_i^t | z_i^t).
\]

Consider a three state quantity-space \( Q_1 = (-\infty, -1) \), \( Q_2 = (-1, 1) \) and \( Q_3 = (1, \infty) \) and observations subject to Gaussian random errors with standard error of 1. For every observation \( z \in Q_1 \) there are two positions \( x, x' \in Q_1 \) such that for all \( x'' \in Q_2 \), \( P(z | x') \leq P(z | x'') \leq P(z | x) \). So, a preponderance of observations in \( Q_1 \) will not produce a clear preference ordering between \( Q_1 \) and
1.5 Reasoning Under Imprecision and Uncertainty

$Q_2$ since $[P_*(Q_1 \mid z^j), P_*(Q_1 \mid z^j)] \cap [P_*(Q_2 \mid z^j), P_*(Q_2 \mid z^j)] \neq \emptyset$. Clearly, for a symmetric noise distribution, $Q_1$ should be preferred. Constructing a preference ordering based on either $P^*$ or $P_*$ alone would be meaningless.

Although $P^*$ and $P_*$ are guaranteed to be upper and lower probabilities over inference, many problems arise when combining such estimates. Dempster’s famous rule of combination fails to maintain the upper and lower probability interpretation even for binary quantity-spaces. Subsequently, it is no longer fashionable in the literature to interpret the Dempster-Shafer plausibility and belief functions\(^2\) as upper and lower probabilities.

1.5.4 The Point Estimation Paradigm

We demonstrate the point estimation and data fusion paradigm using, perhaps, the most popular method: the Kalman filter. The reason for presenting, in detail, the Kalman filter theory will become clear when the main contribution of this thesis, the Qualitative filtering (QF) paradigm, is presented for reasoning under imprecision. The QF shares a number of salient concepts of the Kalman filter.

The Kalman filter (KF) is the classical low-pass filtering solution to the problem of estimating the state of a linear dynamic system in the presence of noise (Morrell and Stirling, 1991) and it is the optimal linear, recursive, minimum variance point estimation and data fusion method. If it is assumed that the error distributions are Gaussian, then each Kalman estimate defines a Gaussian distribution representing the true pdf conditioned on the measurement sequence (Jaswinski, 1970). It is critical to observe, however, that the assumptions underlying this interpretation are not necessary for the rigorous derivation of the Kalman filter, and such Gaussianity assumptions generally do not hold in practice: the Kalman filter is deemed to be non-parametric as precise models of the pdf are not required.

The key requirement of the Kalman filter is the ability to apply process and observation models to transform means and covariances. In particular, the estimate of the current state of the system must be obtained by prediction from a previous estimate before a new observation can be assimilated. If the state estimate is not in the same coordinate system as the observation estimate, then the state estimate must be transformed to observation coordinates (see Figure 1.2). The Kalman filter estimates the state of a system by linearly combining a prediction of the state with a set of measurements which are obtained by a suite of sensors. The uncertainty in the estimate is expressed by a covariance matrix whose value evolves through time and reflects the expected mean-squared-error of the estimated state at any given time.

The KF is one of the most widely used estimation algorithms. The great popularity of the KF arises from a number of important and well-known properties. These include:

- **Optimality.** The filter is the optimal linear mean-squared-error estimator. Given a linear update rule, no other estimator yields an estimate which has a smaller mean-squared-error.

- **Recursiveness.** In each iteration only the current sensor observations are used and it is not necessary to maintain and re-process previous observations. This minimises the storage necessary and allows estimates to be made on-line with modest computational resources.

- **Linearity.** The state estimate is obtained through a weighted linear sum of information from the predictions and the observations. This provides a simple and efficient means of combining

---

\(^2\)Plausibility and belief are the generalisations of $P^*$ and $P_*$ to nonstatistical problems.
Figure 1.2: The Kalman filter (KF) cycle. A new estimate \( \hat{x}(k) \) of \( x \) is inferred from an observation \( z(k) \) at time \( k \) and is combined with an estimate of \( x \) obtained by prediction from information gathered until time \( k-1 \) to give \( \hat{x}(k | k) \).

- **Covariance Modelling.** The uncertainties associated with the system behaviour and observations are explicitly modelled using random variables with specified means and covariances. This provides a confidence measure in the estimates provided by the algorithm.

Figure 1.2 shows the basic operation of the filter. It is first initialised - an initial state of the system and an associated covariance matrix are specified. The filter then operates recursively, performing a single cycle each time a new set of observations becomes available. Each iteration propagates the estimate from the time the last observation was obtained to the current time. The propagation process consists of two stages: *prediction* followed by *update*. Suppose the state of the system has been estimated at time \( k-1 \) and new observations become available at time step \( k \). Prediction calculates the expected state of the system at \( k \) given the estimate at \( k-1 \) and a process model which describes the evolution of the system. In general the prediction contains errors due to unmodelled dynamics and unknown control inputs. To reduce or eliminate these errors the prediction is updated or corrected. The estimate is formed from a linear combination of the prediction and an observation-dependent correction term. The update is conducted in such a fashion that the mean-squared estimation error is minimised.

The entire state of the physical system at any time step \( k \) is expressed using an \( n_T \) dimensional state space vector \( x_T(k) \). The components of this vector may include the position and velocity of a vehicle, various internal states and possibly relevant environmental effects (such as ambient temperature). The subscript \( T \) denotes the fact that this is the state of the true system. In practice the structure and form of the true system is not known and approximations must be used. However, for simplicity, it is often assumed that the filter uses the same state space and process model as the true system.

The state of the system changes through time due to its dynamics. This time evolution is described using the discrete-time process model:

\[
x_T(k) = f_T[x_T(k-1), u_T(k-1), v_T(k), k-1].
\]  
(1.10)
1.5 Reasoning Under Imprecision and Uncertainty

The state space “velocity” is a function of four parameters: current state, control inputs, the process noise and the current time. The first two terms are self-explanatory. The third is the process noise vector, $\mathbf{v}_T(k)$. This random vector contains all the perturbations which act on the system which are not described deterministically. These include unmodelled dynamics and noise corrupted (or unmeasured) control inputs. Finally, the current time $k$ is included because the process model might be time-varying.

It is assumed that the process noise is a zero-mean random variable and has covariance $\mathbf{Q}_T(k)$. It is further assumed that the process noise at any time is independent of the process noises or the state of the system which have occurred at any previous time:

$$E[\mathbf{v}_T(i)\mathbf{v}_T^T(j)] = \mathbf{Q}_T(i)\delta_{ij},$$
$$E[\mathbf{v}_T(i)\mathbf{x}_T^T(j)] = \mathbf{0}$$

where $\delta_{ij}$ is the Dirac delta function.

A suite of sensors are used to measure quantities which are related to the pose of the system. The measurements acquired at time $k$ are aggregated into the $m$-dimensional observation vector $\mathbf{z}(k)$. The values are related to the state of the system according to the sensor model:

$$\mathbf{z}(k) = \mathbf{h}_T[\mathbf{x}_T(k), \mathbf{u}_T(k), \mathbf{w}_T(k), k].$$
(1.11)

This is a function of the true state of the system, the control inputs and the time index. The observation noise $\mathbf{w}_T(k)$ encompasses all the unmodelled effects which act on the observations, but not on the underlying state of the system itself. For example, disturbances of this type arise from thermal noise in components of sensor circuits and the discretisation of analogue to digital converters. It is assumed to be a zero-mean random variable with variance $\mathbf{R}_T(k)$ which is independent of the observation noises at all previous time steps:

$$E[\mathbf{w}_T(i)\mathbf{w}_T^T(j)] = \mathbf{R}_T(i)\delta_{ij}$$

and independent of the process noise:

$$E[\mathbf{v}_T(i)\mathbf{w}_T^T(j)] = \mathbf{0}.$$

The filter estimates the system state using the process model, the observation model and a sequence of observations. The inferred value of $\mathbf{x}_T(i)$, using all observations up to and including time step $j$ (or $\mathbf{Z}_i$ where $\mathbf{Z}_i = \{\mathbf{z}(1), \ldots, \mathbf{z}(j)\}$), is $\hat{\mathbf{x}}(i | j)$. The covariance of the inferred value is defined as follows. Defining the estimation error as:

$$\tilde{\mathbf{x}}(i | j) \triangleq \mathbf{x}_T(i) - \hat{\mathbf{x}}(i | j),$$

so that the expected error is zero:

$$E(\tilde{\mathbf{x}}(i | j)) = E(\mathbf{x}_T(i) - \hat{\mathbf{x}}(i | j)) = 0$$
(1.12)

the covariance of $\tilde{\mathbf{x}}(i | j)$ is:

$$\mathbf{P}(i | j) = E[\tilde{\mathbf{x}}(i | j)\tilde{\mathbf{x}}^T(i | j) | \mathbf{Z}_i].$$

---

3This assumption is made for convenience. The filter can be generalised to include noise terms which have a non-zero mean and are correlated through time and with respect to the system state or observation noises.

4For stochastic processes the true state $\mathbf{x}_T(i)$ is a random variable.
1.5 Reasoning Under Imprecision and Uncertainty

Consider a single iteration which propagates an estimate from time step \( k - 1 \) to time step \( k \). At \( k - 1 \) the estimate of the system is \( \hat{x}(k - 1 \mid k - 1) \) with covariance \( P(k - 1 \mid k - 1) \). The new estimate at \( k \) is \( \hat{x}(k \mid k) \) with covariance \( P(k \mid k) \). The filter accounts for this change through the prediction step. Using the discrete-time process model (i.e. Equation 1.10), the state of the system at \( k \) is predicted from the estimate at \( k - 1 \). The predicted state vector \( \hat{x}(k \mid k - 1) \) is:

\[
\hat{x}(k \mid k - 1) = E[x_T(k) \mid Z_{k-1}^1]
= E[f_T[x(k - 1 \mid k - 1) + \hat{\epsilon}(k - 1 \mid k - 1), u_T(k - 1), v_T(k - 1), k - 1] \mid Z_{k-1}^1]
\]

The covariance of the prediction is:

\[
P(k \mid k - 1) = E[\hat{x}(k \mid k - 1) \hat{x}^T(k \mid k - 1) \mid Z_{k-1}^1].
\]

The estimate is obtained by updating the prediction with the current observation. The Kalman filter uses a particular form for its update rule. The estimate is equal to the prediction plus the weighted sum of the innovation vector:

\[
\hat{x}(k \mid k) = \hat{x}(k \mid k - 1) + W(k)\nu(k).
\]

(1.13)

The innovation vector \( \nu(k) \) is defined, in this dissertation, to be the difference between the system state, \( x(k) \) and the predicted observation \( \hat{x}(k \mid k - 1) \):

\[
\nu(k) = x(k) - \hat{x}(k \mid k - 1).
\]

\( W(k) \) is the Kalman weighting matrix. It determines the degree to which the innovation affects the new estimate. Its value is chosen so that the minimum mean-squared-error in the estimate (the trace of \( P(k \mid k) \)) is minimised. The error in the estimate is:

\[
\begin{align*}
\hat{x}(k \mid k) &= x_T(k) - \hat{x}(k \mid k) \\
&= x_T(k) - [\hat{x}(k \mid k - 1) + W(k)\nu(k)] \\
&= x_T(k) - [\hat{x}(k \mid k - 1) + W(k)(x(k) - \hat{x}(k \mid k - 1))] \\
&= (I - W(k))x_T(k) - [(I - W(k))\hat{x}(k \mid k - 1) + W(k)x(k)] + W(k)x_T(k) \\
&= (I - W(k))\hat{x}(k \mid k - 1) + W(k)\hat{x}(k).
\end{align*}
\]

Taking outer products and expectations and assuming \( E[\hat{x}(k \mid k - 1)\hat{x}(k)] = 0 \):

\[
P(k \mid k) = (I - W(k))P(k \mid k - 1)(I - W(k))^T + W(k)P(k)W(k)^T.
\]

(1.14)

The gain matrix \( W(k) \) which minimises the mean-squared-error is obtained via the roots of the Jacobian of \( P(k \mid k) \):

\[
\nabla_W P(k \mid k) = -2(I - W(k))P(k \mid k - 1) + 2W(k)P(k)
\]

\footnote{In the literature it is more common to define the innovation vector in terms of the actual sensor observation: \( \nu(k) \) where the observation vector \( \hat{z}(k \mid k - 1) \) is calculated from:

\[
\begin{align*}
\hat{z}(k \mid k - 1) &= E[z(k) \mid Z_{k-1}^1] \\
&= E[h_T[x_T(k), u_T(k), w_T(k), k] \mid Z_{k-1}^1] \\
&= E[h_T[x(k \mid k - 1) + \hat{x}(k \mid k - 1), u_T(k), k] \mid Z_{k-1}^1].
\end{align*}
\]
which is:

\[ W(k) = P(k | k-1)[P(k) + P(k | k-1)]^{-1}. \quad (1.15) \]

Define \( S(k) = P(k) + P(k | k-1) \) so that \( W(k) = P(k | k-1)S(k)^{-1} \). Substituting this expression back into Equation 1.14 gives the Kalman update equations:

\[
P(k | k) = (I - W(k))P(k | k-1)(I - W(k))^T + W(k)P(k)W(k)^T
\]
\[
= P(k | k-1) - W(k)P(k | k-1) - P(k | k-1)W(k)^T
\]
\[
+ W(k)[P(k | k-1) + W(k)P(k)]W(k)^T
\]
\[
= P(k | k-1) - W(k)S(k)W(k)^T.
\]

The Kalman filter information form may be derived using the following identities:

\[
I - W(k) = [P(k | k-1) - W(k)P(k | k-1)]P(k | k-1)^{-1}
\]
\[
= [P(k | k-1) - W(k)S(k)S(k)^{-1}P(k | k-1)]P(k | k-1)^{-1}
\]
\[
= [P(k | k-1) - W(k)S(k)]W(k)[P(k | k-1)]W(k)^T
\]
\[
= P(k | k)^{-1}
\]

and, by Equation 1.15:

\[ W(k)[P(k) + P(k | k-1)] = P(k | k-1) \]

so using Equation 1.16:

\[ W(k) = P(k | k)P(k)^{-1}. \quad (1.17) \]

Substituting Equations 1.16 and 1.17 into Equation 1.14:

\[ P(k | k) = P(k | k)[P(k | k-1)]^{-1}P(k | k) + P(k | k)P(k)^{-1}P(k | k) \]

Pre and post multiplying this expression by \( P(k | k)^{-1} \) we obtain the Kalman filter covariance information form:

\[
P(k | k)^{-1} = P(k | k-1)^{-1} + P(k)^{-1}. \quad (1.18)\]

Also, substituting Equation 1.17 into Equation 1.13 and pre-multiplying by \( P(k | k)^{-1} \) we obtain the information form of the state update equation:

\[ P(k | k)^{-1}\dot{x}(k | k) = P(k | k)^{-1}[\dot{x}(k | k-1) + W(k)(x(k) - \dot{x}(k | k-1))] \]
\[
= P(k | k)^{-1}[(I - W(k))\dot{x}(k | k-1) + W(k)x(k)].
\]

Thus:

\[ P(k | k)^{-1}\dot{x}(k | k) = P(k | k-1)^{-1}\dot{x}(k | k-1) + P(k)^{-1}\dot{x}(k). \quad (1.19)\]
A covariance measure for a propagated estimate is obtained by scaling the prior covariance measure according to the gradient of the process or observation function. This produces a variance measure accurate to third order in the Taylor expansion of the propagation function. The success of the Kalman filter depends on the accuracy of its propagated estimates. Inaccurate estimates degrade filter performance and can lead to divergence. In linear systems an exact closed form solution exists. However, when the process or observation models are nonlinear there are significant problems with accurately predicting the state of the system (Kushner, 1967). In such cases the mean and covariance can only be calculated if the entire distribution of the state error is known (Julier and Uhlmann, 1997). A potentially unbounded number of parameters are required and virtually all filters which have been implemented use approximations of some kind. The simplest and most widely used approach is the Extended Kalman filter (EKF) which is accurate only to third order in the process and observation models.

In summary, the Kalman filter technique estimates the true state using quantitative models to propagate observations into a common state space. The estimate is unbiased in the sense that the ensemble expected value is identical to the true state of the system: \( E(\hat{x}(k)) = x(k) \). The Kalman filter weights individual estimates in accordance to the confidence assigned to the estimate. The confidence measure is the covariance form of the Fisher Information: \( P^{-1}(k) \). The Kalman filter exhibits the following properties:

- it is an unbiased estimator of mean for zero-mean random observation noise and stochastic process and observation models,
- it makes no assumptions of the details of the probability density functions. The first two moments (mean and covariance) are maintained independent of underlying probability densities,
- it performs data-fusion via quantitative inter-parameter association,
- it maintains a measure of confidence in the estimate (Fisher Information measure) and weights evidence according to its confidence,
- it is the minimum mean-squared-error estimator and is, in this sense, optimal,
- it can fail for highly non-linear models.

The KF suffers from a number of fundamental problems. First, and most importantly, precise models of the system are required as the Kalman filter maintains a single point-value estimate of the state of the physical reality and is unable to represent ignorance when modelling information is incomplete (Morrell and Stirling, 1991). The Kalman filter requires that the system be *observable* and have the ability to *uniquely* determine the state of a free linear dynamic system from observations of linear combinations of the output of the system in finite time (see (Bryson and Ho, 1969), page 458). The models employed in data fusion methods are used explicitly to estimate the state of the system and to provide measures of estimator performance. In essence, quantitative algorithms develop a complete internal model of a system which is taken as the "truth" by which observations are referenced. Thus, if the specified system model is inaccurate, the estimates of state and all measures of performance will be incorrect.

In practice it is often necessary to increase the amount of assumed process noise in order to account for systematic errors, numerical imprecision and any other noise source that is not accounted for explicitly. This additional noise is referred to as *stabilising noise* and is typically subsumed in
1.5 Reasoning Under Imprecision and Uncertainty

An over-estimate of the true noise covariance, $Q(k)$, is said to be conservative. The amount of stabilising noise is empirically determined by testing the filter and increasing $Q$ until it produces state covariance estimates that are consistent with the covariance of the distribution of the actual errors committed by the filter. A filter that yields consistent state estimates is said to be non-divergent. When the filter generates covariance estimates that are smaller than the true errors in its mean estimates, then the filter is said to be divergent.

Developing a precise or complete model of a system and then encoding systematic errors that encompasses all possible behaviors is not generally practical and, in many situations, there is no conservative covariance matrix which satisfies all possible behaviors. To illustrate this fact consider the following example. Two variables $x_1$ and $x_2$ are imprecisely (and noiselessly) measured and an upper bound on the systematic error is known to be unity for each variable. The covariance matrix for $(x_1, x_2)^T$ is partially specified:

$$P_{1,2} = \begin{pmatrix} 1 & C \\ C & 1 \end{pmatrix}$$

and the aim is to identify appropriate values for the cross-terms $C$ which produce appropriate systematic errors for the inferences $x_3 = x_1 + x_2$ and $x_4 = x_1 - x_2$. The observation model for our system is

$$\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

from which we obtain the covariance matrix:

$$P_{3,4} = 2 \begin{pmatrix} 1+C & 0 \\ 0 & 1-C \end{pmatrix}.$$ 

Suppose $x_1$ and $x_2$ are known to lie within some interval: $x_1 \in (y_1 - 1, y_1 + 1)$ and $x_2 \in (y_2 - 1, y_2 + 1)$. Then it is natural to assign estimates $\hat{x}_1 = y_1$ and $\hat{x}_2 = y_2$ with variances $Var(\hat{x}_1) = Var(\hat{x}_2) = 1$ in which case the inference $x_3 = x_1 + x_2$ lies within the interval $T_3 = (y_1 + y_2 - 2, y_1 + y_2 + 2)$. To accommodate $T_3$ we must have $(P_{3,4})_{1,1} = 2^2$ which requires $C = 1$ and:

$$P_{1,2} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$ 

Similarly, the inference $x_4 = x_1 - x_2$ lies within the interval $T_4 = (y_1 - y_2 - 2, y_1 - y_2 + 2)$ and to accommodate $T_4$ we must have $(P_{3,4})_{2,2} = 2^2$ which requires $C = -1$ and:

$$P_{1,2} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$ 

Clearly, there is a disagreement over the appropriate value for $C$. In general, no matrix may exist that is consistent with all inferences. In cases when consistent cross-term values can be found their dependency on the underlying process and observations models is a burden we wish to avoid.

The previous example demonstrated problems of inference for two and higher dimensional systems. Inference is also a problem for one dimensional non-linear systems as the Kalman filter is mathematically unsound and incomplete even when the model is known precisely. Consider the mapping $y = x^3$ and a noiseless observation $x = 1$ with a systematic error $\sigma_x = 1$ so that the true state (which is guaranteed to lie within 1 of the observation is covered). The EKF would predict a standard-error $\sigma_y = 3 \times 1^2 \times \sigma_x = 3$ and the EKF estimate would cover $(1^3 - 3, 1^3 + 3) = (-2, 4)$ in
1.5 Reasoning Under Imprecision and Uncertainty

$Y$ space. This estimate is far removed from the actual possible values for the true state in $Y$ which can take any value in $(0, 8)$ (This range is precisely that predicted by the IIF). The Kalman filter is incomplete as it includes values $(-2, 0)$ which are not possible and it is also unsound as it excludes those which are possible (i.e., $(4, 8)$). Thus, the Kalman filter is both unsound and incomplete in the mathematical sense and can fail in situations of high uncertainty when non-linear models are assumed (Julier and Uhlmann, 1997). Even if it were possible to account for more than the first two moments, the soundness and completeness problems would still not be solved since there is a basic flaw in the Kalman filter philosophy: it is derived under the basic assumption that the mean of its state estimate corresponds to the true state of the system. It is clear that this assumption cannot apply to both an observed parameter and its propagation via a non-linear function.

Inference is not the only problem with point estimation methods. Fusion can also lead to over-confident estimates. Consider the two parameter system depicted in Figure 1.3 which shows the imprecise mapping of $X$ onto $Y$ constrained only by three corresponding value pairs $C_0$, $C_1$ and $C_2$ and the knowledge that $Y$ is a monotonic increasing function of $X$. Observations of $X$ are subject to a Gaussian noise distribution and $x_T \in Q_B$ such that 40% of the observations of $X$ appear in $Q_A$. We demand a false-positive error-rate for $Y$ no greater than 20%. In $Y$'s quantity-space $Q'_A$ receives 40% of our predicted estimates of $Y$ from observations of $X$. Since the true state, which lies in $Q_B$ can map arbitrarily closely to landmark $y_T$ in $Y$'s quantity-space then, in order not to exclude the true state 40% of the time the standard deviation must be at least the size of region.

Figure 1.3: Mapping $X$ onto $Y$ with corresponding values $C_0$, $C_1$ and $C_2$. Dotted line shows actual function $y = f(x)$ which is only known to be a monotonic increasing function.

6Completeness and soundness are defined this way in much of the Qualitative Reasoning literature. However, some authors (Struss, 1997) switch the definitions of these terms and use completeness to describe reasoning which does not exclude real solutions.
$Q'_B$. Repeating this argument when the true state is in $Q_A$ and, say 40% of observations of $x$ are in $Q_B$, the required false-positive error-rate is obtained only if the standard deviation is no less than the size of region $Q'_A$. So, the standard deviation must not exclude any part of $Q'_A$ or $Q'_B$. Any indication of preference encoded in $X$ space is lost through propagation into $Y$ space.

If quantitative methods suffer from such fundamental problems, the question must be asked as to why they have been so persuasive in so many previous applications. The main reason for this simply stems from the underlying complexity of the data fusion problems that have been addressed (Gao and Durrant-Whyte, 1994). If a system is relatively simple, and sufficient constraints can be placed on its operation that it can be guaranteed that only well known and well understood processes can occur, then it is clearly possible to develop a model of operation which is exact and accurate enough to satisfy the requirements of a quantitative data fusion algorithm. However, as a system becomes more complex and as constraints and our understanding of the system diminishes, building such models rapidly becomes intractable.

1.5.5 Hybrid Methods

There has been an attempt to reconcile the Interval Intersection and the Kalman filter approaches to estimation. The Extended Interval Kalman Filter (EIKF) (Chen et al., 1997; Siouris et al., 1997; Chen et al., 1998) uses standard Kalman filter matrix prediction and update equations but their entries are generalised to interval values. Each point value and point valued operation in the standard Kalman Filter equations are replaced by their interval representation counterparts (i.e. $x^I = [x_L, x_U]$) and the expected value of an interval is the interval formed by the expected values of the random limits (i.e. $E(x^I) = [E(x_L), E(x_U)]$). Thus, imprecise models may be represented using the interval representation in a manner similar to the III.

Extending well known point-valued numerical methods to interval representations by simple one-to-one conversion of point valued algebra to interval valued algebra is a straightforward and often lucrative area of research. For example, Hofbaur (Hofbaur, 1997) extends Ljapunov’s algebraic stability test to qualitative models in this way and the semi-quantitative simulator NISIM (Kay and Kuipers, 1993) uses an interval extension of the Runge-Kutta method. However, an interval extension approach is inappropriate for the Kalman filter and, despite the claims of the authors, the EIKF is not optimal. Consider the fusion of two estimates, $\hat{x}^I(k | k-1) = (l_{k-1}, u_{k-1})$ and $\hat{x}^I(k) = (l_k, u_k)$, each with known variance $P(k | k-1) = P(k) = (\delta, \delta)$. Equation 1.18 yields $P(k | k) = \frac{1}{2}$ and, therefore, by Equation 1.19, the EIKF gives:

$$\hat{x}^I(k | k) = \frac{\hat{x}^I(k | k-1) + \hat{x}^I(k)}{2}.$$ 

The interval width of the combined estimate $\hat{x}(k | k)$ is a convex function of the input estimate widths and is larger than the tightest input estimate. When imprecision dominates uncertainty the optimal solution is obtained using the III, namely the intersection of the input intervals, which yields an estimate smaller than each of the input estimates.

1.5.6 Summary

In this thesis, we are concerned with situations when process and observation models and noise models are imprecisely known. Non-parametric point estimation methods such as the Kalman filter fail to represent imprecision adequately within the covariance matrix representation. Probabilistic
methods fail to maintain uncertainty when the system models are imprecise and interval intersection methods fail to assimilate degrees of uncertainty.

The ideal filter would include the advantages of each approach presented in this section. It would utilise the interval intersection approach when imprecision dominates uncertainty and point estimation methods when uncertainty dominates imprecision. The Dempster-Shafer Theory of Evidential Reasoning represents imprecise models of uncertainty by upper (Plausibility) and lower (Belief) bounds. Unlike Bayesian methods, where the evidence is divided between possible future states on ambiguous inference, the Dempster-Shafer approach assigns the evidence as a whole to the conjunction of the possible states. The origin of the Dempster-Shafer theory lies in frequency-based probabilistic estimation (Dempster, 1967; Dempster, 1968) but theoretical emphasis since this seminal work was published has focussed on the technical details of the belief function in the context of subjective probability assignments (Smets, 1988a; Hummel and Landy, 1988). Unfortunately, unlike Bayesian probability theory, frequency-based semantics for the DS representation are not preserved by the Dempster rule of combination. In this thesis we develop state ordering semantics for the interpretation of mass. The mass values will have no (and will need no) intrinsic probabilistic interpretation. Mass will be assigned non-parametrically and uncertainty will be encoded as its standard deviation. The Dempster rule of combination will operate under this state order interpretation, thus avoiding the probabilistic clash with Bayes rule. Ultimately, the strength of the state order may be converted to statements of probability for decision making. In developing this theory we endeavour to incorporate the non-parametric approach of point estimators.

1.6 The Robotics Test Domain

Applications in mobile robot localisation will be used to test the filter methodology presented in this thesis. *Localisation* is the determination of position in the environment and is a fundamental competence for mobile robot navigation. The robot must determine its position in the environment in relation to the goal location. This relationship is maintained in an internal representation (*map*) and is updated in accordance with the robot’s beliefs obtained from sensory information (*sensor cues*). Localising in unstructured environments using metric information is problematic (Sutherland and Thompson, 1994). Odometric errors can be significant and unpredictable especially when slippage is inconsistent with the internal control model. Unpredictable environmental effects on sensor behaviour can induce unmodelled metrical relocalization errors. Further, the inherent operation of the sensors themselves may induce distortion through, for example, the discretization of data.

A qualitative map is divided into qualitatively recognisable regions. In (Kuipers and Byun, 1991), the robot maintains a network description of *distinct places* in the environment. Distinctiveness measures are metrical and are not invariant within the boundaries of the distinct place. An example of a distinctiveness measure is the range and bearing information measured from a point from which there is an equal distance to near objects. Inter-region travel paths are encoded as procedural information between nodes (e.g. move along object to the left). Dai and Lawton (Dai and Lawton, 1993; Levitt and Lawton, 1990) describe an algorithm for region acquisition and robot navigation using visual sensors. Their approach requires either the presence of distinctive landmarks in the environment or a compass. For the case of compassless navigation distinctive regions are separated

---

7In this thesis, and in the AI literature, the term “landmark” is used to denote quantity-space partitioning points in qualitative representations and distinctive features in the environment. It is assumed that the appropriate meaning will be identified by the context.
by *landmark boundary pairs* (LBP). For example, in Figure 1.4, the LBP ‘a to b’ distinguishes two regions X and Y on either side. When a is observed to be to the left of b the robot is in region X. Distinctive LBP regions are identified by the conjunction of all LBPs between locally observed features and the LBP signature is invariant within the LBP region. The robot navigates by following LBPs.

\[ \begin{array}{c}
 & Y \\
 a & \quad b \\
 X & 
\end{array} \]

*Figure 1.4: LBP regions.*

Landmark navigation requires the ability to recognise individual landmarks. The robot used in this research can sense its environment and determine landmark characteristics using a variety of modalities: ultrasonic time-of-flight, frequency modulated (FM) sonar, infrared time-of-flight and gyroscope data. In general, it is difficult to model the behaviour of sonar and infrared observations quantitatively. For a typical in-air sonar there is no direct correspondence between the environment and the recorded range measurements. The manner in which the physical properties of the sensor (wave-length, beam-width) interact with the physical properties of the environment (characteristic size, object geometry, absorption, reflection and transmission characteristics) produces a wide variety of qualitatively different measurements. Sonar and laser sensing differ qualitatively since, under normal conditions, the sonar wave-length is much larger than the characteristic size of the reflector and, for a laser, the opposite is true. Figure 1.5 shows a partial 2D sonar and laser scan of the Robotics Laboratory in Oxford comprising walls, corners and a doorway. The views look totally different. On first examination, the laser range information adheres more closely to the way the room looks, and the sonar may be thought to be redundant. However, the wall partition join at X, which is not an obviously significant feature to the laser sensor, is a significant feature to the sonar. This permanent feature is a valuable landmark as it can be used to disambiguate the robot’s position along the otherwise homogeneous wall. Conversely, the wall running through X is significant in geometric terms, but is less so to a sonar. Thus sonar and laser time-of-flight modalities offer complementary information about the structure of the local environment.

Simple model-free methods for interpreting and combining sonar and infrared data have been proposed. The Occupancy Grid method (Elfes, 1987; Matthis and Elfes, 1988; Courtney and Jain, 1994) assumes a naive statistical correspondence between the observed sonar range reading and the true range to the target and, unsurprisingly, this method has not met with great success. Referring again to Figure 1.5 we can appreciate Kuc’s comment on ultrasonic sensing (Kuc, 1990):

*However, problems arise in the straightforward, but naive, interpretation of time-of-flight (TOF) readings: objects that are present are not always detected and range readings produced by the TOF system do not always correspond to objects at that range. Because of these problems, many researchers abandon sonar-only navigation systems …*  

However, all features in the ultrasonic scan in Figure 1.5 can be *explained* using the physical properties of the sensor and the environment. Typically, sonar scans are taken from a simple time-of-flight acoustic sensor with single chirp at 47.5KHz and with a beam-width of approximately 30°. At this frequency the wave-length is approximately 1cm. For many sonar devices, especially those used in
this research, the chirp is emitted and received by circular piston transducers. The angular dependence of the radiated intensity, shown in Figure 1.6, exhibits diffraction rings and the main beam is accompanied by less intense but often observed side-lobes. The characteristic roughness of the wall is much smaller than the acoustic wavelength and the specular (i.e. mirror-like) reflection from a relatively smooth surface (shown in Figure 1.5, part (a)) forms a circular arc with radius corresponding to the perpendicular range return. Alternatively, the diffuse reflector has surface texture with a much larger characteristic size and this return is qualitatively different from the specular return in that it approximates a straight line rather than a circular arc (Figure 1.5, part (b)).

Kuc and Siegel (Kuc and Siegel, 1987) developed a mathematical model for sonar reflections from smooth, planar surfaces and corners using virtual images of the transmitter (see Figure 1.7) and later, reflections from rough surfaces were modelled by Bozma and Kuc (Bozma and Kuc, 1994). The latter demonstrated that as a robot approaches a diffuse plane the form of the sonar return can change and begin to approximate a circular arc similar to the specular return towards the centre of the image while maintaining the characteristic diffuse straight-line approximation further out. This is a result of the wide-beam characteristics of the sonar sensor and the reflectance properties of the diffusing material.

Figure 1.8 shows a sonar scan of a simple environment for a robot placed at 'A'. The scan is a record of the half distance estimate of the product of the speed of sound in air and the time taken
Figure 1.6: Radiation intensity from circular piston acoustic transducer (wavelength 1cm, aperture radius 2cm) measured relative to the paraxial direction \( \theta = 0 \).

Figure 1.7: Equivalent sonar sensor configurations for planes and corners. A configuration comprising sonar emitter \( s \in \{a, b\} \) is equivalent to a reflector-less environment with (virtual) sensor \( s' \).

for a signal to reflect and return to the sensor:

\[
\text{Range} = \frac{\text{speed of sound} \times \text{time of flight}}{2}. \tag{1.21}
\]

However, from this scan it can be seen that the range measurements for reflections perpendicular to the walls at \( a, b, c \) and \( d \) in the figure, the edge at \( e \) and corners at \( f, g \) and \( h \) are in close correspondence to distance between the sensor at \( + \) and the reflectors. Furthermore, range values are equal over a range of bearings each side of these points. This phenomenon has been noted by (Kuc and Siegel, 1987) and (Leonard and Durrant-Whyte, 1992) and the latter have coined the phrase “region of constant depth (RCD)”. An RCD is a contiguous sequence of bearings with equal range values (RCD formations in the sonar map for the room in Figure 1.8, part (a) and a robot at ‘+’, are shown in bold). They are formed because the sonar beam is wide and during a sweep scan a tiny part of the reflector is visible over a discrete range of bearings.
As a robot moves through the environment, the RCDs move predictably. RCDs formed by specular reflections from walls move tangentially with the wall and abreast of the robot. Corner and edge RCDs rotate about the point of reflection. This is apparent in Figure 1.8, part (b) which shows the overlay of a set of RCDs taken from various positions in the environment. Leonard (Leonard and Durrant-Whyte, 1992) demonstrated how a robot can navigate by tracking RCDs as the robot moves. Imagine walking through the room depicted in Figure 1.8, part (a). When walking towards location Y, for example, the RCD at a would move abreast of the observer while edge at h would appear to move away from the observer. The essential point here is that the information obtained by a sonar does not correspond well with the underlying geometry of the environment. However, the information (i.e. RCDs) are predictable and have motion patterns which are well understood consequences of the underlying physics of the sensing process. The RCD is an appropriate ontological cue for first return time-of-flight sonar sensing and is adopted in this thesis. However, such primitives are not pertinent for continuous frequency-modulated sonar sensing and laser time-of-flight sensing which are able to accumulate more depth information as Figure 1.5 testifies.

Recently, Durrant-Whyte (Durrant-Whyte, 1991) has argued the need to model the dynamic sensing process and proposes that a qualitative representation would capture a deep understanding of the physics of the sensing process as it is a more powerful representational mechanism for describing complex sensor information and interactions. Qualitative representations are sufficiently flexible that a significant corpus of physical laws can be included in the sensor interpretation process. This thesis proposes the use of “sensor frames” for landmark navigation which are viewpoint representations of mapped features obtained by qualitative reasoning about the way in which the robot interacts with the environment (the correspondence of sensory cues derived by reasoning about the physics of the sensing modalities) and the orientation of these objects relative to the robot (maintained using the LBP representation). The sensor frame is a realisation of Minsky’s frame representation (Minsky, 1975):

A frame is a data-structure for representing a stereotyped situation ... Some of this
information is about how to use the frame. Some is about what one can expect to happen next... For visual scene analysis, the different frames of a system describe the scene from different viewpoints, and the transformations between one frame and another represent the effects of moving from place to place.

To illustrate the use of sensor frames in the sonar domain a table is observed (or actually its four legs are observed) by a frequency modulated ultrasound sensor which scans in the 2D plane parallel to the floor. The FM sonar sensor displays echo range and pressure data. We assume that odometry has led the robot to believe that it is in one of regions I, J, K or L in Figure 1.9. These regions are bounded by LBPs. LBPs are defined by the table legs 1, 2, 3 and 4 and by the intersection of other LBPs. The robot's positional uncertainty can be reduced by localising with the table. An anti-clockwise sonar scan across the room identifies four legs labelled in order of appearance "a", "b", "c", and "d". Each distinctive place is labelled with the relative ranges of each leg (called the BR order). For example, if $R(b) < R(d) < R(a) < R(c)$ then “bdac” is attached to that distinctive place indicating that the third closest leg was observed first, followed by the closest leg, then the furthest leg and finally, the second closest leg. Consistent labellings for the configuration depicted in Figure 1.9 are shown in Table 1.1.

Figure 1.10 shows the FM sonar echo profile for five bearings. Significant echoes originate at ranges 0.6, 0.8 and 0.9 metres respectively. By reasoning about the pressure trends for each echo we infer that the peak at range 0.8 m is observed first, followed by that at range 0.9 m, then at 0.6 m and then a second peak at 0.8 m. This corresponds to a BR-labelling $CXYB$ where $\{X, Y\} = \{A, D\}$. Ambiguity is induced by the fact that there were two peaks at 0.8 m. Matching this labelling with those in the table above we conclude that the robot is in region C, E or K. The robot concludes that it is location K from its odometric information.

This example demonstrates the kind of representation we envisage encoding in the robot's cognitive map. This thesis is not directly concerned with the map building task. It is concerned with a more fundamental competence of the map building process, namely the modelling of landmarks in
Table 1.1: Ordinal table leg spatial relationships for various robot locations. See text for explanation.

Figure 1.10: Frequency-modulated sonar 2D amplitude-range profile of a four legged table.

the environment which, ultimately can be recognised and assigned to LBPs for the localisation task.

1.7 Conclusions and Thesis Overview

This chapter has investigated various representations of imprecision and uncertainty for use in the robot navigation task when the physical behaviour of the robot sensing modalities are hard to describe quantitatively. The numeric interval representation was chosen to represent imprecision as it captures many different information types (i.e. nominal, ordinal, interval and point). Reasoning about uncertainty using probabilistic models or point estimation techniques were discounted and a
non-parametric mass formalism was proposed.

This thesis will be chiefly concerned with the development of a reasoning framework for both imprecision and uncertainty and will involve the reconciliation of the two distinct filtering frameworks described in this chapter:

- Point estimation, namely the Kalman filter, for eliminating uncertainty (i.e. random error).
- Numeric interval methods, such as the IIF, for eliminating imprecision (i.e. systematic error).

We seek a single framework which is commensurate with both paradigms:

- when imprecision dominates the new filter approximates to the IIF.
- when uncertainty dominates, the filter approximates to the Kalman filter.

The new filter would then approximate to optimal filtering methodologies in extreme cases when imprecision or uncertainty dominate.

Chapter 2 describes the sonar and laser time-of-flight range finders and develops qualitative spatial and physical models based on the idea of optical flow from Vision theory (Sonka et al., 1993) and the ENDURA method of Bozma and Kuc (Bozma and Kuc, 1994). Chapters 3 and 4 address the problem of parameter estimation under noisy observations. Reasoning that Cox's proof for additive theories of evidence is incorrect, we introduce a generalisation of the Dempster-Shafer Theory of Evidential Reasoning (Shafer, 1976) to a continuum of P-norm calculi. From this and the concept of Biscay information and the non-parametric Biscay distribution, we develop the Qualitative filtering framework for data fusion. Mathematical properties of this framework are presented. Chapter 5 demonstrates the extent of application of the Qualitative filtering framework and develops a non-parametric continuous domain estimation technique for applications where the system models are imprecise. This new filter is demonstrated on the Ballistic Missile Problem. Finally, in Chapter 6, the Qualitative filter is applied to various problems in the robotics sensor fusion domain and we see how model-free scale-space abstraction techniques can be used with the qualitative filtering framework for hypothesis suggestion. The relevant literature is reviewed throughout the thesis.

The main contributions of this thesis are:

- The extension of the specular motion model of sonar reflectors developed by Kuc and Leonard (Kuc, 1990; Leonard and Durrant-Whyte, 1992) to a sensor-centred representation of sensor one flow. This can be used for first-order reflections from surfaces with arbitrary curvature and the motion of higher-order reflections from planar surfaces.
- A novel argument against Cox's paper "Probability, Frequency, and Reasonable Expectation" (Cox, 1946) and the subsequent introduction of the P-Norm generalisation of the Dempster-Shafer Theory of Evidential Reasoning. The proposal of the Lagrangian approximation for filtering and definition of unbiased filtering for statistical estimation including unbiased mean and median estimators. The introduction of Biscay information and the non-parametric Biscay distribution.
- Proof of COMOC (morphological scale-based filtering) unbiased conditions and demonstration of its utility for hypothesis generation within the Qualitative filtering framework.
- The introduction of a new approach to filtering in the domain using dynamic landmarks.
The results of this research have been combined into a system which has been implemented and tested on the “OxNav” robot (Stevens et al., 1995) (shown in Figure 1.11) which is equipped with four time-of-flight Polaroid sonar sensors (Everett, 1995), a Gyrostar gyroscope (Everett, 1995) and a SICK time-of-flight laser sensor (Optik-Elektronik, 1995). The “OxNav” robot, shown in Figure 1.11 is a product of the EPSRC funded projects GR/J/46067 and GR/J/57773. These methods have also been tested on a continuous time, frequency modulated (FM) time-of-flight ultrasound sensor (Lee, 1997). Experimental results for both real and simulated data are presented throughout this thesis.

Figure 1.11: The “OxNav” robot.
Chapter 2

The Qualitative Sensor Model

2.1 Introduction

Much research has been expended on qualitative map representations (Kuipers, 1978; Kuipers and Byun, 1991; Mataric, 1990; Dai and Lawton, 1993; Hernandez, 1994; Lee, 1995) for which a robot localises and plans its motion relative to distinct feature landmarks. However, very little work has been undertaken on the qualitative modelling of the features themselves. This chapter is concerned with the development of the robot centred ontologies for RCDs and the physical behaviour of Frequency modulated (FM) sonar cues: (i) qualitative acoustic flow models for RCD sonar range sensors and gyroscopes are derived for single and multiple reflections and (ii) qualitative physical models relating energy, duration and range (QENDURA) of FM sonar data are derived. QSim (Kuipers, 1994) is applied to these models to generate sensor frames and the utility of the qualitative approach is investigated for target identification in the robot navigation domain. These models are exploited with real data in Chapter 6.

We build on the work presented in (Kuc and Siegel, 1987) and (Leonard and Durrant-Whyte, 1992) for robot navigation by RCD tracking and develop RCD motion models for multiple reflections from planar surfaces. Probably as a consequence of Leonard’s book RCD models have been used, almost exclusively, in conjunction with absolute coordinate frame quantitative maps of the environment. Such approaches perform target classification by verifying the consistency of the estimated position of the robot, the absolute (mapped) position of the target and the quantitative RCD behaviour. An abstract qualitative framework cannot rely on such geometry and, in this chapter, we develop sensor-centred, coordinate-independent models which do not require an absolute geometric representation of the environment.

2.2 Sensor Configuration and Experimental Technique

The robot sensor suite comprises a servo-mounted binocular RCD (RCD Differential) tracking sonar, a Gyrostar gyro and a SICK time-of-flight infrared sensor. The sonar RCD unit is able to track normal surface reflections as the robot moves through the environment by turning so as to minimise the differential time-of-flight of a ultrasonic pulse emitted by one transducer and subsequently received by both (Manyika and Durrant-Whyte, 1994). The Acoustic Flow model relating the curvature of the reflector to the motion of the robot and RCD range values and turn rates will be developed in the next section. As we will see, the qualitative form of this model requires that the gyro is affixed on top of the RCD unit (see Figure 2.1) so that it can measure the combined turn rate of the RCD unit and robot.
The SICK laser range finder is a scanning sensor which returns 360 range readings over a 180° field of view in front of the robot. Whereas the robot must be moving to solicit useful information from its tracking RCDD unit, the robot is stationary when it uses the laser to scan the reflector. Both sonar and laser sensors are necessary since (i) the sonar range value is uncertain to 1cm whereas the laser range value is sensitive to 5cm. The latter is inadequate for measuring small range changes necessary for the differential Acoustic Flow model; (ii) the SICK laser sensor head is unable to track the normal reflection. Thus, it is not possible nor would it be useful to affix the gyro to the laser head. (iii) Both sonar and laser sensors measure different aspects of the qualitative Acoustic Flow model and, as we will see later in this chapter, both are necessary to remove incompleteness problems within the qualitative approach.

![Gyro mounted on a servo-mounted binocular RCDD (RCD Differential) tracking sonar. The RCDD unit is able to track normal surface reflections by turning so as to minimise the differential time-of-flight of a ultrasonic pulse emitted by one transducer and subsequently received by both. The Gyrostar gyro is affixed to the RCDD unit so that the gyro measures the turn-rate of the RCDD unit (i.e. the gyro measures \( \delta \theta + \delta \phi \)).](image)

2.2.1 Noise Characteristics

Noisy sonar and gyro measurements can degrade the performance of both qualitative and quantitative feature recognition systems. This section presents a sample data set as the robot tracks a convex surface. A piezoelectric gyroscope (a Murata Gyrostar Model ENV Piezoelectric Vibrating Gyroscope (Barsban and Durrant-Whyte, 1995; Everett, 1995)) is mounted on a binocular sonar tracking system (Maryika and Durrant-Whyte, 1994) (see Figure 2.1).

Figure 2.2 shows real data from the gyro and sonar system for the sonar orientation change \( \delta \phi \), sonar bearing relative to robot chassis \( \theta \) and range to feature \( R \). The feature is a cylinder of (near) constant radius \( r = 0.12m \). The figure shows the distribution of calculated feature radius-of-curvature \( r \) and demonstrates the degree to which the data is corrupted by noise. Estimation using qualitative models and noisy data will be the focus of Chapters 3 and 4.
2.3 Towards a Qualitative Model: The Acoustic Flow Equations

2.3.1 1st Order Reflections from Arbitrary Curved Surfaces

Leonard (Leonard and Durrant-Whyte, 1992) introduces a unified quantitative description of plane, edge and corner sonar reflectors. Each feature can be regarded as a cylindrical surface in which planes are cylinders with infinite radius and edges and corners are cylinders with zero radius. In this section, we develop a qualitative generalised first order \(^1\) RCD model for the complete class of continuous, curved, specular surfaces. We develop the equations which relate the radius of a cylinder \(r\) and the perpendicular distance from its surface \(R\) to a robot moving with translational speed \(v\) and rotational speed \(\phi\) (see Figure 2.3). This extends the models in (Kuc and Siegel, 1987; Leonard and Durrant-Whyte, 1992) to include arbitrary continuous (though not necessarily continuously differentiable) 2D surfaces. For coincident emitter and receiver transducers, only reflections normal to the surface are observed. In Figure 2.4 the robot is initially at \(a\) and then moves with speed \(v\) for a time period \(\delta t\) to position \(c\) while changing its heading smoothly through an angle \(\delta \phi\). When the robot is at \(d\) it observes the reflection from \(e\) at a bearing \(\theta\) relative to the direction of motion of the robot. Subsequently, when the robot is at \(c\) it observes the reflection from \(b\) on the surface at a bearing of \(\theta + \delta \theta\). The surface is parameterised by tangent direction \(\alpha(s)\) at arc length \(s\) (Carmo, 1976). \(^2\) The curvature \(\kappa\) at \(s\) is defined to be the inverse of the radius \(r\) for a generalised cylinder

---

\(^1\)First order refers to the number of reflections the sonar pulse undergoes before returning to the transducer.

\(^2\)In (Carmo, 1976) \(\alpha\) describes the tangent line and \(\alpha'(s)\) is the tangent direction.
2.3 Towards a Qualitative Model: The Acoustic Flow Equations

Figure 2.3: System geometry. Beacon is at bearing $\theta$ and robot turn rate is $\dot{\phi}$.

Figure 2.4: System Geometry.

which touches the curve at $s$ (Carmo, 1976):

$$r = \frac{1}{\kappa} = \frac{1}{\alpha'(s)}.$$  \hfill (2.1)
2.3 Towards a Qualitative Model: The Acoustic Flow Equations

Referring to Figure 2.4:

\[
\begin{align*}
\angle dhc &= \delta \phi - \gamma \\
\therefore \angle bcd &= \pi - (\delta \phi - \gamma) - \theta - \delta \theta \\
\therefore \angle bjc &= \delta \phi - \gamma + \theta + \delta \theta - \mu - \delta \mu \\
\therefore \angleaji &= \pi - (\delta \phi - \gamma + \theta + \delta \theta - \mu - \delta \mu).
\end{align*}
\] (2.2)

Also:

\[
\begin{align*}
\angle edi &= \pi - (\theta - \gamma) \\
\therefore \angle eid &= \theta - \gamma - \mu
\end{align*}
\] (2.3)

Combining Equations 2.2 and 2.3:

\[
\delta \alpha = \angle ai j = \delta \theta + \delta \phi - \delta \mu
\]

and therefore:

\[
\frac{1}{r} \delta = \dot{\alpha} = \dot{\theta} + \dot{\phi} - \dot{\mu}.
\] (2.4)

Referring to Figure 2.4, let \(\angle beg = \beta\), where \(\vec{t}\) is tangential to the surface at \(\epsilon\):

\[
\angle ai j = \pi - [\angle edi + \mu] = \theta - \gamma - \mu
\]

Since \(\vec{t}\) is parallel to \(\vec{n}\) then \(\angle aef = \angle ai j\) and:

\[
\begin{align*}
\angle feg &= \pi - \angle aef - \angle gei = \frac{\pi}{2} - \angle ai j = \frac{\pi}{2} - \theta + \gamma + \mu \\
\therefore \angle bcf &= \beta - \frac{\pi}{2} + \theta - \gamma - \mu
\end{align*}
\]

From \(\triangle bcf\) in Figure 2.4 \(|cf| = |dc| - |da| - |cm|\) and therefore:

\[
\delta s \cos(\beta - \frac{\pi}{2} + \theta - \gamma - \mu) = v \delta t - R \cos(\theta - \gamma) + (R + \delta R) \cos(\theta + \delta \theta + \delta \phi - \delta \mu - \gamma),
\] (2.5)

Since, \(\lim_{\delta t \to 0} \beta = 0\) and \(\lim_{\delta t \to 0} \gamma = 0\) then to first order linear approximation:

\[
\delta s \sin(\theta - \mu) = v \delta t - [R \cos(\theta + R \gamma \sin(\theta)] + \delta R \cos(\theta + \delta \theta - \delta \phi - \delta \mu - \gamma)]
\]

Therefore:

\[
\delta \sin(\theta - \mu) = v \dot{\theta} - R \cos(\theta - \mu) \sin(\theta + \dot{\phi} - \dot{\mu}).
\] (2.6)

Similarly, \(|bf| = |bm| - |el|\):

\[
\delta s \sin(\beta - \frac{\pi}{2} + \theta - \gamma - \mu) = - R \cos(\theta + \delta \phi - \gamma) - R \sin(\theta - \gamma)
\] (2.7)

and therefore:

\[
\delta \cos(\theta - \mu) = - \dot{R} \sin(\theta - R \cos(\theta + \dot{\phi})
\] (2.8)

Equation 2.6 \(\times \sin \theta + \) Equation 2.8 \(\times \cos \theta\) gives:

\[
\delta \cos \mu = v \sin(\theta - R \cos(\theta + \dot{\phi})
\] (2.9)

and substituting for \(\delta\) from Equation 2.4 gives:
Similarly, Equation 2.6 × cosθ – Equation 2.8 × sin θ gives:

\[-\dot{\delta} \sin \mu = v \cos \theta + \dot{R}\]  

(2.11)

and substituting for \( \delta \) from Equation 2.4 gives:

\[
\dot{R} = -v \cos \theta - R(\dot{\theta} + \dot{\phi}) \sin \mu.
\]  

(2.12)

Equations 2.10 and 2.12 are the Acoustic Flow Model. We will see that full feature discrimination can be performed using pure qualitative counterparts of the Acoustic Flow Model and a static laser range sensor in conjunction with a moving sonar sensor.

**The Case when \( v = 0 \)**

In this section we derive the model for the static laser sensor, the case when \( v = 0 \) in Equations 2.10 and 2.12.

Substituting for \( \mu \) from Equation 2.10 into Equation 2.12 and setting \( v = 0 \):

\[
\dot{R} = R(\dot{\phi} + \dot{\theta}) \tan \mu.
\]

Defining \( \Phi = \phi + \theta \), then by the chain rule:

\[
\frac{dR}{d\Phi} = R \tan \mu.
\]  

(2.13)

and differentiating with respect to \( \Phi \):

\[
\frac{d^2R}{d\Phi^2} = \frac{dR}{d\Phi} \tan \mu + R \sec^2 \mu \frac{d\mu}{d\Phi}.
\]  

(2.14)

From Equation 2.10 and the chain rule:

\[
\frac{d\mu}{d\Phi} = 1 + \frac{R}{r \cos \mu}.
\]  

(2.15)

Substituting Equations 2.13 and 2.15 into Equation 2.14:

\[
\frac{d^2R}{d\Phi^2} = R \tan^2 \mu + \frac{R}{\cos^2 \mu} \left(1 + \frac{R}{r \cos \mu} \right) (1 + \frac{R}{r \cos \mu}),
\]

Using \( \sin^2 \mu + \cos^2 \mu = 1 \) then:

\[
\frac{d^2R}{d\Phi^2} = R \tan^2 \mu + R(1 + \tan^2 \mu) \left(1 + \frac{R}{r \cos \mu} \right).
\]  

(2.16)
2.3.2 Multiple Reflections from Planar Surfaces

Figure 2.5 shows the first, second and third order reflections obtained from experimental data for an acute angle formed by two non-parallel specular, plane surfaces. Specular profiles are often bedevilled with higher order returns. The usual solution is to ignore this type of return. The following section extends Kuc’s theoretical investigation using virtual images for right-angled corners to non-right angled surfaces called acute corners. A qualitative RCD ontological model for acute corners is derived. This model is verified for a circular transducer with primary and secondary power beams using the simulation analysis presented in (Kuc and Siegel, 1987). We confirm the real sonar data presented in Figure 2.5, which was taken from an acute corner of 0.87 rads made from perspex. The aim of this analysis is to investigate the mutual occlusion effects of transducer emission and reception since, for the multi-order case, the outward and inward ray paths do not coincide. The

![Diagram](image)

Figure 2.5: First, second and third order RCDs for an acute corner (0.87 rads.). Circular dashed lines denote the theoretical ranges of RCDs (Experimental data).

A robot-centred model for the motion of multiple order reflections is derived next.

![Diagram](image)

Figure 2.6: System Geometry.

Figure 2.6 shows a robot moving with speed $v$ and turning with angular rate $\phi$ between two
2.3 Towards a Qualitative Model: The Acoustic Flow Equations

planar surfaces $p1$ and $p2$. The planar surfaces can be extended to form a virtual corner at $O$. This corner has acute angle $\lambda$. In Appendix A, Equations A.5 and A.6 are derived for the orientation of the principal ray for even and odd order reflections:

$$
\mu_n^{\text{even}} = \pm \frac{\pi - n\lambda}{2},
$$

$$
\mu_n^{\text{odd}} = \pm \left( \frac{\pi - (n + 1)\lambda}{2} \right) + \epsilon.
$$

The motion of the virtual corner relative to the robot is obtained from Equation 2.10. Let $P$ be the distance from the robot to the virtual corner $O$. The motion of $O$ relative to the robot is obtained by substituting $R = P$ and reassigning $\theta \rightarrow \theta - \mu_n$ in Equation 2.10:

$$
v \sin(\theta - \mu_n) = P(\dot{\theta} - \dot{\mu_n} + \dot{\phi}),
$$

(2.17)

Rearranging:

$$
\dot{\theta} = \frac{v \sin(\theta - \mu_n)}{P} + \dot{\mu}_n - \dot{\phi}.
$$

(2.18)

Similarly, from Equation 2.12 we have:

$$
\dot{P} = -v \cos(\theta - \mu_n).
$$

(2.19)

Differentiating Equation A.7 with respect to time gives:

$$
\dot{R}_n = \dot{P} \cos \mu_n - P \dot{\mu}_n \sin \mu_n.
$$

(2.20)

Substituting $\dot{\epsilon} = \dot{\theta} - \dot{\mu}_n + \dot{\phi}$ (see Figure 2.3) into Equations A.5 and A.6:

$$
\mu_n^{\text{even}} = 0,
$$

(2.21)

$$
\mu_n^{\text{odd}} = -\frac{v \sin(\theta - \mu_n)}{P}.
$$

(2.22)

Substituting the expressions for $\mu_n$ into Equations 2.18 and 2.20 gives:

$$
\theta_n^{\text{even}} = \frac{v \sin(\theta - \mu_n)}{P} - \dot{\phi},
$$

(2.23)

$$
\theta_n^{\text{odd}} = -\dot{\phi},
$$

(2.24)

$$
R_n^{\text{even}} = -v \cos(\theta - \mu_n) \cos \mu_n,
$$

(2.25)

$$
R_n^{\text{odd}} = -v \cos \theta.
$$

(2.26)

Figure 2.7 shows actual and theoretical RCDs for orders 1 to 3 for an environment comprising two non-parallel perspex planes. The RCDs have been observed at positions separated by 0.2$m$ on the $x$-axis. The initial parameters are:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot position $P$</td>
<td>3.6 m</td>
</tr>
<tr>
<td>Corner angle $\alpha$</td>
<td>0.7 rads</td>
</tr>
<tr>
<td>Robot angular position $\epsilon$</td>
<td>0.35 rads</td>
</tr>
</tbody>
</table>

Qualitatively, the higher order RCDs behave thus:
2.4 The Qualitative Sensor Model

The sensors we investigate (sonar, laser time-of-flight) offer a 2D view of the environment (1D signal propagation and 1D rotation of sensor head). The robot identifies features through the intersection of a plane with the environment and Leonard (Leonard and Durrant-Whyte, 1992) shows that feature curvature is a good discriminant in such circumstances. Hoffman (Hoffman and Richards, 1988) contemplates encoding shape by curvature and concludes that a qualitative description of curvature suffices to capture the essential discriminants of the environment. Hallam (Hallam, 1986) suggests that the qualitative curvature of specular targets can be determined by a qualitative description of their visibility sets (i.e. a qualitative description of the environment from which the target can be observed by a sonar sensor). He suggests that the behaviour of view vectors (i.e. normal incident reflections) when moving between two locations may be sufficient to determine the visibility sets. The Spatial Reasoning community (see for example (Freksa, 1992), (Hernandez, 1994) and (Mukerjee and Mittal, 1995)) has proposed ego-centred, allo-centred and hybrid qualitative representations of space which uses topology, arrangement, orientation and order-of-magnitude distance information. However, according to Forbus’ Poverty Conjecture (Forbus, 1991) “there is no purely qualitative,
general-purpose representation of spatial properties". To elaborate, although qualitative spatial information is useful, qualitative representations by themselves are inadequate to draw many classes of predictions. The aim of this chapter is to develop and investigate the predictive power of a qualitative formulation of the Acoustic Flow Equations.

In this section, we investigate how qualitative information about curvature can be computed using qualitative descriptions of the view vectors and we demonstrate how a mobile robot, equipped with a sonar tracking sensor, a laser time-of-flight range sensor and a gyroscope (see Figure 2.1) can distinguish feature curvature using qualitative models inter-relating its sensor cues \(^3\). In later sections, we will investigate the problem of noise filtering using qualitative models. In Figure 2.8, the robot (indicated by \(\triangle\)) passes in front of a curved surface (shaded). For specular (i.e., mirror-like) surfaces the robot sees only the reflection normal to the surface (i.e \(\mu = 0\)) (Kuc, 1990; Leonard and Durrant-Whyte, 1992). The robot tracks the feature by following this normal reflection (see Figure 2.9).

The robot is able to lock on to the normal reflection by measuring the differential time-of-flight of the signal emitted by one transducer and received by both transducers (see Figure 2.1 and (Maryaka and Durrant-Whyte, 1994)). If the sonar (indicated by \(\blacksquare\) in Figure 2.8) turns with rate \(\dot{\varphi}\) while tracking a beacon with curvature \(\frac{1}{2}\) lying a distance \(R\) from the robot and at a bearing \(\theta\) relative to the forward motion of the robot then:

\[
\dot{\varphi}(r + R) = -\dot{r}\tan\theta.
\]

The robot is able to determine the surface curvature \(r\) of a feature in the environment by fusing qualitative information about \(\tan\theta\), \(\dot{r}\) and \(\dot{\varphi}\). The natural landmarks for the system are shown in Table 2.1. The zero angle landmark means forward in QSim notation. The reflector type landmarks 0 and in \(f\) denote edges (and corners) and planes respectively. Table 2.2 shows qualitative behaviours

\(^3\)This section builds on work presented elsewhere (Reece and Durrant-Whyte, 1995b).
consistent with Equation (2.27)\footnote{The qualitative magnitude and derivative operators are:}

\[
\begin{align*}
Q\text{mag}(X) &= \text{sign}(X) \subseteq (\infty, \infty) \\
Q\text{dir}(X) &= \text{sign}(X) \subseteq (\infty, \infty).
\end{align*}
\]

To illustrate the interpretation of this table consider the fourth entry in conjunction with the close-concave surface depicted in Figure 2.8. When the feature lies left-frontal (i.e. $Q\text{mag}(\tan(\theta)) = +$) and the range $R$ is decreasing (i.e. $Q\text{dir}(R) = -$) and the gyro is turning clockwise (i.e. $Q\text{dir}(\phi) = -$) then the feature must be a concave surface (i.e. $Q\text{mag}(\phi) = -$).

This table can be constructed from Equation (2.27) using qualitative knowledge of the qualitative algebraic building blocks (Kuipers, 1994; Reece and Durrant-Whyte, 1995b). For example, qualitative multiplication: $[Q\text{mag}(A) = +] \land [Q\text{mag}(B) = +] \Rightarrow [Q\text{mag}(A) \otimes Q\text{mag}(B) = +]$ or qualitative negation: $[Q\text{mag}(A) = +] \Rightarrow [\circ Q\text{mag}(A) = -]$.

To implement this interpretation of qualitative sensor cues, QSim was extended to deal with persistently infinite variables. QSim treats infinity as a point value and, therefore, a variable cannot remain at infinity and be decreasing simultaneously. It is not possible to represent plane reflectors in such a system since $(R + r)$ in the Acoustic Flow equations may be infinite but not necessarily constant. In general, we want to allow behaviours for $A(t)$ when $A(t) = \lim_{x \to \infty} x + B(t)$ and $Q\text{mag}(A(t)) = \infty$ but $Q\text{dir}(B(t)) = -$ or $Q\text{dir}(B(t)) = +$. This is achieved by allowing the infinity landmarks to be both successors and predecessors of themselves in QSim quantity spaces.

Figures 2.10 and 2.11 show the qualitative behaviours obtained from the Acoustic Flow equations.
### 2.4 The Qualitative Sensor Model

<table>
<thead>
<tr>
<th>Qmag(\tan \theta)</th>
<th>Qlin(R)</th>
<th>Qlin(\phi)</th>
<th>Qmag(\pm)</th>
<th>Surface Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>concave</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+/−</td>
<td>concave/convex</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+/−</td>
<td>concave/convex</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+/−</td>
<td>concave/convex</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>concave</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>concave</td>
</tr>
<tr>
<td>+/0/−</td>
<td>+/−</td>
<td>0</td>
<td>0</td>
<td>plane</td>
</tr>
</tbody>
</table>

Table 2.2: Curvature inferred from sonar and gyroscope cues.

<table>
<thead>
<tr>
<th>⨿</th>
<th>+</th>
<th>0</th>
<th>−</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>0</td>
<td>+</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>−</td>
<td>?</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>⨿</th>
<th>+</th>
<th>0</th>
<th>−</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>−</td>
<td>?</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>⨿</th>
<th>+</th>
<th>0</th>
<th>−</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>∞</td>
<td>−</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>?</td>
<td>0</td>
</tr>
<tr>
<td>−</td>
<td>−</td>
<td>−∞</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>⨿</th>
<th>+</th>
<th>−</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 2.3: Qualitative sign relationships for binary addition, multiplication, division and unary negation operations.

For a robot moving towards a plane and past a cylinder, an edge and a concave surface. The top three graphs in Figure 2.10 (beh1) show the qualitative behaviour for the case of the plane RCD. In the top left graph the generalised cylinder radius is infinite and steady corresponding to a plane reflector. The top middle graph shows that the range value decreases and the top right graph shows that the bearing \( \theta \) to the plane remains constant. For plane reflectors the RCD moves tangentially to the reflector and abreast of the robot and this is the reason why \( \theta \), in this case, is constant. The second row of graphs (beh2) describe the behaviour of an edge (or corner) RCD. In the middle left graph the generalised cylinder radius is zero and steady corresponding to an edge reflector. In the middle graph we see that the range gradually decreases until some time point \( t1 \) and then increases indefinitely. This corresponds to a robot moving towards a point like object fixed in space, passing close to it and then moving away from it. In the middle right graph we can see that this RCD moves gradually further to the left and then reedoes behind. The third row of graphs shows the RCD behaviour beh3 for an arbitrary cylinder of finite radius. The qualitative behaviour of \( R \) and \( \theta \) are the same for a corner, only the qualitative curvature estimate is different.

Figure 2.11 is dedicated entirely to the concave surface RCD behaviours. Two behaviours emerge since \( R + r \) can take either positive or negative values. The top row (beh4) describes the situation when \( R < −r \) (called the close-concave behaviour) for which range and bearing both decrease towards zero. This is a distinct behaviour of concave surfaces. The second row (beh5) describes the situation when \( R > −r \) (called the far-concave behaviour) for which range and bearing behaviours are identical to those of the convex and edge targets. We will demonstrate later how concave and convex surfaces can be discriminated when behaviours beh2, beh3 and beh5 are encountered by considering qualitative laser time-of-flight information in conjunction with the qualitative sonar RCD behaviours.
2.4 The Qualitative Sensor Model

Figure 2.10: QSim generated behaviours for plane, edge and convex surface targets. Ordinates are the ordered variable quantity spaces; Abscissa in each case represents the time axis with landmarks denoting boundaries between qualitatively different behaviours. Parameter trends are represented by ↑, ↓ and ○ indicating increasing, decreasing and steady parameters respectively. The first column of graphs describe the constant curvature reflector type: planar in the top left graph; an edge in the middle left graph and a convex surface in the bottom left graph. The middle column represents the sonar sensor range cue behaviour as a function of time: range is continuously decreasing for a plane (top middle graph); range initially decreases but then increases for the edge and convex surface cues (middle and bottom middle graphs respectively). The right column represents the bearing cue behaviour: the top right graph exhibits the constant bearing characteristic of a plane surface; the right middle graph shows monotonic increasing behaviour for convex and edge surfaces.

Since the qualitative model uses only variable values and first derivatives, any continuous surface can be captured by a simple qualitative description irrespective of the detailed geometry (i.e. irrespective of the higher order variable derivatives such as changing curvature). For example, the “stone” shape in Figure 2.12 can be modelled as a sequence of convex, concave and edge type features (Hoffman and Richards, 1988). Although the micro-curvature of the surface varies in value (and therefore, so does its geometric sensor cue description) the qualitative surface description remains constant within a region (e.g. $Q_{mag}(r) = +$).

QSim generates parameter behaviours from an initial state; each behaviour is a sequence of discrete states and state transitions. State transitions can be described by a finite state machine (FSM) and for first order reflectors the FSM comprises 532 states. To illustrate, the finite state machine (FSM) segment in Figure 2.13 describes the sonar RCD behaviour for a convex surface as the robot circumnavigates the surface. The state is described by a single six digit integer and each digit encodes the qualitative value of a variable. Thus, 352230 represents a convex surface (3)
lying frontal-left (5) of a robot moving forwards (2) and turning to the left (2). The target bearing increases (3) but range decreases (0).

We mentioned earlier that *far-concave* behaviours for concave surfaces are identical to those of convex surfaces. Thus, the robot must deliberately avoid *far-concave* situations when it is attempting to localise. In general the behaviour of the robot depends on the model representation employed! However, we may overcome this particular problem using a laser time-of-flight sensor in conjunction with the sonar sensors. Since the wavelength of near infra-red laser light (typically 800nm) is significantly shorter than that of ultra-sound (typically 0.5cm), imperfections in the surface cause the laser signal to scatter back towards the sensor for any incident angle $\mu$. The qualitative behaviour
of range against bearing for a complete scan of the surface using laser light is able to disambiguate 
\textit{far-concave} and \textit{convex} surfaces when \( r + R > 0 \). To see this, if \( \mu \) is the reflectance angle of the laser signal relative to the feature normal (see Figure 2.3), then for any surface, \( R \) is minimum at \( \mu = 0 \) by Equation 2.13. Substituting \( \mu = 0 \) into Equation 2.16:

\[
\frac{d^2R}{d\Phi^2} = R \left( \frac{r \cos \mu + R}{r \cos \mu} \right).
\]  

(2.28)

<table>
<thead>
<tr>
<th>Qmag(( \frac{dR}{d\Phi} ))</th>
<th>Qmag(( \frac{d^2R}{d\Phi^2} ))</th>
<th>Qmag(( \frac{d\Phi}{d\Phi} ))</th>
<th>Surface Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>convex</td>
</tr>
<tr>
<td>0</td>
<td>+</td>
<td>0</td>
<td>plane</td>
</tr>
<tr>
<td>−</td>
<td>+</td>
<td>−</td>
<td>close-concave</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>−</td>
<td>far-concave</td>
</tr>
</tbody>
</table>

Table 2.4: Curvature inferred from sonar, gyroscope and laser cues.

Since \( R \) and \( \cos \mu \) are always positive then, by Equation 2.28, we can see that at the critical point of \( \dot{R} \), \( r \cos \mu + R > 0 \) and \( \frac{d^2R}{d\Phi^2} < 0 \) if and only if the surface is \textit{concave} \((r < 0)\). Thus, the laser is able to resolve the surface ambiguity in the sonar qualitative model when \( r + R > 0 \). However, using qualitative information alone, the laser is unable to disambiguate \textit{planar} and \textit{convex} surfaces. Hence, sonar and laser sensors must be used together to determine the surface curvature. Thus, not only does data fusion eliminate noise but it also helps to overcome ambiguity problems of the qualitative models.

To complete the curvature taxonomy we describe a qualitative method for identifying corners and edges in the \((\theta, R)\) coordinate frame. Although corner and edge RCD acoustic flows are qualitatively identical to convex surface behaviours, they can be distinguished by the laser time-of-flight sensor. Both corners and edges exhibit continuous range \( R \) profiles but discontinuous \( \frac{d\Phi}{d\Phi} \) profiles at corners and edges. The ray which intersects the corner (or edge) induces normal surface incidence angles \( \mu_1 \) and \( \mu_2 \) on each plane bordering the corner. Let \( \alpha \) be the corner (edge) angle (see Figure 2.14). Thus, \( \mu_2 = \mu_1 + \alpha \) and, by Equation 2.13, the magnitude of the step in \( \frac{d\Phi}{d\Phi} \) across the corner is:

\[
R \tan \alpha \left[ \frac{1 + \tan^2 \mu_1}{1 - \tan \mu_1 \tan \alpha} \right].
\]
When the corner (or edge) is right angled then
\[ \left| \frac{dR}{d\theta} \right| - \frac{dR}{d\theta} \right| \geq 2R \]
which is significant when \( R \geq 0.5m \) (i.e. the closest the robot should approach the target). To localise corners and edges, the image \( J \) is formed by conjoining the Gaussian smoothed curvature of the range profile \( R(\theta) \) with an image identifying positive-negative qualitative changes of gradient of \( R(\theta) \):

\[ J(\theta) = (G'' \ast R) \times |G' \ast S(G' \ast R)| \]

where \( G \) and \( S \) are the Gaussian function and sign function respectively. Edges and corners are located at, respectively, the maxima and minima of \( J \). Figure 2.15 shows this algorithm applied to two scenes; the first scene comprising an edge and corner and the second comprising a corner and two neighbouring plane walls. Figure 2.16 shows the edge-corner recognition algorithm operating on a sequence of frames obtained at robot locations separated by about 0.2m.

In summary, one approach to determine a target's curvature type is to calculate the curvature from observations using the precise Acoustic Flow model and then abstract the value obtained to a linguistic label. This approach would require complete, precise information from all sensors whose parameters appear in the model. The qualitative Acoustic Flow equations are maximally informative. Using these qualitative models a robot is able to infer surface curvature using very simple sensing devices; a ranging device which is able to detect the direction of range change and track the nearest point of a surface (which is equivalent to tracking surface normals for specular targets) and detect the surface at non-normal incidence angles, and a gyro sensor which is able to detect the direction of rotation only and locate four direction landmarks at front, left, behind and right of the robot. Targets may be discriminated by these landmarks and the RCD QDE. QR guarantees that certain combinations of sensor cues are specific to each curvature type. For example, the concept of convexity is defined in terms of sensor cues:  5

\[ \{ \text{target is left-front, sonar range is decreasing, robot has forward movement, gyro bearing increasing,} \]
\[ \text{laser has } Q \text{mag} \left( \frac{\partial \mu}{\partial \theta} \right) = + \} \lor \ldots \]

Further, by qualitatively reasoning over time QSim identifies state transitions specific to target types.

\[ 5 \left( \frac{1}{\tan \mu} \right)^2 = \frac{\tan \mu}{\sin \mu} = \frac{\cos \mu}{\cos \mu} = \frac{\cos \mu}{\cos \mu} \]

\[ 6 \text{In Chapter 7 we offer an insight into how concepts such as surface curvature may emerge automatically using correlations of qualitative sensor cues over time.} \]
Figure 2.15: Images induced by applying various steps of the edge-corner algorithm. In each graph, the dashed plot is \( g_1 \), the smoothed original range \( R \) versus bearing \( \theta \) data. The continuous plot shows the following: (1) \( S(G' * R) \); (2) \( G' * S(G' * R) \); (3) \( G'' * R \) and (4) \( \beta(\theta) \). For clarity ordinate values are scaled appropriately.

Figure 2.16: Dashed graph is the range scan. Solid graph is the intensity \( J \) locating edges (\( J > 0 \)) and corners (\( J < 0 \)). Plots are labelled in temporal order.
For example, when the omnidirectional robot is undergoing forward motion a continuous transition from $Qdir(R) = -$ to $Qdir(R) = +$ would indicate a convex or far-concave surface whereas a reverse transition, $Qdir(R) = +$ to $Qdir(R) = -$, would indicate a close-concave surface. So QR can infer minimal sets of observations which are guaranteed to identify curvature types. This information can be useful when one or more sensors become unreliable or, at the robot design stage, when choosing a sufficient sensor suite.

The following section extends the applicability of the qualitative Acoustic Flow equations to the discrimination of targets of the same surface type but with different curvature sizes.

### 2.4.1 Differential Modelling

![Differential System Geometry](image)

**Figure 2.17:** Differential system geometry.

How do we distinguish between features of the same type? So far we have considered robots that reason about each target individually. Extra information can be obtained by considering relative differences of sensor values associated with pairs of targets such as, for example, the angle subtended by two targets: $\theta_2 - \theta_1$. When $X_1$ and $X_2$ are the values for some parameter $X$ for targets 1 and 2 respectively the qualitative magnitude of the difference of these values is defined:

$$\Delta Q^2 X = Qmag(X_2 - X_1),$$

Expressions describing the qualitative behaviour of variable differentials can be obtained from the QDEs describing individual target behaviours by applying the identities in Table 2.5.\(^7\) For example,

\[
\begin{array}{l}
\Delta Q^2 A \triangleq Qmag(A_2 - A_1) \\
\Delta Q^2 (A \oplus B) = \Delta Q^2 A \oplus \Delta Q^2 B \\
\Delta Q^2 (A \otimes B) = \Delta Q^2 A \otimes Qmag(B_2) \oplus Qmag(A_1) \otimes \Delta Q^2 B \\
\Delta Q^2 M^+(A) = \Delta Q^2 A \\
\Delta Q^2 M^-(A) = \ominus \Delta Q^2 A \\
\Delta Q^2 Abs(A) = \Delta Q^2 A \otimes Qmag(A_1 + A_2)
\end{array}
\]

**Table 2.5:** Differential identities where $A$ and $B$ are qualitative expressions and $A_1$ and $A_2$ are quantitative values.

---

\(^7\)The qualitative absolute value of a parameter is denoted $Abs(A)$ and $\Delta Q^2 Abs(A) = Qmag(A_2^2 - A_1^2) = Qmag((A_2 - A_1)(A_2 + A_1))$.\)
a sonar-gyro mounted RCDD unit mounted on a robot that is moving forward and observing a convex specular target either left-frontal, frontal or right-frontal behaves according to the following QDE obtained from Equation 2.10:

\[ Q_{\text{mag}}(\theta) = Q_{\text{dir}}(\theta) \otimes (Q_{\text{mag}}(\mathcal{R}) \oplus Q_{\text{mag}}(r)). \]

Applying the differential transformation rules we obtain:

\[ \Delta_{R}^{2} \hat{\theta} = \Delta_{\theta}^{2} \hat{\theta} \otimes Q_{\text{mag}}(\theta_{1}) \otimes (\Delta_{R}^{2} \mathcal{R} \oplus \Delta_{R}^{2} r). \]

(2.29)

To illustrate the utility of the differential model we will show that, under certain conditions, the robot can use Equation 2.29 to determine which of two convex targets has the largest curvature. Identifying targets 1 and 2 as in Figure 2.17 suppose the following qualitative values are observed:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Qualitative value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta_{R}^{2} \hat{\theta} )</td>
<td>+</td>
</tr>
<tr>
<td>( \Delta_{\theta}^{2} \hat{\theta} )</td>
<td>-</td>
</tr>
<tr>
<td>( Q_{\text{mag}}(\theta_{1}) )</td>
<td>-</td>
</tr>
<tr>
<td>( \Delta_{R}^{2} \mathcal{R} )</td>
<td>+</td>
</tr>
</tbody>
</table>

then the only value for \( \Delta_{R}^{2} r \) which satisfies Constraint 2.29 is \( \Delta_{R}^{2} r = - \). Thus feature 2 has a smaller surface curvature than feature 1. However, Constraint 2.29 can only be used under quite specific conditions as not all qualitative values for the observation variables yield informative values for \( \Delta_{R}^{2} r \). Two alternative approaches exist and their relative merits depend on assumptions of the operation of the robot and the composition of the environment. For both methods each object is explored separately so that observation generated landmarks can be compared between the targets. We present one method here. The second method is presented in the next section.

Suppose that two convex targets are explored individually by a robot which is constrained to move identically in each case. The robot exhibits identical speed (i.e. \( \Delta_{V}^{2} \hat{V} = 0 \)) and identical turning rates (i.e. \( \Delta_{\theta}^{2} \hat{\theta} = 0 \)) at corresponding times for each exploration. Also suppose that, prior to each exploratory behaviour, the robot positions itself so that the target is at a specific range and bearing (thus initially \( \Delta_{R}^{2} \mathcal{R} = 0 \) and \( \Delta_{\theta}^{2} \hat{\theta} = 0 \)). Figure 2.18 shows the QSIm generated envisionment for the differential Acoustic Flow equations for two convex surfaces:

\[ \Delta_{R}^{2} \hat{\theta} = \Delta_{\theta}^{2} \hat{\theta} \otimes (\Delta_{R}^{2} \mathcal{R} \oplus \Delta_{R}^{2} r) \]

(2.30)

\[ \Delta_{R}^{2} \hat{\mathcal{R}} = \Delta_{\theta}^{2} \hat{\mathcal{R}} \]

(2.31)

operating under the above constraints. In Figure 2.18, label 1 is assigned arbitrarily and without loss of generality to the target with the largest curvature (i.e. \( \Delta_{R}^{2} r = - \)). Also, each target is assumed to be in quadrant 1 at all times so that \( Q_{\text{mag}}(\theta_{1}) = + \).

All behaviours are identical up until time \( t_{2} \) and in the period \( 0 \) until \( t_{2} \):

\[ \Delta_{R}^{2} r = \Delta_{R}^{2} \hat{\theta} \]

(2.32)

and the relative curvature is immediately inferable from the relative sizes of the observable \( \hat{\theta} \). QSIm also informs us that Equation 2.32 persists while \( \Delta_{R}^{2} (R + r) = - \) which is a significant operating region when the target curvatures are distinct.

\footnote{QSIm normally searches for behaviour loops (i.e. repeated states) within a behaviour and terminates simulation of looping behaviours. QSIm was amended so that it detected and truncated behaviours which formed loops operating across different behaviours. This amendment radically reduced the number of states in the envisionment graph. A looping behaviour is marked \( \otimes \) in the envisionment graph.}
2.5 The Physical Reflector Model

The efficacy of Equation 2.32 was examined for two symmetric cylinders, a wide cylinder of radius 12cm and a narrow cylinder of radius 1cm. Figure 2.19 shows the time evolution of $\dot{\theta}$ for each cylinder including the narrow cylinder when observed under "slippery" conditions. Slip was simulated by forcing the robot to move with a speed about 10% slower than normal. For each data set the robot moved in a straight line and the target traversed a significant range of bearings (approximately 0.7 rads to about 1.5 rads). Under conditions of no slip, $E_t(\delta \theta_{\text{narrow}}(t) - \delta \theta_{\text{wide}}(t)) = 0.006 \pm 0.008$ and the cylinders were correctly discerned although with little conviction. However, when the narrow cylinder was reexamined under slip and compared with the wide cylinder under no slip, $E_t(\delta \theta_{\text{narrow}}(t) - \delta \theta_{\text{wide}}(t)) = -0.0120 \pm 0.006$. In this case the wrong target was assigned the largest radius of curvature.

In summary, this method is not too successful at discriminating degrees of curvature. Further, when operating conditions are non-repeatable, for example when slip is a possibility, then this method is a potential hazard. An alternative curvature discrimination method is offered in the next section.

Figure 2.18: QSIM envisionment for QDE 2.30 and 2.31.

2.5 The Physical Reflector Model

The Acoustic Flow model is a kinematic approach to feature recognition and works by matching the motion of a sonar sensor cue behaviour to a sensor frame. Alternatively, feature information can be obtained through detailed investigation of the echo profile and, in this section, we describe a qualitative model for echo response from a frequency-modulated sonar device. Two properties of a target, the shape and the material, exhibit the most important influence over how much acoustic energy scatters back towards the receiver. Bozma and Kuç developed the ENDURA (Energie-DURation-Range) method (Bozma and Kuç, 1994) for discriminating surface texture using variations in echo energy and duration. The energy $P$ and duration $D$ of an echo over a time window $\tau$ is a function
of the acoustic echo pressure $P$:

\[ P = \int_0^T P^2(t) dt \]

\[ D = \frac{\int_0^T (t - \bar{t})^2 P^2(t) dt}{\int_0^T P^2(t) dt} \]

where $\bar{t} = \frac{\int_0^T t P^2(t) dt}{\int_0^T P^2(t) dt}$. The magnitude of $P$ is determined by absorption and geometric spreading and is a function of the target shape, reflectivity and its distance from the transducer (Moran, 1994). Bozma and Kuc demonstrate that planar diffuse and specular targets can be discriminated using a qualitative description of their energy-duration profiles:

The smooth surface ... is easily identified from the rapidly decaying echo-energy curve within $\pm \theta_0$ from normal incidence and the constant echo duration over this range. The rough surface ... is recognised by a steady decrease in the echo-energy curve and an increase in the echo-duration curve. The moderately rough surface ... is identified mainly from the echo-energy curve which has two distinct regions. Within $\pm \theta_0$ from normal incidence, a strong specular reflection is seen, whereas for other incidence, the incoherent component of the signal dominates, causing a steadily decreasing echo-energy curve.

Echo pressure $P$ is also affected by scattering due to inhomogeneities in the propagating medium, target cross-section and the interaction of the echo with the transducer. For example, rain reduces the echo intensity slightly. Rain is a weak diffuse reflector (see Figure 2.20). Echoes from a particular narrow, hollow, plastic cylinder exhibited $E(P) = 94.8 \pm 1.5$ in dry conditions and $E(P) = 90.3 \pm 1.5$ in rainy conditions. The sample standard deviation is 7.0 in both cases. Also moisture affects surface reflectivity. A specular, internally corrugated cardboard cylinder produced an echo $E(P) = 250.6 \pm$
8.8 when dry and \( E(P) = 228.7 \pm 5.3 \) when its surface was wet. The sample standard deviations are 22.8 and 37.6 respectively. The mean pressure value difference in wet and dry conditions is sufficient to warrant a degree of modelling imprecision.

![FM Sonar Response](image)

(a) Dry  
(b) Rain

**Figure 2.20:** Characteristic FM sonar echoes in dry and heavy rain conditions.

### 2.5.1 Qualitative Response Models

Using conservation of energy Keller (Keller, 1962) showed that the electric fields \( E_R \) of optical rays reflected from regular curved surfaces obey:

\[
E_R(s) = E_R(0) \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + s)(\rho_2 + s)}}. 
\]  

(2.33)

where \( \psi(0) \) and \( \psi(s) \) are two reflected wavefronts (surfaces of constant phase), \( s \) is the axial distance between \( \psi(0) \) and \( \psi(s) \) surfaces and \( \rho_1 \) and \( \rho_2 \) (both of which depend on \( R \)) are the principal radius of curvature of the reflected wavefront at \( s = 0 \). Either \( s \) is an elemental distance or, since the rays are axial, then \( s \) is the distance of the surface \( \psi(s) \) from the caustic point and \( (\nabla \psi \cdot \nabla) = \psi \cdot \nabla = d/s \) (see Figure 2.21). McNamara (McNamara and Pisorius, 1990) showed that Equation 2.33 applies to any field \( X_R \) when the following condition holds:

\[
\frac{dX_R}{ds} + K(\nabla^2 \psi)X_R = 0, 
\]  

(2.34)

We will show that the acoustic wave pressure \( P \) satisfies Equation 2.34. For any two surfaces \( \psi(s = 0) \) and \( \psi(s = s) \) with infinitesimal area at each end of an arbitrary tube of volume \( v \) which is centred on a ray:

\[
\int_v (\nabla \cdot (P \nabla \psi)) \, dv = \int_s P(s) \, ds - \int_0 P(s) \, ds = 0. 
\]

Thus, to first order, \( \nabla \cdot (P \nabla \psi) = 0 \) and:

\[
\frac{dP}{ds} = (\nabla \psi \cdot \nabla)P = \nabla \cdot (P \nabla \psi) - P \nabla^2 \psi = -P \nabla^2 \psi 
\]  

(2.35)
2.5 The Physical Reflector Model

![Diagram of an infinitely narrow diverging astigmatic ray tube](image)

**Figure 2.21:** Infinitely narrow diverging astigmatic ray tube, for which both \( \rho_1 \) and \( \rho_2 \) are positive.

which satisfies Equation 2.34. Thus:

\[
\frac{P_R(s)}{P_R(0)} = \sqrt{\frac{\rho_1 \rho_2}{(\rho_1 + s)(\rho_2 + s)}}.
\]

When \( \rho_i \) is a principal radius of curvature of the incident wavefront at the target surface, \( r \) is the local surface radius of curvature and \( \mu \) is the angle of reflection of some ray (see Figure 2.4), the principal radius of curvature \( \rho_r \) of the reflected wavefront at the surface is (McNamara and Pisrciorus, 1990), pp 88–91):

\[
\frac{1}{\rho_r} = \frac{1}{\rho_i} + \frac{2}{r \cos \mu}.
\]  

Thus, if a plane wave is incident on a 2D surface such as a regular cylinder, edge or corner, the radius of curvature of the reflected wave will be an increasing function of the radius of curvature of the target surface. Two simple cases will demonstrate the utility of Equation 2.33. In the far-field the rays are axial and \( r \approx R + \rho_r \). Reflections from a plane, for which \( \rho_1 \rightarrow \infty \), obey \((\forall R)\ P_R(s) = P(0)\), those from a corner or edge, for which \( \rho_i = R \), obey \( \frac{P_R(s)}{P_R(0)} = \frac{R}{R+s} \). The acoustic echo also undergoes lobe spreading due to interference effects and attenuation by the atmosphere. Lobe spreading is inversely proportional to \( R \) (Morse and Ingard, 1968) and attenuation is exponential \((P_A(R) \propto \exp(-kR))\).

How can ultrasound information be used for target identification and discrimination? Two possibilities immediately suggest themselves. We may model all parameters of the reflector and then match an observation with models for each possible target. Alternatively, we may adopt a qualitative approach and match new observations with observations already attributed to targets. The second approach requires that there is a target invariant mapping of sensor values into the truth values \((T, F)\). We will now develop such a mapping for pressure-range observations from the FM sonar. Excluding curved concave surfaces, any two pressure-range observations for a rotationally

---

9A plane wave reflected from a 2D concave cylindrical surface will converge to a point (a caustic point) and subsequently diverge (see Figure 2.22). Thus, within the near-converging zone \( P_R \) increases as one moves away from the surface. However, this knowledge is countered by the fact that \( P \) decreases by lobe spreading and is further degraded by dispersion. Thus, we cannot draw a conclusion about the qualitative trend of \( P \) in the near-converging zone.
symmetric specular target obey:

\[
\frac{P_1}{P_2} = \left( \frac{R_2}{R_1} \right)^n \exp(k[R_2 - R_1]) + \nu
\]  

(2.37)

where \( \nu \) is zero-mean noise, \( n \in (-1.5, -1.0) \) for 2D surfaces (e.g., cylinders, edges, corners) and \( n \in (-2.0, -1.0) \) for 3D surfaces (e.g., spheres, point targets). In air with humidity between 20% and 30%, the absorption coefficient \( k \in (0.4, 3.0) \) over the full range of FM sonar scanning frequencies (45 kHz to 90 kHz (Lee, 1995)) (Bass et al., 1972) and reproduced in (Kinsler et al., 1982)). Figure 2.22 shows echo pressure-range profiles for specular reflectors with various curvatures.

![Figure 2.22: Pressure-range plots for (a) a plane, a wide cylinder (W. Cyl.) of radius 12cm and a narrow cylinder (N. Cyl.) of radius 1cm and (b) a plane using reduced sensor output energy compared with the plane data from part (a).](image)

Equation 2.37 is a function of surface curvature, instantaneous echo frequency and humidity. Obtaining an accurate value for \( n \) can be time intensive. We may avoid specifying \( n \) and \( k \) precisely by reasoning with static envelopes. Static envelopes (Kuipers, 1994) denote the extreme values, \( l \) and \( u \), of an imprecise function. In our case:

\[
l(R_1, R_2) = \sup_{n \in (-1.5, -1.0), k \in (0.4, 3.0)} \left( \frac{R_2}{R_1} \right)^n \exp(k[R_2 - R_1])
\]

\[
u(R_1, R_2) = \inf_{n \in (-1.5, -1.0), k \in (0.4, 3.0)} \left( \frac{R_2}{R_1} \right)^n \exp(k[R_2 - R_1])
\]

Two observations, \((R_1, P_1)\) and \((R_2, P_2)\), can possibly originate from the same target only if \( l(R_1, R_2) \leq \frac{P_2}{P_1} \leq u(R_1, R_2) \) and necessarily originate from different targets when \( l(R_1, R_2) > \frac{P_2}{P_1} \) or \( \frac{P_2}{P_1} > u(R_1, R_2) \). In the next section we develop a purely qualitative approach to target data association for the FM sonar and then compare the qualitative and static envelope models.

### 2.5.2 FM Sonar Sensor Qualitative Rules

When the sensor is uncalibrated (so that readings are ordinal only) Equation 2.37 generalizes to a simple ordinal relationship between range and pressure information: \( P = M^{-}(R) \). In general, the
following rules capture the qualitative dependency of echo pressure on a convex or planar target’s distance, roughness and curvature. Rules R-1 and R-2 were developed above. Rules R-3 and R-4 are from Bozma and Kuc’s paper.

**Pressure-range.** Echo energy decreases with range for convex targets:

(R-1) \( [\kappa > 0] \supset [P = M^- (R)] \)

**Pressure-curvature.** For carrier wavelengths smaller than the radius of curvature, cross-section is inversely proportional to surface curvature:

(R-2) \( [\kappa > 0] \supset [P = M^- (\kappa)] \)

**Energy-roughness.** Echo energy decreases with roughness:

(R-3) \( P = M^- (\sigma_s) \)

These rules may be used for target recognition in conjunction with or in place of the kinematic equations developed earlier in this chapter. The maximum acoustic pressure value and its corresponding range value are good indicators for target recognition:

\[
P = \max_R P(R), \quad R = \text{argmax}_R P(R).
\]

We may use our knowledge of the physics of sonar and previous observations to partition the Range-Pressure space into regions which identify the possible observations consistent with each feature. By Rule 2.5.2, when two observations indexed 1 and 2 are associated with targets \( \text{target}_1 \) and \( \text{target}_2 \) respectively:

\[
(\Delta^2_1 R = -) \land (\Delta^2_1 P = -) \lor (\Delta^2_2 R = +) \land (\Delta^2_2 P = +) \supset [\text{target}_1 \neq \text{target}_2]. \quad (2.38)
\]

Each observation partitions range-pressure space regions denoting range-pressure values which are possibly consistent with the observation and those denoting range-pressure values which are necessarily inconsistent with the observation. This information can be used to simultaneously associate observations with targets and refine the model for each target. This is illustrated using data in Figure 2.22; specifically, observations 5 and 7 from the wide cylinder and 5 and 6 from the narrow cylinder. Suppose that these observations were obtained in order w-7, n-6, n-5 and then w-5 (see Table 2.6). Figure 2.23 shows the stages of partitioning given observations in this order.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Range (±2)</th>
<th>Pressure (±20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (w7)</td>
<td>150</td>
<td>137</td>
</tr>
<tr>
<td>2 (n6)</td>
<td>129</td>
<td>82</td>
</tr>
<tr>
<td>3 (n5)</td>
<td>112</td>
<td>121</td>
</tr>
<tr>
<td>4 (w5)</td>
<td>117</td>
<td>323</td>
</tr>
</tbody>
</table>

**Table 2.6: CTFM sonar data.**

The qualitative ordinal model (Equation 2.38) is compared with the semi-quantitative model (Equation 2.37) in Figure 2.24. For the semi-quantitative approach, interval curvature dependence
2.5 The Physical Reflector Model

![Pressure vs range graphs](image)

Figure 2.23: Pressure versus range response for FM sonar optimally directed at cylinders with marginally different curvatures. Hatching indicates "is not consistent with features observed"; dark shading indicates "consistent with only one feature" and white space indicates "consistent with two or more features observed".

\[ n \in (-1.5, -1.0) \] and absorption coefficient \( k \in (0.4, 3.0) \) were used in Equation 2.37 and \( \nu \) was assumed to have a standard error \( \sigma = 0.1 \). Two observations are deemed to belong to different targets when \( \nu > 2\sigma \). According to Figure 2.24 the qualitative method would correctly associate all observations with their respective targets.

Although both ordinal and semi-quantitative models are successful in this example they could easily have not succeeded for separate reasons. If observations \( n_5 \) and \( n_6 \) had not been made then the qualitative method, unlike the full numeric method, would have been unable to choose between associating \( \{n_1, n_2, n_3, n_4\} \) with \( \{n_7, n_8, n_9, n_10\} \) or with \( \{n_5, n_6, n_7\} \).

This approach to target discrimination is less likely to be successful with irregular surfaces which exhibit a large variation of corresponding pressure-range values. A combination of kinematic and differential physical modelling is required in these circumstances to identify distinctive parameter trends over a target surface. For example, the radius of curvature of a particular convex, inwardly spiralling surface under investigation varied gradually from about 1m to about 0.3m. At a distance of 1m from the surface the echo pressure varied significantly between different locations (as shown in Figure 2.25). However, since range \( R \) is fixed then the gradual increase in curvature as observed at location 1 though to 4 can be inferred using \( P = M^{-\alpha} \).

So far, we have shown how qualitative reasoning can be useful for target type recognition and sub-type discrimination using the Acoustic Flow model. We have also seen how a physical qualitative model of the echo pressure dependency on target curvature and range can be used to simultaneously associate observations to targets and refine target models. However, not all of these models are
Figure 2.24: The lower and upper diagonal matrices describe the qualitative and semi-quantitative model observation associations respectively. For an arbitrary row \( r \) and column \( c \) a ■ indicates a possible association between observation \( r \) and \( c \), a □ indicates that \( r \) and \( c \) necessarily belong to different targets and a × indicates those associations qualitatively allowed but excluded by the semi-quantitative method.

![Diagram]

<table>
<thead>
<tr>
<th>Location</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure</td>
<td>2456</td>
<td>2400</td>
<td>2162</td>
<td>1846</td>
</tr>
</tbody>
</table>

Figure 2.25: Pressure measurements for various locations \( 1m \) away from a 2D, cardboard, smooth spiral surface.

guaranteed to be informative in all circumstances and the robot must act strategically in order to gather information.

### 2.6 Sensor Management

Three cases where qualitative and semi-quantitative reasoning can be useful for sensor management and robot control are illustrated in this section. No general reasoning algorithm is offered. The first example illustrates the problem posed at the end of the previous section, namely that of informative sensor placement given that all reasoning is done using qualitative models. Often, the robot is limited in the amount of time it can spend exploring an object before it must decide on a classification, and often sensor imprecision dictates that a decision can be made only after a distinct change is observed.
in the sensor data. The second example illustrates semi-quantitative reasoning used as a refinement process which searches for robot behaviours which satisfy both time and data constraints. The third example illustrates how QR can be used to verify robot navigation and obstacle avoidance feedback control laws.

2.6.1 Qualitative Reasoning Towards Action

A robot is equipped with an FM sonar and a map of the environment which contains pressure-range values for a number of objects. The map hints that the current target in the robot's view is one of two objects. However, the current observation does not distinguish between objects as the observation is consistent with the qualitative models for both. The robot must determine how to act to recognise the target.

Two arbitrary corresponding values are chosen, one from each internal object model. These are labelled 1 and 2. We want to place the sensor so that the echo pressure unambiguously identifies the target. In other words, if \(3\) denotes the new observation, we want to find \(\Delta^1_3 R\) and \(\Delta^2_3 R\) so that \(\Delta^1_3 P\) or \(\Delta^2_3 P\) is guaranteed to indicate that one (or more) of the mapped objects is inconsistent with the new observation. Each potential robot position is a possible world, \(W_{i,j,k}\) where \(i, j, k \in \{1, 2, 3\}\), comprising differential range, pressure and curvature information:

\[
W_{i,j,k} = \langle \Delta^1_3 R, \Delta^2_3 R, \Delta^1_3 P, \Delta^2_3 P, \Delta^1_3 \kappa, \Delta^2_3 \kappa \rangle .
\]

Possible worlds are limited by the differential consistency constraint:

\[
(\forall i, j, k \in \{1, 2, 3\}) \Delta^1_3 X = \Delta^2_3 X \uplus \Delta^3_3 X
\]

and the physical model \(\Delta^1_3 P \land \Delta^2_3 R \supset \Delta^3_3 \kappa\).

A possible world \(pw_1\) is accessible to another \(pw_2\) if they share the same differential values for the observables \(R\) and \(P\). A possible world is discarded if the same observable differentials are consistent with multiple values for both \(\Delta^3_3 \kappa\) and for \(\Delta^3_3 \kappa\), since these discarded worlds are unable to disambiguate the objects. The remaining worlds describe placements and pressure differential observations which unambiguously disassociate the new observation from an object. The remaining worlds are grouped so that, within each group, each world has the same range \(R\) differentials but one or more different \(P\) differentials and one or more different curvature \(\kappa\) differentials. The largest groups have the greatest discrimination power.

Two action descriptions are generated by this algorithm:

1. When \(\Delta^1_3 P = +\) and \(\Delta^1_3 R = +\) then choose \(\Delta^3_3 R = -\) and \(\Delta^3_3 R = +\) for which:

\[
[\Delta^1_3 P = +] \land [\Delta^3_3 P = +] \supset [\Delta^3_3 \kappa = -]
\]

and:

\[
[\Delta^1_3 P = +] \land [\Delta^3_3 P = +] \supset [\Delta^3_3 \kappa = +]
\]

2. When \(\Delta^1_3 P = +\) and \(\Delta^1_3 R = -\) then choose \(\Delta^3_3 R = +\) and \(\Delta^3_3 R = -\) for which:

\[
[\Delta^1_3 P = +] \land [\Delta^3_3 P = +] \supset [\Delta^3_3 \kappa = -]
\]

and:

\[
[\Delta^1_3 P = +] \land [\Delta^3_3 P = +] \supset [\Delta^3_3 \kappa = +].
\]
Thus, the robot has the best chance of recognising the target when the robot is placed between ranges $R_1$ and $R_2$.

### 2.6.2 Action Refinement

A robot uses a semi-quantitative Acoustic Flow model (Equations 2.10 and 2.12) to determine whether a target is specularly convex or concave by observing the total change ($\Delta \theta$) in range (i.e. $\Delta R$) and bearing (i.e. $\Delta \theta$) to the target after moving a distance $X$. Setting $\mu = 0$ in Equations 2.10 and 2.12 and integrating over $X$ (where $dX = Vdt$) we obtain:

$$
\Delta \theta = \int_0^X \frac{\sin \theta}{R + r} dX,
$$

$$
\Delta R = -\int_0^X \cos \theta dX.
$$

In order to determine the surface curvature, the noise of the system and mechanical properties require that the RCDD must turn in excess of 0.1 rads before a definite classification is assured. The robot is initially 5m away from a reflector which, if a convex surface, is known to have a radius of curvature varying between 0.8 and 1.0 metres. The robot contemplates its possible actions to determine whether the surface is concave or convex. The robot requires only that by moving less than 10m the gyro will turn through at least 0.1 radians to indicate a convex surface. The parameter values in Equations 2.39 and 2.40 are thus constrained throughout the exploratory behaviour of the robot:

$$
\begin{align*}
 r & \in [r_1, r_2] = [0.8, 1.0], \\
 \theta & \in [\theta_0, \theta_0] + [0, \Delta \theta], \\
 R & \in [R_0, R_0] + [-1, 1] \times [0, \Delta R] = [5, 5] + [-1, 1] \times [0, \Delta R]
\end{align*}
$$

where $R_0$ and $\theta_0$ is the initial range and bearing to the target respectively. Combine Equations 2.39 and 2.40:

$$
\Delta \theta = \frac{[0, 1] \times \sin([\theta_0, \theta_0] + [0, 1] \times [0, \Delta \theta])}{[0.8, 1.0] + [R_0, R_0] + [-1, 1] \times \cos([\theta_0, \theta_0] + [0, \Delta \theta])}.
$$

Equation 2.41 is a constraint satisfaction problem which is solved by iteratively recalculating $\Delta \theta$ and $\theta_0$ and combining the results with the previous estimates using interval intersection. The goal is to find values for $\theta_0$ which guarantee that $\Delta \theta \geq 0.1$. The following table shows the initial parameter value ranges and final stable ranges obtained after 5 iterations:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Interval</th>
<th>Final Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>$(1.0, 1.0)$</td>
<td></td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>$(0.0, 2\pi)$</td>
<td>$(0.563, 2.447)$</td>
</tr>
<tr>
<td>$R_0$</td>
<td>$(5, 5)$</td>
<td></td>
</tr>
<tr>
<td>$\Delta \theta$</td>
<td>$(0.1, \infty)$</td>
<td>$(0.100, 0.132)$</td>
</tr>
<tr>
<td>$X$</td>
<td>$(0.0, 1.0)$</td>
<td></td>
</tr>
</tbody>
</table>

Thus, moving with an initial bearing between 0.563 and 2.447 rads to the target, a convex surface will yield a bearing shift greater than 0.1 radians over 1.0m of travel.
2.6.3 Control Verification

Finally, we demonstrate the use of QRs and specifically non-numeric landmarks and qualitative simulation, to investigate the consequences of a motion control law. Suppose a robot, navigating using the laser sensor, is moving along a corridor and approaching a doorway through which it intends to pass. A simple door navigation rule is proposed and investigated for potential collision behaviours. The robot turns in an attempt to minimise the difference between the angles subtended at it by each vertical segment of the door frame:

$$\frac{d\phi}{dt} = C(\theta_1 + \theta_2)$$  \hspace{1cm} (2.42)

where $C$ is some positive constant and the segments are labelled 1 and 2. The forward looking laser sensor is assumed to maintain both targets in its field of view at all times. Both door-frame targets are assumed to be edge type and the robot is assumed to maintain continuous forward motion so that:

$$\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}, \text{ Qmag}(V) = +, \text{ Qmag}(r_1) = 0, \text{ Qmag}(r_2) = 0.$$

Since the targets are in different places:

$$\begin{align*}
[\text{Qmag}(R_1) = 0] & \land [\text{Qmag}(R_2) = 0] \quad \lor \quad \bot, \\
[\Delta_2^3 \theta = 0] & \land [\Delta_2^3 R = 0] \quad \lor \quad \bot.
\end{align*}$$

**Acoustic Sonar Flow QDE.** The acoustic sonar flow Equation 2.10 with $r = 0, \mu = 0, \text{ Qmag}(\sin \theta) = \text{ Qmag}(\theta)$ and $Qdir(\sin \theta) = Qdir(\theta)$ becomes:

$$\begin{align*}
Qdir(\theta_1) \oplus Qdir(\phi_1) &= \text{ Qmag}(\theta_1), \\
Qdir(\theta_2) \oplus Qdir(\phi_2) &= \text{ Qmag}(\theta_2).
\end{align*}$$

The qualitative differential operator applied to Equation 2.10 for two edge type targets gives:

$$\begin{align*}
Qdir(\Delta_2^3 \theta) &= Qdir(\theta_2) \circ Qdir(\theta_1), \\
Qdir(\Delta_2^3 \theta) &= \Delta_2^3 \theta \circ \Delta_2^3 R \circ \text{ Qmag}(\theta_1), \\
Qdir(\Delta_2^3 \theta) &= \Delta_2^3 \theta \circ \Delta_2^3 R \circ \text{ Qmag}(\theta_2).
\end{align*}$$

Similarly, the qualitative differential operator applied to Equation 2.12 gives:

$$Qdir(\Delta_2^3 R) = \circ \Delta_2^3 (\cos \theta) = \Delta_2^3 \text{ Abs}(\theta).$$

Finally, by Table 2.5:

$$\text{ Qmag}(\Delta_2^3 \text{ Abs}(\theta)) = \Delta_2^3 \theta \circ \text{ Qmag}(\theta_1 + \theta_2)$$

and:

$$\begin{align*}
Qdir(\Delta_2^3 \text{ Abs}(\theta)) &= \text{ Qmag}(\theta_2) \circ Qdir(\theta_2) \circ \text{ Qmag}(\theta_1) \circ Qdir(\theta_1), \\
Qdir(\Delta_2^3 \text{ Abs}(\theta)) &= \Delta_2^3 \theta \circ Qdir(\theta_1) \circ Qdir(\Delta_2^3 \theta) \circ \text{ Qmag}(\theta_2), \\
Qdir(\Delta_2^3 \text{ Abs}(\theta)) &= \Delta_2^3 \theta \circ Qdir(\theta_2) \circ Qdir(\Delta_2^3 \theta) \circ \text{ Qmag}(\theta_1).
\end{align*}$$
2.7 Conclusions

Control QDE. The QDE control law is:

\[ Q_{dir}(\phi) = Q_{mag}(\theta_1) \oplus Q_{mag}(\theta_2), \]
\[ Q_{dir}(\phi) = Q_{mag}(\theta_1 + \theta_2). \]

The QDE is simulated using QSim from the following initial conditions which are observed by the robot when approaching the doorway:

\[ Q_{mag}(\theta_1) = -, \quad Q_{mag}(\theta_2) = -, \quad Q_{mag}(R_1) = +, \quad Q_{mag}(R_2) = +, \]
\[ \Delta^2_t \theta = -, \quad \Delta^2_t R = +, \quad \Delta^2_t R = +. \]

After approximately 11.17 hrs QSim returns an environment comprising 1445 behaviours built from 392 individual states. These behaviours were examined for potential collisions with the walls. A simple query tool was developed which operated on the environment. A partial “goal” state query is declared which matches one or more states in the environment. All non-looping behaviours which reach a goal state are returned. The number of behaviours is often prohibitive but they can be aggregated by removing irrelevant distinctions. The user declares a set of relevant parameters. State counts are pruned of all non-relevant parameters and then each behaviour is pruned so that no neighbouring states are identical. Finally, duplicate behaviours are removed.

A head-on collision with target 1 was investigated with the following search query:

\[ \text{Query}([Q_{mag}(\theta_1) = 0] \land [Q_{mag}(R_1) = 0]) \lor ([Q_{mag}(\theta_2) = 0] \land [Q_{mag}(R_2) = 0]) \]

and relevant variables \( \theta_1, \theta_2, \Delta^2_t \theta, \Delta^2_t R \) and \( \Delta^2_t \text{Abs}(\theta) \). Behaviours which involved passing through walls were suppressed by insisting that all behaviours were not searched beyond a state containing \( \Delta^2_t \theta = 0 \). No behaviours corresponding to collision with target 1 and one aggregate behaviour was found which includes a collision with target 2 (which is shown in Figure 2.26). Also, by posing the question:

\[ \text{Query}([\Delta^2_t \theta = 0] \land [\Delta^2_t R = +]) \]

it was found that no behaviours lead to a collision with the wall connected to target 1 and, using:

\[ \text{Query}([\Delta^2_t \theta = 0] \land [\Delta^2_t R = -]), \]

only one aggregate behaviour leads to a collision with wall 2.

Both collision behaviours arise when the turn rate is slow and the sign of \( \Delta^2_t \text{Abs}(\theta) \) persists over consecutive states (in this case states 1, 2 and 3). Thus, a large value for \( C \) will reduce the possibility of collision. Of course, we have investigated an ideal robot with no delays between environment conditions, perception and action.

2.7 Conclusions

This chapter has demonstrated the utility of qualitative representations for various robot navigation tasks such as object recognition, model building and control. We began by developing the Acoustic Flow equations which are sensor centred differential equations relating observation parameters to target curvature or reflector order number. This extends the work on sonar sensing. "Regions of
Constant Depth" introduced by (Kuc and Siegel, 1987; Leonard and Durrant-Whyte, 1992) to surfaces of non-constant curvature. These models were verified using the in-air ultrasonic Polaroid transducer and an infrared laser range finder.

The Acoustic Flow model was transformed into an abstract language of differential equations based on the sign algebra and directions of change of variable values. This yielded a logical Acoustic Flow model based on natural landmarks of the system; the critical points of the underlying mathematical equations. It was shown how a few (well chosen) landmarks can be sufficient for the qualitative disambiguation of surface curvature by ordinal reasoning. However, certain qualitative operations, especially sign addition, required careful treatment for the model to be informative. A troublesome constraint was removed by adjusting the hardware layout of the robot and mounting the gyroscope on the servo-mounted RCDD unit. 10 One further source of ambiguity was removed by sensor data fusion; by combining qualitative laser (diffuse sensing) time-of-flight information with qualitative sonar (specular sensing) kinematic time-of-flight information. Thus, the qualitative Acoustic Flow equations are maximally informative with respect to surface curvature recognition. So, in theory, a robot is able to infer surface curvature using very simple sensing devices; a ranging device which is able to detect the direction of range change and track the nearest point of a surface (which is equivalent to tracking surface normals for specular targets) and detect the surface at non-normal incidence angles, and a gyro sensor which is able to detect the direction of rotation only and locate four direction landmarks at front, left, behind and right of the robot. Application issues for this model are explored in Chapter 6.

To extend the robot's recognition abilities, differential models for comparing the magnitudes of target attributes was introduced. Such models require no further domain information. Although these models can be used to rank target curvature sizes their range of applicability is limited. This demonstrates a common problem with qualitative representations, that only a subset of the robot's allowable actions generate informative inferences. FM ultrasonic sensors supply a greater range of information types (pressure, energy, duration and range) and we used this information, along with the differential modelling technique, to distinguish convex surfaces with different surface curvatures. Qualitative rules were introduced which can be used for simultaneous recognition and model building.

10 An interesting analogy/question arises in natural evolution. To what extent is our physical make-up constrained by our mental representation? A creature with a mental representation in harmony with its physical characteristics has the ability to take more advantage of each and therefore one would expect a more favourable chance of survival. Perhaps our vestibular and visual systems are co-mounted on the head for similar reasons that we chose to mount the gyro sensor on the servo-mounted sonar sensor?
This approach demonstrates the utility of reasoning with a combination of qualitative “text-book” physics information and case specific dynamic landmarks.

Finally, we explored the use of semi-quantitative sonar models for action planning and control. The robot can specify desirable parameter operation ranges and QDE system constraints and bounding envelopes relating their values. Through gradual refinement, a range of consistent parameter values are obtained which indicate the optimal robot control actions.

Early in this chapter we noted that, since curvature is a second order derivative, it is noisy (Asada and Brady, 1986). Real data from the robot domain demonstrated the extent to which it is corrupted by noise. Noise filtering and parameter estimation using qualitative models to govern sensor data fusion are the focus of Chapters 3, 4 and 5.
Chapter 3
The Qualitative Filter

While research in qualitative physics is aimed at uncovering how humans reason about continuous physical quantities, relatively little research has been done on how they process the uncertainties connected with those quantities. For example, while it is common knowledge that both temperature and time have a decisive effect on the quality of your barbecued hamburger, it is not quite clear how the colour of the flame and the sound of sizzling translate into uncertainty about the temperature, and how this uncertainty translates into a decision about when to come back to inspect the progress of your dinner. Less esoterically, if the time shown on your car clock is different than that shown on your watch, how do you combine the two readings to determine if you have time to visit a friend on the way home?

J. Pearl
Probabilistic Reasoning in Intelligent Systems [1988]

3.1 Introduction

The previous chapter demonstrated that qualitative representations are potentially very useful for robot navigation. In a noise-free environment qualitative behaviours that are inconsistent with observations can be pruned. This is the mechanism underlying Forbus’ measurement interpolation algorithm (Forbus, 1986). However, if the observations or the dynamics of the system are subject to random fluctuations then this step is not valid as no behaviours are truly inconsistent. In such a situation, measurements serve to change the likelihoods of various behaviours but never rule them out. This problem has been noted by (Wellman, 1990; Pearl, 1988) and investigated in (Berleant, 1991).

In this chapter we investigate the problem of probabilistic synergy for probability density functions (pdfs) defined over the domain of each observable. The problem is further complicated when a full probabilistic model of the situation is either unavailable or too complex to construct. Motivated by the Kalman filter (KF) which is able to estimate values of states with hardly any information about the actual underlying probabilistic models of noise, we develop the Qualitative Filter (QF) ¹ which uses quantitative, non-parametric statistical filtering techniques on qualitative representations of state. The Dempster-Shafer theory is adopted as the philosophical premise for the QF and a statistical framework for the assimilation of many types of uncertainty is proposed.

¹The use of the term “qualitative” in Qualitative filtering refers to the state representation and not to the representation of degree of belief.
3.2 Rational Decision Making

The philosophical basis underlying the development of the Qualitative filter is the desire of an enquiry to avoid error (Morrell and Stirling, 1991). Contrast this with the goal of classical estimation theory which is to minimise some error metric such as the mean-squared error of the estimate. A decision is made according to some filter theory specific criterion:

Definition 1 Let $D(A\mid\{C_1,\ldots,C_n\})$ be the decision that $A$ is true given that conditions $\{C_1,\ldots,C_n\}$ hold.

Since observations of the system are random we can only guarantee that a decision is correct for a proportion of all possible trials. To quantify the success of the filter, the false-positive error rate $Err$ is useful:

Definition 2 Inclusion Principle. If $x_T$ is the true state of the system, $Q$ is some qualitative state, $Err$ is an upper limit on the false-positive error rate and $D$ is some decision criterion, then the inclusion principle is satisfied when: \(^2\)

$$\Pr(D(x_T \in Q \mid x_T \notin Q)) \leq Err.$$ 

In the following sections we will derive conditions for which the Inclusion Principle applies for all (unknown) probability distributions with specific first and second moments. Ideally, we would like to guarantee a lower bound on the true-positive decision fraction as well. However, knowledge of first and second moments is insufficient to calculate such a bound. Instead, we will insist that the filter has a propensity towards the true-state hypothesis. That is, the filter is unbiased.

3.3 A Framework for Qualitative Model-Based Estimation

The Kalman filter exhibits some desirable properties which we aim to adopt in our Qualitative filtering approach: the Kalman filter maintains unbiased estimates without detailed knowledge of the underlying probability density distribution of the sensor noise or of the stochastic processes of the system. It can combine correlated (i.e. dependent) information and it can optimally reduce uncertainty by selectively combining estimates in accordance to their information content. In the estimation literature these concepts are defined in terms of point estimates.

Unlike the Kalman filter, where parameter point estimates are inter-related using precise quantitative models, no such precision exists within the qualitative approach. Consider the way in which estimates $\hat{x}$ and $\hat{y}$ for two variables $x$ and $y$ are combined using the Kalman filter (see Figure 3.1). A function $f$ (i.e. process or observation model) transforms $\hat{x}$ into Y-space and the point estimates $f(\hat{x})$ and $\hat{y}$ are combined by selective averaging and weighted according to the confidence in the individual estimate. When $f$ is not known precisely we no longer have the requisite point estimates and the best we can say about $f(\hat{x})$ is that it must take some value between landmarks $Y_0$ and $Y_1$.

Before offering our solution to the problem of estimation when models relating parameters are highly imprecise, we will reconsider the Interval Intersection filter (IIF) introduced in Chapter 1. For a noisy observation $z$ of $X$, with standard deviation $\sigma_x$, we compute the interval estimate (Papoulis,

\(^2\)Of course, it is always possible to guarantee this holds by never declaring positive! However, there is no non-parametric bound on the true positive recognition rate.
A similar construct can be obtained for the probability of \( y \) given an observation of \( x \). We then map the interval of \( x \) through to \( Y \) space. If the interval of \( x \) crosses the landmark then both regions \( Y_1 \) and \( Y_2 \) must be covered by the mapped interval. However, if the landmark lies outside the \( x \)-interval then only one region of \( Y \) will be consistent with the observation of \( X \). Thus, by intersecting the interval formed by \( z_y \) with the interval formed by mapping \( z_x \) we are able to fuse the information. Unfortunately, the mapped interval can be excessively large (Gao and Durrant-Whyte, 1994). If it covers both regions in \( Y \) then it offers no extra information about \( y \). Alternatively, if the \( x \)-interval is constrained to one side of the landmark then it, and it alone, asserts a single region in \( Y \) and the \( z_y \) is ineffectual. If both intervals straddle the landmark then no information is gained by fusing the mapped \( x \) interval. Consequently, this approach relies on outliers to aid fusion, a rather risky business, and valuable information about the strength of an observation is not utilised.

In our approach to estimation using qualitative models, we transform our point estimate \( \hat{x} \) into a belief that the point estimate belongs to each region, the region \((X_0, X_1)\) in Figure 3.1 for example. This mass value is then assigned to the region \((Y_0, Y_1)\) according to our imprecise description of \( f \) (that any point in the region \((X_0, X_1)\) maps into \((Y_0, Y_1)\)) and fused with mass estimates generated by \( y \) using the Dempster rule of combination. The degree of uncertainty is captured by a variance measure of the mass function. This approach to filtering is explained in detail in the remainder of this dissertation. We develop a definition of bias in terms of expected likelihood for \( \infty \)-norm estimators: we will use \( \infty \)-norm Dempster-Shafer theory to separate estimate and confidence information; derive basic probability functions for mean and median estimators; derive probabilistic confidence measures in the estimate about basic probability assignments (BPA) ratios; and describe a new measure of information for the Dempster-Shafer reasoning framework (Klir, 1994).

This chapter and the following chapter is dedicated to exploring a framework for recursive unbiased estimation. The investigation proceeds as follows:

1. The Dempster-Shafer theory of evidential reasoning is generalised to the \( p \)-norm theories of evidential reasoning. One particular theory, the \( \infty \)-norm theory, is selected and its conceptual and computational advantages over the other theories are presented.

2. A criterion for unbiased filtering based on mass convergence is presented from which non-parametric basic probability assignment functions for mean and median unbiased estimators are derived.
3. An argument (path) centred approach is advocated and the \( \infty \)-norm theory is shown to remain unbiased after fusion of new estimates and after propagation of mass estimates.

4. Considerations of informationally optimal fusion of non-parametric basic probability assignments using knowledge of the standard deviation of the mass values leads to the definition of the \textit{information filter}. It is shown how covariance information between BPAs can be used to filter noise from dependent information sources.

5. Some problems with this approach are identified and addressed.

### 3.3.1 The Qualitative Filter Philosophy

![Diagram](image)

**Figure 3.2: The Qualitative Filter Cycle.** The state mass estimate is propagated to successor states \( X \). Mass \( m(O_{t+1}) \) assigned by a new observation in quantity space \( O \) is transferred to the state quantity space and fused with \( m(X_{t}) \).

In accordance with most quantitative filters, the Qualitative filter (QF) is a recursive estimator comprising a \textit{prediction stage} for transferring evidence between logically equivalent representations of the domain and an \textit{update stage} for pooling (assimilating) new evidence. States are inferred and assigned evidence during the prediction (inference) stage. Evidence from new observations are transferred using the observation transfer function (i.e., FSM) and pooled with the predicted state evidence in the \textit{update stage}. Figure 3.2 depicts the filter cycle diagrammatically. Using notation first introduced in Section 1.5.2, a single observation is denoted \( z_i \) and a finite sample sequence of \( t + 1 \) observations indexed \( a \) to \( b = a + t \) inclusive is denoted \( z_{i,t}^a \). The sequence \( z_{i,t}^a \) may include observations from different sensors.

Referring to Figure 3.2, two directed graphs \( R_p \) and \( R_o \) take the role of the process and observation models of the Kalman filter respectively. These graphs describe the allowable evolutions of a
3.3 A Framework for Qualitative Model-Based Estimation

system over time and the transformation of observation estimates into state estimates. A directed
graph has an associated set $T \subseteq S \times S$ where $S$ is the set of states. Thus, $(Q_i, Q_j) \in T$ iff there
exists a single-step transition from $Q_i$ to $Q_j$. The set of successor states $R(Q)$ from any state $Q$ is
defined:

$$R(Q) = \{Q' | (Q, Q') \in T\}.$$  

The set of possible future states $R^*$ is defined as the logical closure over the successor states:

$$R^*(Q) = \{Q'' \exists Q' \in R^*(Q) \wedge Q'' \in R(Q')\}.$$  

A basic path in the graph is a sequence of states $Q_0, Q_1, \ldots, Q_{m-1}, Q_m$ such that $\forall i \in 0, \ldots, m - 1. (Q_i, Q_{i+1}) \in T$. A compound path which ends at $Q_m$ is the union of basic paths, each of which end at $Q_m$. A path (denoted $\varphi$) is either basic or compound. The assertion that a path $\varphi$ leads to state $Q$ will be denoted $\varphi \rightarrow Q$. A compound path may be the union of paths drawn from the process model $R_p$ and the observation model $R_o$. The mass assigned to a state $Q$ by the set of $t$ observations $z_1^t$ along path $\varphi$ ending at $Q$ is denoted $m_{\varphi}(Q|z_1^t)$.

Berger (Berger, 1985) (page 275) states that it is “generally hard to carry out extensive probability modelling in complicated situations and many (crude) approximations may be needed”. We agree with this sentiment but disagree with the following comment: that it is “preferable to approximate a correct answer than to adopt a completely ad hoc approach”. The ad hoc approach refers to two methodologies for pooling evidence which are prevalent in the data fusion literature: the linear and the independent opinion pools (Berger, 1985). Evidence $\pi_i$ from a sequence of observations $z_1^t$ is weighted and pooled linearly by the linear opinion pool

$$\pi_1^t = \sum_{i=1}^{t} w_i \pi_i$$

where $w_i > 0$ and $\sum_{i=1}^{t} w_i = 1$. The weights are used to model the reliability of the source and information is combined as if from a disjunctive set of possibly conflicting sources. Methodological approaches for obtaining the weights $w_i$ have been proposed, such as an entropy-based measure (Basir and Shen, 1993). Linear opinion pool methods always average the evidence. This thesis advocates a logarithmic opinion pool paradigm for data fusion (Berger, 1985). This is simply an independent opinion pool paradigm with evidence variably weighted $K_i$ according to its confidence:

$$\pi_1^t = \prod_{i=1}^{t} \pi_i^{K_i}.$$

The prefix “independent” will become redundant and we will show, in Section 3.7, that the logarithmic opinion pool approach can be extended to dependent (correlated) evidence.

The evidence for the compound state $Q$ due to the observation set $z_1^t$ is the weighted sum over the evidence offered by each alternative path leading to proposition $X$ where $X \subseteq Q$. The opinion of each path is reinforced by each observation. This section introduces a general class of estimators with free parameters which assign variable weight to individual paths (i.e., the $p$-weights $p \geq 0$) and to individual observations (i.e., the $K_i$-weights $K_i \geq 0$).

$$\pi_1^t = \left[ \sum_{\varphi} \prod_{i=1}^{t} \left( \pi_i^{K_i} \right)^p \right]^\frac{1}{p}.$$

---

3These graphs are generated from qualitative constraints (i.e., confluences in the QFT literature) by, for example, QSIM (Kuipers, 1994).

4The weights $p$ and $K_i$ must be positive so that the evidence is assigned to the estimate with appropriate polarity.
The quantity-space representation is set-wise but the evidence model is incomplete to interpret $\pi$ as a probability measure. The Dempster-Shafer theory of Evidential Reasoning (Shafer, 1976) (also known as evidence theory or the theory of belief functions) is based on set operations and is intended for use when a complete probabilistic model is unavailable. The theory is intended to maintain upper (plausibility) and lower (belief) bounds on the actual probability and operate using a constructive theory of probability in which probability judgements are made by comparing the particular situation at hand to abstract canonical examples in which the uncertainties are governed by known chances (Shafer, 1981). However, it has been shown that, when renormalisation is required, this approach does not necessarily yield a belief function which assigns accurate probability ranges to each proposition, even if the sources of information are independent (Lemmer, 1986; Pearl, 1986; Pearl, 1990; Voorbraak, 1991; Liu and Bundy, 1994; Murphy, 1996; Wu et al., 1996). So the interpretation of belief as a probability measure must be abandoned.

In this thesis, the interpretation of belief and plausibility functions as bounds on the underlying probability is abandoned and replaced by sufficient criteria for our estimation and hypothesis testing goals. A probability distribution tells us two things: the relative ordering of the probabilities tells us which state to choose; the magnitude of the probability tells us how often our decision is the right one. Thus, the mass function need only capture the preference ordering of states and the corrigibility of this ordering. In this approach, the meaning of the numerical values of the mass function is unimportant. Since the underlying pdf cannot be known a priori the QF must be able to operate with incomplete probability models but sufficient statistics of the noise process. This thesis advocates a non-parametric approach similar to the Kalman Filter and uses only the first and second moments of the noise model. In summary, this thesis will use an extension of the Dempster-Shafer theory of evidential reasoning but:

- Abandon the probabilistic interpretation of Dempster-Shafer mass. A new theory for state preference orderings and confidence in these is required.
- Build non-parametric BPA functions so that proposition preference ordering is correct.
- Use the statistics of the mass function to encode uncertainty in the preference ordering.
- Finally, for decision making, reconstruct statements of probability using the Chebyshev inequality.

Throughout the remainder of the thesis real robot domain data and simulated data will be used to illustrate the abilities of various Qualitative filters. I make no apologies for using simulated data as this provides a direct way to measure the performance of a filtering algorithm under strict control when the true state is known. The next section develops a generalisation of the Dempster-Shafer theory to an infinite class of $p$-norm theories. Conceptually, the $p$-norm generalisation of the Dempster-Shafer theory is no different to the original Dempster-Shafer theory. The $p$-norm formalism generates an infinite class of theories of which the Dempster-Shafer theory is a special case.
3.4 P-Norm Dempster-Shafer Theories of Evidential Reasoning

If \( X \) is a model parameter and \( y \) is a qualitative value then a propositional value in QR is of the form:

- \( Q\text{mag}(X) = y \) meaning that the qualitative magnitude of the parameter \( X \) is the value \( y \).
- \( Q\text{dir}(X) = y \) meaning that the qualitative direction of change of the parameter \( X \) is the value \( y \).

A propositional language for QR expressions is constructed using the following syntax:

1. The alphabet:
   - Logical constants: \( \top \) (truth) and \( \bot \) (falsity).
   - Propositional variables: \( \{a, b, c, \ldots\} \).
   - Classical connectives: \( \neg \) (negation), \( \lor \) (or).

2. The well-formed formulas (wffs):
   - All propositional variables and propositional constants are wffs, also called atomic formulas.
   - If \( a \) and \( b \) are wffs, so are \( \neg a \) and \( a \lor b \).

3. Some abbreviations:
   - \( ab \equiv \neg(a \lor \neg b) \).
   - \( a \supset b \equiv \neg a \lor b \).

Let \( \mathcal{L} \) denote the set of all wffs of the propositional language. For example, in the electronic systems ontology, where voltage \( V \), current \( I \) and resistance \( R \) are related by the familiar expression \( V = IR \) the following is a valid proposition in \( \mathcal{L} \):

\[
[Q\text{dir}(V) = -] \land [Q\text{mag}(R) = +] \supset [Q\text{dir}(I) = -].
\] (3.2)

The frame of discernment \( \Theta \) is a set of mutually exclusive and exhaustive atomic propositions. The frame of discernment is a \( \sigma \)-algebra (or “Borel algebra”) over subsets of \( \Theta \) in that:

1. If \( A \subseteq \Theta \) then its complement \( \neg A \subseteq \Theta \) where \( \neg A \equiv \Theta \setminus A \).
2. If \( \{A_n\} \) is any countable collection of sets in \( \Theta \), then also their union \( \bigcup A_n \) and intersection \( \bigcap A_n \) belong to \( \Theta \).

An expression in \( \mathcal{L} \) may be represented as a subset of \( \Theta \) provided the closed world assumption holds (de Fériet, 1982):

\[
\top \equiv \Theta,
\quad \bot \equiv \emptyset,
\quad a \equiv \{a\}
\quad a \lor b \equiv \{a, b\},
\quad \neg a \equiv \Theta \setminus \{a\}.
\]
So where:

\[ \Theta = \{ \text{Qdir}(V) = -, \text{Qdir}(V) = + \} \times \{ \text{Qmag}(R) = + \} \times \{ \text{Qdir}(I) = -, \text{Qdir}(I) = + \} \]

Expression 3.2 becomes in set-theoretic notation:

\[ \{ \text{Qdir}(V) = +, \text{Qmag}(R) = +, \text{Qdir}(I) = + \} \lor \{ \text{Qdir}(V) = +, \text{Qmag}(R) = +, \text{Qdir}(I) = - \} \lor \{ \text{Qdir}(V) = -, \text{Qmag}(R) = +, \text{Qdir}(I) = + \} \]

An observation assigns to each proposition evidence \( m \) which is a measure over \( \Theta \). The crisp nature of interval arithmetic adopted in this thesis obeys the rules of first order logic. R. T. Cox (Cox, 1946) showed that any theory of uncertainty operating on a first order logic (i.e., an algebra closed under complementarity, union and intersection) and fuses evidence by multiplication (i.e., independent opinion pooling), is an instance of probability theory. In Appendix D we use Cox’s proof to show that a necessary condition for any such theory of uncertainty is:

\[
\left[ \sum_{A \in \Theta} m_p(A)^p \right]^{\frac{1}{p}} = 1
\]

where \( m_p(A) \) is that evidence assigned to \( A \) in theory \( p \). Adding two extra axioms, namely, that inconsistent propositions are not allowed and that all evidence has a positive real value then we obtain the \( p \)-norm theories of evidential reasoning.

**Definition 3** The \( p \)-Norm Theory of Evidential Reasoning:

1. A \( p \)-norm basic probability assignment (BPA) on \( \Theta \) is a function \( m_p : \Theta \rightarrow [0, 1] \) such that:

\[
m_p(\emptyset) = 0 \quad \forall A \subseteq \Theta \quad m_p(A) \geq 0
\]

\[
\left[ \sum_{A \in \Theta} m_p(A)^p \right]^{\frac{1}{p}} = 1
\]

2. The belief function \( \text{Bel}_p \), induced by the BPA \( m_p \) on \( \Theta \) is defined by:

\[
\text{Bel}_p(A) \triangleq \left[ \sum_{B \subseteq A} m_p(B)^p \right]^{\frac{1}{p}}
\]

3. The plausibility function \( \text{Pl}_p \) induced by the BPA \( m_p \) is defined by:

\[
\text{Pl}_p(A) \triangleq \left[ \sum_{B \cap A \neq \emptyset} m_p(B)^p \right]^{\frac{1}{p}}
\]

4. The commonality number \( Q_p(A) \) induced by the BPA \( m_p \) is defined by:

\[
Q_p(A) \triangleq \left[ \sum_{A \subseteq B} m_p(B)^p \right]^{\frac{1}{p}}
\]
3.4 P-Norm Dempster-Shafer Theories of Evidential Reasoning

$m_p(A)$ corresponds to the measure of belief that is committed exactly to the proposition $A$. $Bel_p(A)$ measures the total belief committed to $A$, including measures of belief committed to subsets of $A$. $Pl_p(A)$ indirectly codes information concerning the measures of belief committed to the negation of $A$. $Q_p(A)$ measures the total belief which is committed to $A$ but is not assigned to any proper subset of $A$.

**Definition 4** Let $m_p$ be the BPA of $Bel_p$:

1. If $m_p(A) > 0$ then $A$ is called a focal element of $Bel_p$.
2. The union of all focal elements of $Bel_p$ is called the core of $Bel_p$.
3. A belief function is called vacuous if $\Theta$ is its only focal element.
4. A belief function is called consonant if the core elements are nested: $A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n$.
5. If all focal elements of $Bel$ are singletons, then $Bel_p$ is called Bayesian.

The p-norm theories have the following properties. Let $m_p$ be the BPA of $Bel_p$:

1. $(\forall A \subseteq \Theta) Pl_p(A) \geq Bel_p(A)$.
2. $(\forall A \subseteq \Theta) [Bel_p(A)^p + Pl_p(\neg A)^p]^{\frac{1}{p}} = 1$.
3. $(\forall Q \in \Theta) Pl_p(\{Q\}) = Q_p(\{Q\})$.

**Definition 5** (The P-Norm Dempster Rule of Combination). Let $Bel_p$ and $Bel'_p$ be belief functions induced by the p-norm BPAs $m_p$ and $m'_p$, respectively. The combination of $Bel_p$ and $Bel'_p$ by the p-norm Dempster's rule is the belief function $Bel_p \oplus Bel'_p$ induced by $m_p \oplus m'_p$, where:

$$m_p \oplus m'_p(A) \triangleq K \left[ \sum_{X \subseteq Y} m_p(X)^p m'_p(Y)^p \right]^{\frac{1}{p}}. \quad (3.3)$$

The factor $K$ is the renormalizing constant and is defined so that $\left[ \sum_{A \subseteq \Theta} (m_p \oplus m'_p(A))^p \right]^{\frac{1}{p}} = 1$:

$$K \triangleq \left[ \sum_{A \subseteq \Theta} \sum_{X \subseteq Y, A = X \cap Y} m_p(X)^p m'_p(Y)^p \right]^{\frac{1}{p}}. \quad (3.4)$$

The original Dempster-Shafer theory corresponds to $p = 1$.

Inference, which is a fundamental component of the filtering cycle prediction and observation transfer functions, involves propagating mass through logical links between frames of discernment. Thus, we also introduce the p-norm inference operator $\triangleright$:

**Definition 6** (The p-norm Inference Operator). Let $Bel_p$ be a belief function induced by the p-norm BPA $m_p$ on the frame $\Theta$. The mapping of $Bel$ via the inference graph $R$ is the belief function $\triangleright R Bel_p$ induced by $\triangleright m'_p$, where:

$$\triangleright m'_p(A) \triangleq \left[ \sum_{X \subseteq \Theta, A = X \cap \bigcup_{Y \in X} R(Y)} m_p(X)^p \right]. \quad (3.5)$$
Also, following Shafer (Shafer, 1976) we will consider transformations between frames of
discriminant which discern propositions at various levels of granularity. Let \( \Theta \) and \( \Omega \) be frames of
discriminant. \( \Omega \) is called a refinement of \( \Theta \) and \( \Theta \) is called a coarsening of \( \Omega \). Shafer defines a
refining to be a function \( \omega : 2^\Omega \rightarrow 2^\Theta \) such that for all \( A \subseteq \Theta \), \( \omega(A) \) denotes those propositions in
\( \Omega \) into which \( A \) is split and:

1. \( \omega(A) = \bigcup_{\emptyset \neq A} \omega(\{\theta\}) \).
2. the sets \( \omega(\{\theta\}) \) with \( \theta \in \Theta \) constitute a disjoint partition of \( \Omega \), i.e.:

\[
\omega(\{\theta\}) \neq \emptyset \quad (\forall \theta \in \Theta),
\omega(\{\theta\}) \cap \omega(\{\theta'\}) = \emptyset \quad \text{if } \theta \neq \theta',
\bigcup_{\theta \in \Theta} \omega(\{\theta\}) = \Omega.
\]

An inverse mapping to \( \omega \) is the outer reduction of \( \omega \) and is \( \bar{\omega} : 2^\Theta \rightarrow 2^\Omega \) given by:

\[
\bar{\omega}(A) = \{ \theta \in \Theta \mid \omega(\theta) \cap A \neq \emptyset \}.
\]

### 3.4.1 The \( \infty \)-Norm Theory

Letting \( p \rightarrow \infty \) in Definition 5:

\[
\forall A \subseteq \Theta \quad m_\infty(A) \geq 0, \quad Bel_\infty(A) = \max_{B \subseteq A} m_\infty(B),
\]

\[
m_\infty(\emptyset) = 0, \quad PL_\infty(A) = \max_{B \cap A \neq \emptyset} m_\infty(B),
\]

\[
\max_{A \subseteq \Theta} m_\infty(A) = 1.
\]

with the following rule of combination:

\[
m_{\infty,1} \oplus m_{\infty,2}(A) = K \max_{A = X \cup Y} m_{\infty,1}(X)m_{\infty,2}(Y).
\]

For the remainder of the thesis, the \( \infty \) suffix will be implicitly assumed whenever the norm \( p \) index
is not specified.

In the original \( 1 \)-norm theory, all evidence pointing towards a proposition contributes to the belief
in that proposition whereas, in the \( \infty \)-norm formalism, only the least corrigible evidence contributes.
Effectively, a filter employing this theory performs a multiple hypothesis test (MHT) over all possible
arguments and identifies the proposition which has the single most believable argument.

As we will see, there are a number of advantages to this approach. Firstly, by considering single
paths, we can consider each argument in isolation of the network structure. Thus, we may assign
mass so that the “correct explanation” is preferred on average more than another explanation. Since
distinct arguments never merge and only new branches of sub-arguments are appended to complete
the argument then, irrespective of the structure of the reasoning network, if all sub-arguments
are unbiased then so too should the full argument be unbiased. A second advantage for \( \infty \)-norm
filtering is the computational efficiency of the approach. Combinatorial complexity is always a
problem with the original Dempster-Shafer approach as, for a frame of discernment with \( n \) singleton
propositions, \( 2^n - 1 \) propositions are possible. However, in the next section, we will develop a filtering
framework which utilises only plausibility value ordering of singleton propositions. In anticipation of this development we will show that, under the ∞-norm formalism, both fusion and inference have recursive plausibility expressions.

**Theorem 1** Let Θ₁ and Θ₂ be defined on a frame Θ then:

\[
Pl_{Θ ∪ Θ}(\{Q\}) = \max_{q ∈ Q ∈ R(Θ)} Pl_{Θ}(q),
\]

\[
Pl_{Θ₁ ∪ Θ₂}(\{Q\}) = \frac{Pl_{Θ₁}(\{Q\})Pl_{Θ₂}(\{Q\})}{max_{Q ∈ Θ₁ \cup Θ₂} (Pl_{Θ₁}(Q)Pl_{Θ₂}(Q))}.
\]

**Proof** First we show that theorem holds for propagation and begin by deriving a relationship for the plausibility values between apriori frame Θ and posteriori frame Θ₁Θ:

\[
Pl_{Θ₁Θ}(\{Q\}) = \max_{X ⊆ P ∩ Θ E X} m_{Θ₁Θ}(X)
= \max_{X ⊆ P ∩ Θ E X} \{ \max_{Y ⊆ X ∩ Θ ∪ R(Θ)} m_{Θ}(Y) \}
= \max_{Y ⊆ Θ ∪ R(Θ)} m_{Θ}(Y)
= \max_{q ∈ X ⊆ Θ ∪ R(Θ)} \{ \max_{Y ⊆ X ∩ Θ} m_{Θ}(Y) \}
= \max_{q ∈ X ⊆ Θ ∪ R(Θ)} Pl_{Θ}(q).
\]

The fusion identity relationship can be derived almost directly by Theorem 3.3 in (Shafer, 1976). The p-norm generalisation of this theorem is presented as Theorem 21 in Appendix E. By Theorem 21, and the fact that, for singleton propositions their commonality numbers are identical to their plausibility values then for singleton propositions:

\[
(∀ Q ∈ Θ) Pl_{Θ₁ ∪ Θ₂}(Q) = K Pl_{Θ₁}(Q)Pl_{Θ₂}(Q).
\]

The normalisation constant K satisfies:

\[
K = \max_{A ⊆ C, X ∪ Y = A} m₁(X)m₂(Y)
= \max_{Q ∈ Θ} \{ \max_{A ⊆ C, X ∪ Y = A} \max_{X, Y} m₁(X)m₂(Y) \}
= \max_{Q ∈ Θ} \{ \max_{X, Y} m₁(X)m₂(Y) \}
= \max_{Q ∈ Θ} \{ \max_{X ⊆ C} m₁(X) \} \max_{Y ⊆ C} m₂(Y)
= \max_{Q ∈ Θ} Pl_{Θ₁}(Q)Pl_{Θ₂}(Q).
\]

Refinement and coarsening are instances of propagation with \( R = ω \) for refinement and \( R = ̂ \) for coarsening (see Theorem 22 in Appendix E).

The next section extends the ∞-norm theory to accommodate random statistical evidence and begins by developing a definition of statistical bias for the ∞-norm filter. From this, unbiased non-parametric basic probability assignment functions (BPAs) for mean and median estimation are derived and methods for maintaining the measure of uncertainty in the estimate are introduced.
3.5 Statistical Estimation

In (Shafer, 1976), Shafer argues that the statistical plausibility function for an observable parameter should be equated to the underlying distribution of the observations of this parameter:

\[ Pl(\{\theta\}|z) = f(z|\theta). \]  

(3.6)

There are three reasons why this approach is problematic. Firstly, the path likelihoods are correlated and their ratios (obtained by independent opinion pooling) cannot be interpreted as the relative likelihoods of the paths. Secondly, the pdf may not be known apriori and it may vary significantly between behaviours which are described by a common qualitative model. It is often the case that a statistic such as the mean behaviour is sought. For example, the model of a plane is described in terms of the mean servo-mounted gyro rotation rate for an acoustic tracking sensor. Although the variance and degree of symmetry of the noise may vary between targets of different textures, it is the mean turn rate of the gyro that is used to identify the target type and this statistic is common for all textures. This leads to the third reason to abandon Equation 3.6; that it is undesirable to obtain a complete model of the pdf \( f(z|\theta) \) when only an average statistical feature of \( f \) is sought.

In the light of this discussion, this thesis abandons Equation 3.6 and instead advocate an approach to reasoning based on the relative ordering of elements according to their relative plausibility values. This is in contrast to the standard approach which bases decisions on the absolute numeric values of the mass function. The mass function should be independent of the underlying random behaviour of the observation or process noise but it should reflect the actual behaviour of the physical system. Thus, if the mean of the physical behaviour or of the observed behaviour is sought then we should endeavour to use non-parametric BPAs which are unbiased with respect to the mean for all pdfs.

The remainder of this section demonstrates that long term behaviour of the plausibility function may be inferred with some confidence from observations and knowledge of the statistical properties of the noisy observations. We will now derive a sufficient condition for unbiasedness in terms of the expected value of the BPA \( m(Q|z) \) for a single observation. This will identify constraints on the form of the non-parametric BPA and also identify the limits of the approach. The remainder of this chapter and the next chapter will address the consequences of the following analysis.

3.5.1 Aleatory Data and Statistical Unbiasedness

The sequence of true states \( x_t^1 \) describing the actual physical situation of the system at \( t \) instances are subjected to stochastic evolution and are also observed by noisy sensors. For any ensemble of observations derived from \( x_t^1 \) the filter must remain unbiased: the filter, given sufficient numbers of observations should eventually prefer the qualitative state containing the true state. We have seen that the plausibility values for singleton propositions can be maintained efficiently within the \( \infty \)-norm theory. We will show that rational decisions can be drawn using only the plausibility values for singleton propositions.

Decision Making Using Plausibility Values

The qualitative filter attempts to order the propositional states in accordance with the evidence \( E \) for each. When the evidence is imprecise then \( Bel_p \) and \( Pl_p \) are lower and upper bounds of the
3.5 Statistical Estimation

\[ (\forall A \subseteq \Theta) \text{Bel}_p(A) \leq \mathcal{E}(A) \leq \text{PL}_p(A) \]

However, for singleton elements \( Q \) and \( Q' \), whenever \( \text{Bel}_p(\{Q\}) > \text{PL}_p(\{Q'\}) \) then \( \mathcal{E}(\{Q\}) > \mathcal{E}(\{Q'\}) \) necessarily. Thus, for distinct singleton elements \( Q \) and \( Q' \), we decide \( x \in Q \) when for all \( Q' \neq Q \), the belief of \( Q \) exceeds the plausibility of \( Q' \). For noisy observations the system is unbiased if it has a propensity towards eliminating incorrect decisions as information accrues:

**Definition 7** A belief function \( \text{Bel}_p \) is unbiased if and only if:

\[ (\forall Q, Q', \epsilon > 0)(\exists T > 0)(\forall t > T)(x \in Q \supset \text{Pr}_{z_1}(\text{Bel}_p(Q'|z_1) > \text{PL}_p(Q|z_1)) < \epsilon). \tag{3.7} \]

The following (trivial) theorem forms the basis for a criterion for unbiased filtering:

**Theorem 2** \( \text{Bel}_p \) is an unbiased belief function if:

\[ (\forall Q, Q', \epsilon > 0)(\exists T > 0)(\forall t > T)(x \in Q \supset \text{Pr}_{z_1}(\text{Bel}_p(Q'|z_1) > \text{PL}_p(Q|z_1)) < \epsilon). \]

**Proof** Since \( (\forall p \geq 1)\text{PL}_p(Q|z_1) \geq \text{Bel}_p(Q|z_1) \) then \( [\text{Bel}_p(Q'|z_1) \geq \text{PL}_p(Q|z_1)] \supset [\text{PL}_p(Q'|z_1) \geq \text{PL}_p(Q|z_1)] \) and, therefore:

\[ (\forall Q, Q', \epsilon > 0)(\exists T > 0)(\forall t > T)(x \in Q \supset \text{Pr}_{z_1}(\text{Bel}_p(Q'|z_1) > \text{PL}_p(Q|z_1)) < \epsilon). \]

Thus, the long term behaviour of the plausibility function alone is sufficient to discriminate and eliminate those propositions which do not contain the true state.

Unfortunately, Appendix F informs us that there is no generic BPA for a noisy observable parameter which is unbiased for all qualitative models. The best we can do is to aim to find a method for constructing unbiased BPAs which conforms to a minimal set of criteria:

1. The BPA should be chosen independent of the network structure.
2. The BPA should be unbiased in the limit of zero observation noise.

The following analysis aims to satisfy these criteria.

By the definition of the plausibility function \( \text{PL}_p \), for all \( p \):

\[ \text{PL}_p(X | z_1) = \left[ \sum_{Q' \cap X \neq \emptyset} m_{\varphi}(Q'|z_1)^p \right]^\frac{1}{p} \geq m_{\varphi}(Q|z_1) \]

where \( \varphi \to Q \) and:

\[ \text{PL}_p(X | z_1') = \left[ \sum_{Q' \cap X' \neq \emptyset} m_{\varphi'}(Q'|z_1')^p \right]^\frac{1}{p} = m_{\varphi'}(Q'|z_1') \left[ \sum_{Q' \cap X' \neq \emptyset} \left( \frac{m_{\varphi'}(Q'|z_1')}{m_{\varphi'}(Q'|z_1')} \right)^p \right]^\frac{1}{p} \]
where \( \psi' \to Q' \) and \( \psi'' \to Q' \). Thus:

\[
\log \frac{P_{\psi'}(X | z_i)}{P_{\psi''}(X | z_i)} \geq \log \frac{m_\psi(Q | z_i)}{m_{\psi'}(Q' | z_i)} - \log \left[ \sum_{Q \cap X \neq 0} \left( \frac{m_{\psi'}(Q' | z_i)}{m_{\psi'}(Q | z_i)} \right)^p \right]^{\frac{1}{p}}.
\]

Equality holds when \( X \) is fed by a single path. Consider the fusion of an ensemble \( z_1' \) of observations from \( t \) trials repeated under the same deterministic conditions:

\[
\log \frac{P_{\psi'}(X | z_1')}{P_{\psi''}(X | z_1')} \geq \sum_{i=1}^t \log \frac{m_\psi(Q | z_i)}{m_{\psi'}(Q' | z_i)} - \log \left[ \sum_{Q \cap X \neq 0} \left( \frac{m_{\psi'}(Q' | z_i)}{m_{\psi'}(Q | z_i)} \right)^p \right]^{\frac{1}{p}}.
\]

By the strong law of large numbers (Grimmett and Stirzaker, 1993):

\[
\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^t \log \frac{m_\psi(Q | z_i)}{m_{\psi'}(Q' | z_i)} = a.s. tE \left( \log \frac{m_\psi(Q | z)}{m_{\psi'}(Q' | z)} \right)
\]

and, therefore:

\[
\lim_{t \to \infty} \frac{1}{t} \log \frac{P_{\psi'}(X | z_1')} P_{\psi''}(X | z_1') \geq a.s. tE \left( \log \frac{m_\psi(Q | z)}{m_{\psi'}(Q' | z)} \right) - \Lambda
\]

where:

\[
\Lambda = \log \left[ \sum_{Q \cap X \neq 0} \left( \frac{m_{\psi'}(Q' | z_i)}{m_{\psi'}(Q | z_i)} \right)^p \right]^{\frac{1}{p}}.
\]

The magnitude of \( \Lambda \) is minimised when \( p \to \infty \) in which case:

\[
\Lambda = \log \max_{Q \cap X \neq 0} \left( \frac{m_{\psi'}(Q' | z_i)}{m_{\psi'}(Q | z_i)} \right).
\]

Further, in the limit of zero noise (i.e. \( z_1' = x_1' \)), by choosing \( m_{\psi'}(Q' | x_1') = \max_{\psi'} \{ m_{\psi'}(Q' | x_1') \} \) then \( \Lambda \to 0 \) for any number of paths only when \( p \to \infty \). By Theorem 2 a Qualitative filter is unbiased when:

\[
(\exists X : x \in X) (\forall X' \neq X) \lim_{t \to \infty} \log \frac{P_{\psi'}(X | x)}{P_{\psi''}(X | x)} \geq 0.
\] (3.9)

Theorem 2 is guaranteed to hold if the right hand side of Equation (3.8) is bounded below by zero. We have seen that such a bound exists in a network structure free manner for zero noise when \( p \to \infty \). Therefore the following condition for \( \infty \)-norm BPAs satisfies Criteria 1 and 2:

\[
(\exists Q : x \in Q, X \cap Q \neq \emptyset, \psi \to Q) (\forall Q' \neq Q, \psi' \to Q') E \left( \log \frac{m_\psi(Q | z)}{m_{\psi'}(Q' | z)} \right) \geq 0.
\] (3.10)

Inequality 3.10 is a statement of unbiased filtering based on pair-wise comparisons of path mass values over time. This is the path-centred (or the Lagrangian) view. Alternatively, the original condition for unbiased filtering (Equation 3.9) involves comparisons of BPAs assigned to particular states over time. This will be referred to as the state-centred (or the Eulerian) view. The terms Lagrangian and Eulerian are borrowed from fluid dynamics. In fluid dynamics the Lagrangian view
3.6 The Lagrangian $\infty$-Norm Interpretation

describes the dynamics of a particular element of fluid as it travels down stream. The Eulerian view describes the behaviour of different elements of fluid as they flow past a fixed point relative to the river bank.

Of course, for state filtering purposes the Lagrangian view is an approximation to the Eulerian view. The differences between these views highlights the relative ease with which unbiased BPAs can be assigned by the model builder. The Eulerian interpretation requires consistency over random path assignments to states whereas the Lagrangian interpretation requires consistency of path pairs only. The Eulerian interpretation is the preferred interpretation as it describes the observed statistical behaviour of the mass network. However, as the observation noise decreases the probability that the same path will be assigned to the state increases towards unity and therefore, the Lagrangian interpretation converges towards the Eulerian interpretation as the system noise decreases in strength. This is in contrast to the 1-norm approach for which pair-wise unbiased mass assignments to paths needn’t lead to unbiased filter behaviour even in the limit of zero noise. The convergence of Eulerian and Lagrangian filter behaviours as noise is reduced, the ease with which unbiased estimators can be built and the self-consistent behaviour of the interpretation with the $\infty$-norm theory are reasons for preferring the $\infty$-norm theory over the 1-norm theory.

The use of the word “estimate” in the context of hypothesis testing is now obvious as the aim is to estimate the value of the expected mass between pairs of states from random observations. In the next section we will develop non-parametric BPAs for Lagrangian unbiased filtering. In Section 4.5 we determine cases when the Eulerian decision properties of these BPAs are identical to the Lagrangian decision properties.

3.6 The Lagrangian $\infty$-Norm Interpretation

Reiterating Equation 3.10 from the previous section we have the following the condition for unbiased Lagrangian filtering:

**Definition 8** For the class of $\infty$-norm estimators, if $\psi$ and $\psi'$ are paths which end at $Q$ and $Q'$ respectively then an estimate $m(Q|z_i)$ is unbiased if for observation $i$:

$$\exists Q : x \in Q, \psi \forall Q', \psi' \quad E \left( \log \frac{m_{\psi}(Q|z_i)}{m_{\psi'}(Q'|z_i)} \right) \geq 0$$

We will show that the $\infty$-norm theory of evidential reasoning preserves the Lagrangian unbiased property for inference (which includes refinement and coarsening of frames) and update.

**Theorem 3** The fusion of two Lagrangian unbiased estimates using $\infty$-norm Dempster Rule of Combination is Lagrangian unbiased.

**Proof** Suppose, prior to fusion, we have $x \in Q_1$ and $x \in Q_2$ and:

$$\exists Q_1 : x \in Q_1, \psi_1 \forall Q'_1, \psi'_1 \quad E \left( \log \frac{m_{\psi_1}(Q_1|z_i)}{m_{\psi'_1}(Q'_1|z_i)} \right) \geq 0$$

and:

$$\exists Q_2 : x \in Q_2, \psi_2 \forall Q'_2, \psi'_2 \quad E \left( \log \frac{m_{\psi_2}(Q_2|z_i)}{m_{\psi'_2}(Q'_2|z_i)} \right) \geq 0$$
then we can find a state $Q_3 = Q_1 \cap Q_2$ with $x \in Q_3$ such that, for paths $\varphi_3 = \varphi_1 \cup \varphi_2$ leading to $Q_3$ and $\varphi_3' = \varphi_1' \cup \varphi_2'$ leading to $Q_3'$: 

$$\left( \exists Q_3 : x \in Q_3, \varphi_3 \right) (\forall Q_3', \varphi_3') E \left( \log \frac{m_{\varphi_3}(Q_3|z)}{m_{\varphi_3}(Q_3'|z)} \right) = E \left( \log \frac{m_{\varphi_3}(Q_1|z)}{m_{\varphi_3}(Q_1'|z)} \right) + E \left( \log \frac{m_{\varphi_3}(Q_2|z_{s+1})}{m_{\varphi_3}(Q_2'|z_{s+1})} \right) \geq 0.$$


\[ \square \]

**Theorem 4** The inference from a Lagrangian unbiased estimate using the ∞-norm theory is Lagrangian unbiased.

**Proof** If the initial estimate is unbiased then there is a state $Q$ which contains the true state $x$ and paths $\varphi$ and $\varphi'$ which end at $Q$ and $Q'$ respectively. Therefore:

$$\left( \exists Q : x \in Q, \varphi \right) (\forall Q', \varphi') E \left( \log \frac{m_{\varphi}(Q|z)}{m_{\varphi'}(Q'|z)} \right) \geq 0.$$

Now, since the qualitative representation is sound (Kuipers, 1994) then any future true state $x_f$ must lie in a future successor of $Q$ (i.e. $x_f \in R^*(Q)$) and therefore, for any other state $Q^*$:

$$\left( \exists Q^*: x_f \in Q^*, \varphi^* \subseteq \varphi \cup R^*(Q) \right) (\forall Q'^*, \varphi'^*) E \left( \log \frac{m_{\varphi'^*}(Q^*|z)}{m_{\varphi'^*}(Q'^*|z)} \right) = E \left( \log \frac{m_{\varphi}(Q|z)}{m_{\varphi'}(Q'|z)} \right) \geq 0.$$

\[ \square \]

Thus, we have shown that, if the mass assigned to a state given an observation is unbiased then the mass function after inference and fusion will also be unbiased. We now turn to the problem of finding unbiased mass assignment functions for mean and median estimation.

### 3.6.1 Path Centred (Lagrangian) BPA

Two kinds of filtering can be invoked depending on the stability of the observation noise: mean and median filtering. Unlike the median filter, the mean filter is sensitive to the distance between the observation and the landmark. Thus, the mean filter tends to converge more quickly than the median filter but, the mean filter is prone to misclassification due to outliers. We initially obtain a general unbiased estimator for binary quantity-spaces $\{Q, \neg Q\}$ and then build estimators for $N$ region spaces from this.

For a landmark $l$ separating the decision regions $Q$ and $\neg Q$, the one-sided Laplace distribution is an estimator for the mean of an arbitrary pdf $f(z|x)$ (see Figure 3.3):

**Theorem 5** The mean unbiased BPA is:

$$m_{\text{mean}}(Q|z) \triangleq \begin{cases} 1 \\ \exp(-|z - l|) \end{cases} \quad \iff z \in Q \quad \iff z \in \neg Q. \quad (3.11)$$
3.6 The Lagrangian $\infty$-Norm Interpretation

\[ m_{\text{med}}(Q \mid z) \]
\[ \begin{array}{c}
\text{(a) Laplacian} \\
\text{(b) Step}
\end{array} \]

Figure 3.3: BPA $m(Q \mid z)$ for mean and median estimation.

Proof Define $\neg Q = \{-\infty, l\}$ and $Q = \{l, \infty\}$. For any pdf $f$:

\[
E \left( \log \frac{m_{\text{mean}}(Q \mid z)}{m_{\text{mean}}(\neg Q \mid z)} \right) = \int_{-\infty}^{l} dz f(z \mid x)[-z - l] + \int_{l}^{\infty} dz f(z \mid x)[z - l] = E(z) - l. \tag{3.12}
\]

which clearly satisfies Theorem 8 for $E(z) > l$, $E(z) = l$ and $E(z) < l$.

The step function is the BPA for the median of an arbitrary pdf:

Theorem 6 The median unbiased BPA is:

\[
m_{\text{med}}(Q \mid z) \triangleq \begin{cases} 
1 & \iff z \in Q \\
\alpha & \iff z \in \neg Q
\end{cases} \tag{3.13}
\]

where $\alpha$ satisfies $0 < \alpha < 1$.

Proof We demonstrate that $m_{\text{med}}$ is unbiased with respect to the median of the underlying pdf. The median $z_{\text{med}}$ of a pdf satisfies:

\[
z_{\text{med}} \in Q \iff \int_{l}^{\infty} f(z \mid x)dz < \int_{-\infty}^{l} f(z \mid x)dz
\]

\[
\iff \log \alpha \int_{l}^{\infty} f(z \mid x)dz > \log \alpha \int_{-\infty}^{l} f(z \mid x)dz, \quad \text{since } \alpha < 1
\]

\[
\iff \int_{-\infty}^{l} 0f(z \mid x)dz + \log \alpha \int_{l}^{\infty} f(z \mid x)dz > \int_{l}^{\infty} 0f(z \mid x)dz + \log \alpha \int_{-\infty}^{l} f(z \mid x)dz
\]

\[
\iff E(\log m_{\text{med}}(Q \mid z)) > E(\log m_{\text{med}}(\neg Q \mid z)).
\]

\[\square\]
Although these BPAs are unbiased with respect to the true state provided the true state is at the mean or median respectively, it is possible to adjust the BPA accordingly for any fixed bias $\alpha$ (i.e. $l \rightarrow l + \alpha$).

### 3.6.2 N-ary State BPAs

We now turn our attention to statistical fusion for problems involving more than two states: the $N$ state problem.

![Diagram of binary frames corresponding to four region refined frame $\theta$.](image)

**Figure 3.4:** Three binary frames of discernment $\Upsilon$ corresponding to the four region refined frame $\Theta$.

To extend our estimation framework from binary quantity-spaces to $N$ region spaces we show that filtering in a quantity-space comprising $N-1$ landmarks can be formulated as the conjunction of $N-1$ binary decision problems, each comprising a single landmark. We use the notion of coarse frames of discernment from the Dempster-Shafer Theory of Evidence (Shafer, 1976) described in the Section 3.4.1. A coarse frame of discernment $\Upsilon$ comprises propositions which are supersets of propositions in $\Theta$ (in this case called a refined frame). For each member $v \in \Upsilon$ the refining operator $\omega_\Upsilon : 2^\Upsilon \rightarrow 2^\Theta$ gives the subset $\omega_\Upsilon(v)$ of $\Theta$ consisting of those possibilities into which $v$ can be split:

$$x \in \Theta \text{ and } x \in v \Rightarrow x \subseteq \omega_\Upsilon(v).$$

Each of the $N-1$ coarse frame mass assignments $m_{\Upsilon}(Q_\Upsilon|z)$ obtained from an observation using Equation (3.11) or (3.13) can be seen as a source of information pertaining to the true state of the system and the true state must have common consensus for all sources. Therefore, the likelihood that the true state $x$ is in $\Theta$ is represented by the conjunction of a subset of the $N-1$ coarse frame mass distributions. To illustrate, the three coarse frames in Figure 3.4 denoted $A$, $B$, and $C$ are coarsenings of the underlying frame of discernment $\Theta = \{Q_0, Q_1, Q_2, Q_3\}$ such that:

- $\omega_A(\{A\}) = \{Q_0\}$,
- $\omega_A(\{\neg A\}) = \{Q_1, Q_2, Q_3\}$,
- $\omega_B(\{B\}) = \{Q_0, Q_1\}$,
- $\omega_B(\{\neg B\}) = \{Q_2, Q_3\}$,
- $\omega_C(\{C\}) = \{Q_0, Q_1, Q_2\}$,
- $\omega_C(\{\neg C\}) = \{Q_3\}$.
3.6 The Lagrangian \(\infty\)-Norm Interpretation

We treat each coarse frame as an individual expert basic probability mass assignment and combine the (highly correlated) collection of BPAs using the \(\infty\)-norm Dempster rule. We will prove the correctness of this procedure shortly. From each individual coarse mass function we assign mass to the refined frame \(\Theta\) according to (Shafer, 1976):

\[
m_\Theta(\omega_{\Theta}(v)) = m_{\Theta}(v).
\]

Then, \(N\) individual BPAs are combined:

\[
m_\Theta(\theta) = k \max_{\Theta=H_1 \cap \ldots \cap H_N} \prod_{Y=1}^{N} m_{\Theta}(H_Y).
\]

where \(\Theta \subseteq \omega_{\Theta}(H_Y)\) and \(k\) is a normalisation factor. To illustrate, \(\{Q_0\} = \omega_{A}(\{A\}) \cap \omega_{A}(\{B\}) \cap \omega_{A}(\{C\})\) so \(m_\Theta(\{Q_0\}) = k m_A(\{A\}) m_B(\{B\}) m_C(\{C\})\). The complete refined BPA for the example described above is:

\[
m_\Theta(\{Q_0\}) = k m_A(\{A\}) m_B(\{B\}) m_C(\{C\}),
\]

\[
m_\Theta(\{Q_1\}) = k m_A(\{\neg A\}) m_B(\{B\}) m_C(\{C\}),
\]

\[
m_\Theta(\{Q_2\}) = k m_A(\{\neg A\}) m_B(\{\neg B\}) m_C(\{C\}),
\]

\[
m_\Theta(\{Q_3\}) = k m_A(\{\neg A\}) m_B(\{\neg B\}) m_C(\{\neg C\}).
\]

Normalising the mass assigned to the central state \(m_\Theta(\{Q_1\})\), say, we obtain:

\[
\frac{m_\Theta(\{Q_0\})}{m_\Theta(\{Q_1\})} = \frac{m_A(\{A\})}{m_A(\{\neg A\})}, \quad \frac{m_\Theta(\{Q_2\})}{m_\Theta(\{Q_1\})} = \frac{m_B(\{\neg B\})}{m_B(\{B\})}
\]

and:

\[
\frac{m_\Theta(\{Q_3\})}{m_\Theta(\{Q_1\})} = \frac{m_B(\{\neg B\}) m_C(\{\neg C\})}{m_B(\{B\}) m_C(\{C\})}.
\]

In general, we have:

**Theorem 7** For singleton states \(Q\) and \(Q'\) in a refined frame \(\Theta\) and a set \(L\) of the \(N-1\) compatible binary frames of discernment:

\[
\log \frac{m_\Theta(Q|x)}{m_\Theta(Q'|x)} = \sum_{Y \in L} \log \frac{m_{\Theta}(\{Q\}|x)}{m_{\Theta}(\{Q'|x\})}
\]

is an unbiased BPA.

**Proof** The expected contribution to \(\log \frac{m_\Theta(Q|x)}{m_\Theta(Q'|x)}\) for a binary decision problem \(\gamma = \{A_T, \neg A_T\}\) with \(x \in Q\) is:

\[
E \left( \log \frac{m_{\gamma}(A_T|x)}{m_{\gamma}(A_T'|x)} \right)
\]

where:

\[
A_T = \begin{cases} 
A_T & \text{if } Q' \in \omega_{\gamma}(A_T) \text{ in which case } E \left( \log \frac{m_{\gamma}(A_T|x)}{m_{\gamma}(A_T'|x)} \right) = 0 \\
\neg A_T & \text{if } Q' \not\in \omega_{\gamma}(A_T) \text{ in which case } E \left( \log \frac{m_{\gamma}(A_T|x)}{m_{\gamma}(A_T'|x)} \right) > 0.
\end{cases}
\]

\(\square\)
The quality (or confidence) of an estimate is not encoded in non-parametric BPAs and it is necessary to reason about the higher order statistics of the BPAs so that rational decisions can be made. The next section demonstrates how a variance measure over the path centred mass can be maintained within the QF framework.

### 3.7 An Information Centred Approach

In this section confidence intervals are considered and the Decision Rule is derived. The Decision Rule uses the mass value and statistics of the mass value for an ensemble of trials for which the true state is confined to $x_1$ to guarantee a maximum false-positive identification rate. To improve readability we introduce some notation for the log-mass ratio and for its variance:

$$M_t^1 \equiv \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])}, \quad V_t^1 \equiv \text{Var}_{x_1} \left( \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])} \right),$$

$$M_t^{1-1} \equiv \log \frac{m_{\psi}(Q[z_1^{-1}])}{m_{\psi'}(Q'[z_1^{-1}])}, \quad V_t^{1-1} \equiv \text{Var}_{x_1^{-1}} \left( \log \frac{m_{\psi}(Q[z_1^{-1}])}{m_{\psi'}(Q'[z_1^{-1}])} \right),$$

$$M_t \equiv \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])}, \quad V_t \equiv \text{Var}_{x_1} \left( \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])} \right),$$

and:

$$C \equiv \text{Cov}_{x_1,x_2} \left( \log \frac{m_{\psi}(Q[z_1^{-1}])}{m_{\psi'}(Q'[z_1^{-1}])}, \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])} \right).$$

Confidence measures in the log likelihood ratio are required to indicate when a rational decision can be made. We decide $x \in Q$ for a finite set of observations whenever:

$$(\exists Q : x \in Q, \forall Q', y') \ E(M_t^1) \equiv E_{x_1} \left( \log \frac{m_{\psi}(Q[z_1])}{m_{\psi'}(Q'[z_1])} \right) \geq 0, \quad (3.20)$$

In general we won’t know $E(M_t^1)$ but we can estimate its value with some probability of confidence using $M_t^1$ and $V_t^1$.

**Theorem 8** For two states $Q$ and $Q'$, if we make the decision $x_T \in Q$ when:

$$M_t^1 \geq G\sqrt{V_t^1} \quad (3.21)$$

where $G$ is a value referred to as a “gate”, then the probability that an incorrect decision is made across the entire ensemble of trials, is less than $G^{-2}$. That is:

$$Pr(M_t^1 \geq G\sqrt{V_t^1} \mid x_T \in Q') \leq \frac{1}{G^2}$$

**Proof** By Chebyshev’s Inequality (Grimett and Stirzaker, 1993):

$$Pr \left( \left| M_t^1 - E_{x_1} (M_t^1) \right| \geq G\sqrt{V_t^1} \right) \leq \frac{E \left( (M_t^1 - E_{x_1} (M_t^1))^2 \right)}{(G\sqrt{V_t^1})^2} = \frac{1}{G^2}. \quad (3.22)$$
Now, since \( [x_T \in Q'] \supset [E(M^j) < 0] \) then:
\[
Pr \left( M^j \geq G \sqrt{V_T} \mid x_T \in Q' \right) \leq Pr \left( M^j \geq G \sqrt{V_T} \mid E(M^j) \leq 0 \right) \\
\leq Pr \left( |M^j - E(M^j)| \geq G \sqrt{V_T} \right) \\
\leq \frac{1}{G^2}.
\]

The last line follows by Equation (3.22).

\[\square\]

When more information is known about the distribution (for example that it is as good as symmetric about its mean or that the distribution is Gaussian) then a tighter estimate for the error rate may be found. The case of independent symmetric distributions is investigated in Section 4.2.1.

The uncertainty between pairs of states can be captured by a single scalar value: the Biscay information measure.

**Definition 9** The Biscay information \( I^j_1 \) for two states and observation sequence \( z^j_1 \) is:
\[
I^j_1 \triangleq \frac{M^j}{\sqrt{V^j_1}}.
\]

The intuition behind Biscay is that the more information available, the more confident the decision. We shall use Biscay information to guide the construction of efficient estimators which are characterised by their desire to maximise the information value.

In the remainder of this section we describe how to obtain the variance of the log mass ratio for two states \( Q \) and \( Q' \) and an observation sequence \( z^j_1 \) recursively using covariance measures on subsets of \( z^j_1 \). We start by calculating the log mass ratio covariance for two observations \( z_1 \) and \( z_2 \) for a binary decision problem \( \{Q, \neg Q\} \). For a mean estimator the covariance of the log mass ratio is simply the covariance of the observations themselves:

**Theorem 9**
\[
\text{Cov}_{z_1, z_2} \left( \log \frac{m_{\text{mean}}(Q | z_1)}{m_{\text{mean}}(\neg Q | z_1)}, \log \frac{m_{\text{mean}}(Q | z_2)}{m_{\text{mean}}(\neg Q | z_2)} \right) = \text{Cov}_{z_1, z_2}(z_1, z_2).
\]

**Proof** Using Equation (3.12) we obtain:
\[
\text{Cov}_{z_1, z_2} \left( \log \frac{m_{\text{mean}}(Q | z_1)}{m_{\text{mean}}(\neg Q | z_1)}, \log \frac{m_{\text{mean}}(Q | z_2)}{m_{\text{mean}}(\neg Q | z_2)} \right) = \text{Cov}_{z_1, z_2}(z_1 - I, z_2 - I) = \text{Cov}_{z_1, z_2}(z_1, z_2).
\]

\[\square\]

The median estimator is essentially a probabilistic estimator and the covariance can be determined if the probabilistic distribution of observations is known. For our median estimator we have:
3.7 An Information Centred Approach

Theorem 10

\[
\text{Cov}_{x_1, x_2} \left( \log \frac{\text{med}(Q|x_1)}{\text{med}(\neg Q|x_1)}, \log \frac{\text{med}(Q|x_2)}{\text{med}(\neg Q|x_2)} \right) = 2(\log \alpha)^2 (\Pr(z_1 \in Q, z_2 \in Q|x_1, x_2) + \Pr(z_1 \in \neg Q, z_2 \in \neg Q|x_1, x_2) - 1 - 2\Pr(z_1 \in Q|x_1)\Pr(z_2 \in Q|x_2) + \Pr(z_1 \in \neg Q|x_1) + \Pr(z_2 \in \neg Q|x_2)).
\]

(3.23)

Proof

\[
\text{Cov}_{x_1, x_2} \left( \log \frac{\text{med}(Q|x_1)}{\text{med}(\neg Q|x_1)}, \log \frac{\text{med}(Q|x_2)}{\text{med}(\neg Q|x_2)} \right) = \int_\mathbb{R} dz_1 \int_\mathbb{R} dz_2 f(z_1, z_2|x_1, x_2) \left( \log \frac{\text{med}(Q|x_1)}{\text{med}(\neg Q|x_1)} \log \frac{\text{med}(Q|x_2)}{\text{med}(\neg Q|x_2)} \right) - \left( \int_\mathbb{R} dz_1 f(z_1|x_1) \log \frac{\text{med}(Q|x_1)}{\text{med}(\neg Q|x_1)} \right) \left( \int_\mathbb{R} dz_2 f(z_2|x_2) \log \frac{\text{med}(Q|x_2)}{\text{med}(\neg Q|x_2)} \right)
\]

\[
= (\log \alpha)^2 (\Pr(z_1 \in Q, z_2 \in Q|x_1, x_2) - \Pr(z_1 \in Q, z_2 \in \neg Q|x_1, x_2) - \Pr(z_1 \in \neg Q, z_2 \in Q|x_1, x_2) + \Pr(z_1 \in \neg Q, z_2 \in \neg Q|x_1, x_2) - [\Pr(z_1 \in Q|x_1) - \Pr(z_1 \in \neg Q|x_1)] [\Pr(z_2 \in Q|x_2) - \Pr(z_2 \in \neg Q|x_2)])
\]

\[
= (\log \alpha)^2 (2\Pr(z_1 \in Q, z_2 \in Q|x_1, x_2) + 2\Pr(z_1 \in \neg Q, z_2 \in \neg Q|x_1, x_2) - 1 - 2\Pr(z_1 \in Q|x_1) - 2\Pr(z_2 \in Q|x_2) - 1)
\]

\[
= 2(\log \alpha)^2 (\Pr(z_1 \in Q, z_2 \in Q|x_1, x_2) + \Pr(z_1 \in \neg Q, z_2 \in \neg Q|x_1, x_2) - 1 - 2\Pr(z_1 \in Q|x_1)\Pr(z_2 \in Q|x_2) + \Pr(z_1 \in \neg Q|x_1) + \Pr(z_2 \in \neg Q|x_2)).
\]

\[
\text{Var}_x \left( \log \frac{\text{med}(Q|x)}{\text{med}(\neg Q|x)} \right) = 4(\log \alpha)^2 \Pr(z \in Q|x)(1 - \Pr(z \in Q|x)).
\]

(3.24)

By setting \(z_2 = z_1\) and \(x_2 = x_1\) we can obtain an expression for the variance of the median filter:

\[
\text{Var}_x \left( \log \frac{\text{med}(Q|x)}{\text{med}(\neg Q|x)} \right) = (\log \alpha)^2.
\]

When \(\Pr(z \in Q|x)\) is unconstrained then:

\[
\text{Var}_x \left( \log \frac{\text{med}(Q|x)}{\text{med}(\neg Q|x)} \right) \leq (\log \alpha)^2.
\]

and, since in general for arbitrary random variables \(A\) and \(B\), \(\text{Cov}(A, B) \leq \frac{\text{Var}(A) + \text{Var}(B)}{2}\) then:

\[
\text{Cov}_{x_1, x_2} \left( \log \frac{\text{med}(Q|x_1)}{\text{med}(\neg Q|x_1)}, \log \frac{\text{med}(Q|x_2)}{\text{med}(\neg Q|x_2)} \right) \leq (\log \alpha)^2.
\]

The variance of the log likelihood ratio in the refined frame \(\Theta\) can be obtained using Equation (3.19). For two regions \(Q_i\) and \(Q_{i+n}\) in the refined frame \(\Theta\) (see Figure 3.4) we have:

\[
\text{Cov}_{x_1, x_2} \left( \log \frac{\text{med}(Q_i|x_1)}{\text{med}(\neg Q_i|x_1)}, \log \frac{\text{med}(Q_{i+n}|x_2)}{\text{med}(\neg Q_{i+n}|x_2)} \right) = |a|^2 \text{Cov}_{x_1, x_2} \left( \log \frac{\text{med}(Q_i|x_1)}{\text{med}(\neg Q_i|x_1)}, \log \frac{\text{med}(Q_{i+n}|x_2)}{\text{med}(\neg Q_{i+n}|x_2)}, (3.25)
\]

\[
\text{Var}_{x_1, x_2} \left( \log \frac{\text{med}(Q_i|x_1)}{\text{med}(\neg Q_i|x_1)}, \log \frac{\text{med}(Q_{i+n}|x_2)}{\text{med}(\neg Q_{i+n}|x_2)} \right) = (|a|^2)^2 \text{Var}_{x_1, x_2} \left( \log \frac{\text{med}(Q_i|x_1)}{\text{med}(\neg Q_i|x_1)}, \log \frac{\text{med}(Q_{i+n}|x_2)}{\text{med}(\neg Q_{i+n}|x_2)} \right).
\]
Using the covariance values we may obtain a value for the Biscay information using Equation (3.1):
\[ M_i^t = \sum_{i=1}^{t} K_i M_i \]  
(3.26)
and the variance of \( M_i^t \):
\[ V_i^t = \sum_{i=1}^{t} \sum_{j=1}^{t} K_i K_j C_{i,j} \]  
(3.27)
where \( C_{i,j} = \text{Cov}(M_i, M_j) \) or, we may obtain a recursive expression for Biscay information in terms of mass ratios and their covariance values:
\[ M_i^t = K_i^{t-1} M_i^{t-1} + K_i M_t \]  
(3.28)
and the variance of \( M_i^t \):
\[ V_i^t = (K_i^{t-1})^2 (V_i^{t-1})^2 + (K_i)^2 (V_i)^2 + 2 K_i^{t-1} K_i C. \]  
(3.29)
where \( C = \text{Cov}(M_i^{t-1}, M_t) \). The \( \infty \)-norm theory of evidential reasoning and the Lagrangian approximation to Eulerian unbiased filtering allows a straight-forward linear, recursive, non-parametric estimator which requires only information about the first and second moments of the noise distribution.

Our filter parameters, \( M \) and \( V \), describe the statistics of mass ratios between states on two paths. Thus, for \( n \) active paths \( 2n^2 \) statistics are required. We could find equivalent expressions for each path individually for which only \( 2n \) statistics will be required. We could, for example, assign \( m(Q_0) = 1 \) for some state \( Q_0 \) and then use \( \log m(Q_i) = \log \frac{m(Q_i)}{m(Q_0)} \) for all other states \( Q_i \). In this case \( \text{Var}(\log Q_0) = 0 \) and \( \text{Var}(\log m(Q_i)) = \text{Var} \left( \log \frac{m(Q_i)}{m(Q_0)} \right). \) However, ultimately we will want to construct a path preference ordering using state pair mass differences and the variance of these values. The masses assigned to different paths by a single observation are highly correlated (see, for example, Equations 3.15 to 3.18) and an efficient estimate would require information of the path mass correlations.

**Example: Correlated Observations**

We wish to compare two states \( \{Q_1, Q_1, \text{inc}\} \) and \( \{Q_2, Q_2, \text{dec}\} \) given two observations \( z_1 \) and \( z_2 \). We construct the joint mass assignment thus:
\[ \log \frac{m(\{Q_1, Q_1, \text{inc}\}|z_1, z_2)}{m(\{Q_2, Q_2, \text{dec}\}|z_1, z_2)} = \log \frac{m(Q_1|z_1)}{m(Q_2|z_1)} + \log \frac{m(Q_1|z_2)}{m(Q_2|z_2)} + \log \frac{m(\text{inc}|z_1, z_2)}{m(\text{dec}|z_1, z_2)} \]
\[ = (z_1 - z_1) + (z_2 - z_1) + (z_2 - z_1). \]

Thus, the joint variance is \( 4 \text{Var}_{z_2}(z_2) \).

### 3.8 Stochastic Processes: The Stochastic QF (SQF)

To introduce the problem of encoding stochastic processes within the QF framework consider the following problem. Suppose a stationary stochastic system operates on a binary quantity-space

\(^5\)Whenever \( M_i = 0 \) then we may assign \( K_i = 0 \) without affecting the statistical properties of the filter.
\( \{ Q, Q' \} \). The deterministic component of the model allows state transitions \( Q \rightarrow Q \) and \( Q' \rightarrow Q' \). However, suppose that there is a finite probability that the true state can drift from \( Q \) into \( Q' \) or from \( Q' \) to \( Q \) due to random buffeting. We make all transitions to all possible states explicit in the qualitative model transition graph and augment these transitions with transition probabilities when available. This approach assumes that, for any state \( q \) which can propagate to a subset of states \( Q_0, \ldots, Q_n \), the stochastic process transition probabilities \( \Pr(Q_i \mid q) \) are (partially) known (see Figure 3.5). This approach uses ideas from probabilistic network theory (Pearl, 1988; Neapolitan, 1990) which have been investigated in the context of QR stochastic models by (Doyle and Sados, 1989; Bratnik, 1997). Strictly, the QF is unbiased when the expected Biscay-information-induced preference ordering of states prefers the state containing \( x_T \). Thus, an unbiased stochastic process filter cannot be defined in terms of the expected value of some BPA over all trials. Instead, we define the stochastic QF estimator \( m_{\text{stoch}}(Q \mid x_T) \) such that \( m_{\text{stoch}}(Q \mid x_T) > m_{\text{stoch}}(Q' \mid x_T) \iff x_T \in Q \). The stochastic QF estimator is similar to the probabilistic driven median filter:

\[
m_{\text{stoch}}(Q \mid x_T) = \begin{cases} 
1 & \iff x_T \in Q \\
\alpha & \iff x_T \in Q'
\end{cases}
\]

where \( \alpha < 1 \). Define the log mass ratio:

\[
M_{Q,Q'}(x_T) = \log \frac{m_{\text{stoch}}(Q \mid x_T)}{m_{\text{stoch}}(Q' \mid x_T)}.
\]

then, for two paths \( q \rightarrow Q \) and \( q \rightarrow Q' \):

\[
E(M_{Q,Q'}(x_T)) = -\log \alpha \left[ \int_Q f(x_T \mid q) dx_T - \int_{Q'} f(x_T \mid q) dx_T \right]
\]

\[
= -\log \alpha [\Pr(Q \mid q) - \Pr(Q' \mid q)]
\]

and:

\[
\text{Var}(M_{Q,Q'}(x_T)) = \Pr(Q \mid q) \left( \log \frac{m_{\text{stoch}}(Q \mid x_T)}{m_{\text{stoch}}(Q' \mid x_T)} \right)^2 + \Pr(Q' \mid q) \left( \log \frac{m_{\text{stoch}}(Q \mid x_T)}{m_{\text{stoch}}(Q' \mid x_T)} \right)^2
\]

\[
- \left( \Pr(Q \mid q) \left( \log \frac{m_{\text{stoch}}(Q \mid x_T)}{m_{\text{stoch}}(Q' \mid x_T)} \right) + \Pr(Q' \mid q) \left( \log \frac{m_{\text{stoch}}(Q \mid x_T)}{m_{\text{stoch}}(Q' \mid x_T)} \right) \right)^2
\]

\[
= (\log \alpha)^2 [\Pr(Q \mid q) + \Pr(Q' \mid q) - (\Pr(Q \mid q) - \Pr(Q' \mid q))^2].
\]
Thus,

\[
E(P_{Q,Q'}^{stoch}(x_T)) = \frac{E(M_{Q,Q'}(x_T))}{\sqrt{\text{Var}(M_{Q,Q'}(x_T))}} = \frac{\Pr(Q \mid q) - \Pr(Q' \mid q)}{\sqrt{\Pr(Q \mid q) + \Pr(Q' \mid q) - (\Pr(Q \mid q) - \Pr(Q' \mid q))^2}}.
\]

The expected stochastic Biscay information is an estimate of the trial specific stochastic Biscay information \( P_{Q,Q'}^{stoch}(x_T) \). To prove this, we will show that the false-positive error rate of the SQF with a decision value \( G \) is bounded by \( \frac{1}{G^2} \).

The Biscay information between two states \( Q \) and \( Q' \) will comprise observation estimates \( P_{Q,Q'}^{obs}(x_i) \) and expected stochastic estimates \( E(P_{Q,Q'}^{stoch}(x_i)) \). The true Biscay information is the Biscay information when the true state is known:

\[
I_{Q,Q'}^{true} = K^{-1} \left( \sum_i [E(P_{Q,Q'}^{obs}(x_i)) + P_{Q,Q'}^{stoch}(x_i)] \right).
\]

where \( K \) is a normalisation constant:

\[
K^2 = \text{Var} \left( \sum_k P_{Q,Q'}^{obs}(z_k) \right) + \text{Var} \left( \sum_k P_{Q,Q'}^{stoch}(x_k) \right).
\]

This choice of \( K \) will be justified shortly. Since \( E(P_{Q,Q'}^{obs}) \) and \( P_{Q,Q'}^{stoch} \) are unbiased the true Biscay information induces an appropriate preference ordering over the states \( Q \) and \( Q' \), so that \( I_{Q,Q'}^{true} > 0 \iff x_T \in Q \). The combined observation and stochastic process Biscay information \( I_{Q,Q'} \) is:

\[
I_{Q,Q'} = K^{-1} \left( \sum_i [P_{Q,Q'}^{obs}(x_i) + E(P_{Q,Q'}^{stoch}(x_i))] \right).
\] (3.30)

Assuming that observation noise is independent of stochastic noise, \( I_{Q,Q'} \) is an estimate of \( I_{Q,Q'}^{true} \). To see this the probability of false-positive error is:

\[
\Pr(I_{Q,Q'} > G \mid x \in Q) \\
\leq \Pr \left( K^{-1} \sum_k \left( P_{Q,Q'}^{obs}(z_k) + E(P_{Q,Q'}^{stoch}(x_k)) \right) > G \mid K^{-1} \sum_k \left( E(P_{Q,Q'}^{obs}(z_k)) + P_{Q,Q'}^{stoch}(x_k) \right) < 0 \right) \\
\leq \Pr \left( K^{-1} \sum_k \left[ \left( P_{Q,Q'}^{obs}(z_k) - E(P_{Q,Q'}^{obs}(z_k)) \right) - \left[ E(P_{Q,Q'}^{stoch}(x_k)) - E(I_{Q,Q'}^{stoch}(x_k)) \right] \right] < G \right) \\
\leq \frac{\text{Var} \left( \sum_k \left[ \left( P_{Q,Q'}^{obs}(z_k) - E(P_{Q,Q'}^{obs}(z_k)) \right) - \left[ E(P_{Q,Q'}^{stoch}(x_k)) - E(I_{Q,Q'}^{stoch}(x_k)) \right] \right] \right)}{(KG)^2} \\
\leq \frac{\text{Var} \left( \sum_k P_{Q,Q'}^{obs}(z_k) \right) + \text{Var} \left( \sum_k P_{Q,Q'}^{stoch}(x_k) \right)}{(KG)^2} \\
= \frac{1}{G^2}
\]

The penultimate line follows by the Chebyshev inequality. Thus, Equation 3.30 is a decision conservative stochastic estimator.
Example: Simple Stochastic Process

When probabilistic transition information is available the SQF can be both more accurate and faster than the QF, but not necessarily for all problems. To illustrate this, a very simple stochastic system is simulated over 1000 trials and the statistical properties of the SQF are compared with those of the QF. The system comprises a true state \( x \in \mathcal{Q} \) at time \( k = 0 \) transitioning to one of two states \( Q_0 \) or \( Q_1 \) in proportions \( p_0 \) and \( p_1 \) respectively at time \( k = 1 \). States \( Q_0 \) and \( Q_1 \) are separated by a landmark \( l = 0 \) and when in \( Q_0, x_T = 1.0 \) and when in \( Q_1, x_T = -1.0 \). The observation noise is Gaussian with a standard error of 10.0 units. The system is observed at time \( k = 1 \) by a mean unbiased noisy sensor. After 1000 trials for each \( p_0 \in \{0.5, 0.6, 0.7, 0.8, 0.9\} \) the following statistics were obtained (both the QF and the SQF use the same data):

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
Pr(\mathcal{Q}_0) & Pr_{QF}(e \mid \mathcal{Q}_0) & Pr_{QF}(e \mid \mathcal{Q}_1) & Pr_{QF}(\text{error}) & Pr_{SQF}(e \mid \mathcal{Q}_0) & Pr_{SQF}(e \mid \mathcal{Q}_1) & Pr_{SQF}(\text{error}) \\
\hline
0.5 & 7.94 & 4.91 & 6.43 & 1.43 & 1.28 & 1.40 \\
0.6 & 6.29 & 7.11 & 6.62 & 0.97 & 2.89 & 1.74 \\
0.7 & 6.43 & 6.32 & 6.40 & 0.56 & 5.61 & 2.08 \\
0.8 & 6.32 & 5.26 & 6.11 & 0.13 & 7.18 & 1.54 \\
0.9 & 7.05 & 7.48 & 7.10 & 0.00 & 20.56 & 2.06 \\
\hline
\end{array}
\]

Table 3.1: Proportion of false-positive error rates for the QF and the SQF.

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
Pr(\mathcal{Q}_0) & Pr_{QF}(e \mid \mathcal{Q}_0) & Pr_{QF}(e \mid \mathcal{Q}_1) & Pr_{QF}(\text{error}) & Pr_{SQF}(e \mid \mathcal{Q}_0) & Pr_{SQF}(e \mid \mathcal{Q}_1) & Pr_{SQF}(\text{error}) \\
\hline
0.5 & 7.13 & 9.04 & 8.08 & 1.63 & 3.54 & 2.58 \\
0.6 & 9.03 & 10.00 & 9.42 & 4.68 & 0.53 & 3.02 \\
0.7 & 10.07 & 10.53 & 10.21 & 8.11 & 1.75 & 6.20 \\
0.8 & 9.61 & 9.57 & 9.60 & 12.39 & 0.48 & 10.01 \\
0.9 & 8.96 & 6.54 & 8.72 & 27.32 & 0.00 & 24.59 \\
\hline
\end{array}
\]

Table 3.2: Proportion of correct recognition rates for the QF and the SQF.

As is evident from Table 3.1, the total error rate of the SQF is consistently about 20% that of the QF. The relative correct estimation rates are shown in Table 3.2. It is perhaps interesting to note that the SQF is superior for extreme transition probability differences \( P(\mathcal{Q}_0) > 0.75 \) but the QF is superior for less transition probability values. The reason for this is that the QF structure “averages” the input Biscay information: \( I = K(I_{obs} + E(I_{stoch})) \) where \( K \ll 1 \). When the stochastic information contribution \( E(I_{stoch}) \) is small then the normalisation term \( K \ll 1 \) can reduce the potency of \( I_{obs} \). To quantify this further, the cross-over of QF/SQF potency occurs for independent input information when the Biscay information is the same for the QF and the SQF:

\[
I_{obs} = I_{obs} + E(I_{stoch}) \sqrt{2}.
\]

For a binary transition from a single state \( E(I_{stoch}) = \frac{2P-1}{\sqrt{2P(1-P)}} \) where \( P \) is the transition probability to either of the pair of states. Thus, the QF and SQF have identical correct recognition rates when:

\[
I_{obs} \sqrt{2-1} = \frac{2P-1}{\sqrt{2P(1-P)}}.
\]

Rearranging:

\[
P^2 - P + \frac{1}{4 + 2[I_{obs} \sqrt{2-1}]^2} = 0.
\]
Thus:

\[
P = \frac{1 \pm \sqrt{1 - \frac{1}{1 + 0.5 I_{\text{obs}}(\sqrt{2}-1)^2}}}{2}.
\]  

(3.31)

Suppose a decision is drawn when the decision value \( G > 2 \). The potencies of the QF and SQF filters are the same (i.e., identical decisions are drawn) when the combined observation and stochastic Biscay information for the SQF is identical in value to the QF observation information. Setting \( I_{\text{obs}} = 2 \) in Equation 3.31 yields \( P = 0.752 \). So the QF is more potent than the SQF for \( P < 0.752 \) which is consistent with the results in Table 3.2. So it is advantageous in this case to include stochastic information only if the transition probabilities differ by more than 0.5.

This example shows that encoding stochastic information needn't improve the hit rate of the filter. Although it is possible to weight each information source differently so that the SQF is more potent than the QF there is no general weighting scheme which prefers the SQF for all possible true state combinations. This is proved in the next chapter (Theorem 11, Section 4.2). The next chapter also develops conservative QF methods for use when the transition probabilities are only partially known or bounded.  

### 3.9 Comparative Study of the QF and Other Filter Philosophies

The IIIF and QF decision rates and false-positive error rates are compared. A filter must operate with care and, in the author's opinion, the robot should attempt to maximise its decision accuracy even at the expense of its decision rate. The preferred filter would be identified according to the following preference hierarchy:

1. **Ideal**: Smallest error rate and highest decision rate.
2. **Most Risk averse**: Smallest error rate and smallest decision rate.
3. **Reckless**: Largest error rate and highest decision rate.
4. **Worst case**: Largest error rate and lowest decision rate.

The QF and IIIF were compared for a simple static system comprising a three state quantity-space with a pair of close landmarks. This system represents the simplest and most typical filtering problem since (i) a static system demonstrates any filter bias induced by a single observation and (ii) filtering is required only when observations are close to landmarks and thus, filtering is necessary when the true state lies between two close landmarks.

In the experiment, there are two landmarks at 48.0 and 52.0. Observation noise is generated from a Gaussian distribution for 10 experiments. The experiments differ in the standard error \( \sigma \) of the noise which ranges from 1.0 to 10.0 in steps of 1.0. Each experiment comprises sub-experiments distinguished by the location (TS) of the true state which ranges from \( TS = 48.0 - \sigma \) to \( TS = 50.0 \) in steps of \( \frac{\sigma}{2} \). 1000 trials were obtained for each true state/standard error pair. The scatter plots in Figure 3.6 show the relative error and decision rates as a function of the expected information

\[ \text{Error} = \text{Observed} - \text{Expected} \]

\[ \text{Decision Rate} = \frac{\text{Correct Decisions}}{\text{Total Decisions}} \]

\[ \text{Expected Information} = \int_{-\infty}^{\infty} f(x) \log f(x) \, dx \]

where \( f(x) \) is the probability density function of the observation noise.
\[ E(I) = \frac{TS}{\sigma}. \] For each experiment a decision value \( G = 2 \) was used for the QF and interval widths of \( 2\sigma \), \( 2.4\sigma \) and \( 3\sigma \) successively for the IIIF. The following information was recorded:

- when both QF and IIIF estimates were both correct then the number of observations (nobs) required for convergence for each was recorded.

- otherwise, for each filter, when the filter failed to return a result then the filter’s indifference measure (im) was increased by 1.

- else, for each filter, when the filter generated a false-positive the false-positive measure (fpm) was increased by 1.

\[ Pr(\text{false positive} \mid \text{filter}) \] is the false-positive error rate and is the number of incorrect classifications against the number of classifications made.

\[ Pr(\text{false positive} \mid \text{filter}) = \frac{fpm_{\text{filter}}}{1000 - im_{\text{filter}}}. \]

\[ Pr(\neg \text{decision} \mid \text{filter}) \] is the indifference rate of the filter and is the fractional number of trials which did not produce a decision:

\[ Pr(\neg \text{decision} \mid \text{filter}) = \frac{im_{\text{filter}}}{1000}. \]

DR(filter) is the decision rate of correct classifications and is the average number of observations required to come to a decision. It is used to compare the decision rates between filter types and only those experiments when both filter types agree on the correct state are used.

\[ \frac{DR(\text{filter1})}{DR(\text{filter2})} = \frac{\text{nobs}_{\text{filter2}}}{\text{nobs}_{\text{filter1}}}. \]

We now see that the QF has a smaller false-positive error-rate than the IIIF and that the QF makes true-positive decisions more quickly than the IIIF. \( Pr(\text{false-positive} \mid \text{filter}) \) is the false-positive decision rate for those experiments for which a decision was drawn. Graph (a) in Figure 3.6 gives the difference \( \Delta \) of false-positive error rates \( \Delta = Pr(\text{false-positive} \mid \text{IIIF}) - Pr(\text{false-positive} \mid \text{QF}) \). The relative true-positive decision rate for those experiments for which a decision was drawn is simply \( -\Delta \). Part (b) of this figure is \( DR(\text{QF})/DR(\text{IIIF}) \), the relative decision rate of the QF over the IIIF when both filters draw a true-positive decision. Part (b) shows that the QF is, on average, faster than the IIIF (by as much as twice for certain true-states). Thus, the QF is more ‘ideal’ than the IIIF according to our preference hierarchy introduced at the beginning of this subsection.

Of course, increasing the observation estimate width for the IIIF will reduce the IIIF false-positive error rate. Figures 3.7, and 3.8 demonstrate that, increasing the estimate width from \( 2\sigma \) to \( 2.4\sigma \) and \( 3\sigma \) significantly reduces the IIIF decision rate without significantly changing its false-positive error rate.

### 3.9.1 How closely is the QF related to the KF?

The mean BPA QF is a linear update equation similar in form to that of the linearised KF. How similar are the two approaches?

The mean BPA Biscay information \( J \) between two states is a linear function of the observations. The value of \( J \) is an estimate of the value of \( E(I) \) and the variance of \( J \) indicates the possible
Figure 3.6: Comparison of IIF and QF for binary decision problem and various values of the expected Biscay information $E(I)$. $P_r(\text{false-positive} \mid \text{filter})$ is the false-positive decision rate for those experiments for which a decision is drawn. Graph (a) gives the difference $\Delta$ of false-positive error rates $\Delta = P_r(\text{false-positive} \mid \text{IIF}) - P_r(\text{false-positive} \mid \text{QF})$. The relative true-positive decision rate for those experiments for which a decision was drawn is simply $-\Delta$. $D_r(\text{filter})$ is the filter decision-rate which is the inverse of the average number of observations required to make a decision. Graph (b) shows the relative decision rate $D_r(\text{QF})/D_r(\text{IIF})$ for true-positive decisions. The continuous line represents mean rate with bin size 0.1.

Figure 3.7: Comparison of false-positive error rates for experiments leading to a decision: $\Delta = P_r(\text{false-positive} \mid \text{IIF}) - P_r(\text{false-positive} \mid \text{QF})$. QF uses a decision gate $G = 2$ and IIF estimate uses intervals of size $2.4\sigma$ and $3.0\sigma$. 
deviation of the estimate from the true value. Thus, the QF can be seen to be a bank of Kalman filters operating between pairs of states. The QF and KF differ in that the information form of the QF weights observations inversely to their standard errors where as the KF uses the inverse variance. Biscay information is a mahalanobis distance but the optimality of the KF requires an alternative measure. In the next chapter we will show that, unlike the KF, there is no optimal QF. We will argue that the information form of the QF uses information most efficiently for more cases than any other filter and then, in Chapter 5, we will demonstrate that the information form of the QF operating on Biscay distributions is equivalent to the KF when process and observation modelling imprecision is negligible.

3.10 Related Work

In this chapter we have considered an extension to the Dempster-Shafer theory as originally advocated by Shafer (Shafer, 1976). The reinterpretation of the additive probability semantics of the original Dempster-Shafer theory in terms of non-additive maximum probability and minimal explanation overcomes a number of problems with the original approach: computational complexity and incompatibility between frames of various granularities. In this section we explore the relationships between the $P$-norm Dempster-Shafer Theory and theories of uncertainty found in the AI literature.

Possibility Theory (Dubois and Prade, 1988a) is rooted in Fuzzy Set Theory (Zadeh, 1978). A possibility distribution function $\pi_S$ is a measure of the degree of membership of an alternative to a set $S$ or the measure of appropriateness of the labelling $S$ applied to the alternative. For example, if $S$ is the set of young ages then $\pi_{\text{young}}(32) = 0.6$ and $\pi_{\text{young}}(28) = 0.9$ indicate that the appropriateness of the linguistic label “young” for the ages 32 years and 28 years is 0.6 and 0.9 respectively. Fuzzy logic also operates over sets of alternatives. For example, $\Pi_{\text{young}}(\{28, 32\}) = \max\{0.6, 0.9\} = 0.9$ is the
### 3.10 Related Work

<table>
<thead>
<tr>
<th>Possibility Theory</th>
<th>$\infty$-norm Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pi(A \cup B) = \max{\Pi(A), \Pi(B)}$</td>
<td>$PL_\infty(A \cup B) = \max{PL_\infty(A), PL_\infty(B)}$</td>
</tr>
<tr>
<td>$\max{\Pi(A), \Pi(\neg A)} = 1$</td>
<td>$\max{PL_\infty(A), PL_\infty(\neg A)} = 1$</td>
</tr>
<tr>
<td>$N(A \cap B) = \min{N(A), N(B)}$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\min{N(A), N(\neg A)} = 0$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Pi(A) = 1 - N(\neg A)$</td>
<td>$\max{PL_\infty(A), Bel_\infty(\neg A)} = 1$</td>
</tr>
<tr>
<td>$\Pi(A) \geq N(A)$</td>
<td>$PL_\infty(A) \geq Bel_\infty(A)$</td>
</tr>
<tr>
<td>$(\forall \theta \in \Theta) \Pi_1 \oplus \Pi_2(\theta) = k_1 \Pi_1(\theta) \Pi_2(\theta)$</td>
<td>$(\forall \theta \in \Theta) PL_1(\theta) \oplus PL_2(\theta) = k_2 PL_1(\theta) PL_2(\theta)$</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of the structure and properties of Possibility theory and the $\infty$-norm Dempster-Shafer theory. The functions of possibility and $\infty$-norm plausibility are identical. However, relationships for necessity and $\infty$-norm belief differ due to differences in the dual measure.

possibility that “young” is an appropriate label for the age 28 or 32. The possibility distribution is similar to the the $P_{\infty}$ function but the dual measure for $\Pi$ is necessity $N$ where $N(A) = 1 - \Pi(\neg A)$.

The application of Possibility theory to problems of uncertainty has its roots in (Zadeh, 1978) and much work has been expended in relating the semantics of possibility and probability theories (Dubois and Prade, 1982; Dubois and Prade, 1986; Dubois and Prade, 1988b; Dubois and Prade, 1994) subject to the consistency requirement $\pi(u) \geq Pr(u)$. In (Dubois and Prade, 1992; Dubois and Prade, 1994), Dubois and Prade consider data fusion when it is difficult to represent the information supplied by the sources by means of a single probability distribution due to the lack of statistical evidence. They interpret $\Pi$ and $N$ as upper and lower bounds on some unknown probability distribution. Necessity is assigned, initially, to a family of nested subsets (called fractiles) where each subset guarantees coverage of the true state to some probability lower bound. Their method requires reworking the probability distribution into a consonant family of sets and then building an outer possibilistic approximation. A possibility distribution is then obtained from $N$. However, the product rule of combination needn’t be unbiased for conservative possibility distributions and, further, the interval $[N(A), \Pi(A)]$, although initially conservative, needn’t remain so (as we will demonstrate below). The authors assert:

the results of the pooling should be interpreted in an ordinal way, i.e. as a ranking of the values of the parameter under study, in terms of their respective plausibility. It is more difficult to interpret the resulting degrees of possibility with a frequentist approach.

However ordinal information is often sufficient for practical purposes.

The most significant contribution of Dubois and Prade’s recent paper (Dubois and Prade, 1994), in my opinion, is an adaptive combination rule that gradually turns from conjunctive combination into a disjunctive one as the disagreement between the sources increases. The disagreement is measured as the inverse of the height of the intersection of two possibility distributions $\pi_1$ and $\pi_2$: $h(\pi_1, \pi_2) = \max_{\theta \in \Theta} \pi_1(\theta) \pi_2(\theta)$.

Although other ways of interpreting possibility distributions in terms of probability distributions have been proposed (Dubois and Prade, 1982) perhaps the most damning testimony against the probabilistic interpretation of Possibility measures was advanced by Sudkamp (Sudkamp, 1992). Sudkamp demonstrates that, although we might be able to find consistent possibilistic representations of probability values, there is no mechanical mapping between probability distributions and possibility distributions that satisfies natural requirements which can preserve the properties of projection.
(inference), non-interaction (fusion) or conditionalization. I have included Sudkamp's natural conditions to demonstrate their generality. An n-dimensional probability to possibility transformation is a mapping $T_n$ from $P$ to $\Pi$ that satisfies $T1$ to $T4$ below. The image of a probabilistic distribution $P = [p_1, \ldots, p_n]$ under $T_n$ is the possibility distribution $T_n(P) = [t_n(p_1), \ldots, t_n(p_n)] = [\pi_1, \ldots, \pi_n]$.

**T1.** $T_n$ is a bijection.

**T2.** $t_n(p_j) = \pi_j = 1$ if $p_j$ has a maximal probability in $P$.

**T3.** $T_n([p_{\sigma(1)}, \ldots, p_{\sigma(n)}]) = [\pi_{\sigma(1)}, \ldots, \pi_{\sigma(n)}]$ where $\sigma$ is a permutation of $[1, 2, \ldots, n]$.

**T4.** If $P = [p_1, \ldots, p_n] \in P^n$ then $P' = [p_1, \ldots, p_n, 0] \in P^{n+1}$ and $T_{n+1}(P') = [t_n(p_1), \ldots, t_n(p_n), 0]$.

Sudkamp concludes by reiterating Kandel:

> an important aspect of the concept of possibility distribution is that it is non statistical in nature. As a consequence ... the only connection between $\Pi(Y)$ and $P(Y)$ is that impossibility implies improbability but not vice-versa. Thus $\Pi(Y)$ cannot be inferred from $P(Y)$, nor can $P(Y)$ be inferred from $\Pi(Y)$.

We will return to the probabilistic interpretation of the possibility distribution later when we investigate a slightly weaker consistency requirement for probability to possibility mappings. Namely, those which preserve consistent measures of ratios of evidence between pairs of hypotheses. This reflects the approach advocated in this thesis which is a non-probabilistic, but all the same, statistical interpretation of the pair-wise ordering of hypotheses.

It is well known that 1-norm theory consonant plausibility functions are possibility functions (Dubois and Prade, 1988a) in which case $\Pi \equiv P1$ and $N \equiv Bel1$. It is interesting to compare $\infty$-norm and Possibility theories by interpreting $N$ as a degree of belief and $\Pi$ as a degree of plausibility. Table 3.3 compares the properties of Possibility theory and the $\infty$-norm theory for $A, B \subseteq \Theta$ and $\theta \in \Theta$. The relationships between $\infty$-norm plausibility measures are identical to those for $\Pi$ in the Possibility theory. The dual measure between plausibility and belief and between possibility and necessity differ, however, and induce non-conforming statements of belief and necessity. The $\infty$-norm approach extends the interpretation of Possibility theory to non-consonant mass functions. Possibility theory can be derived from $\infty$-norm theory simply by equating $Pl_{\infty}$ and $\Pi$ and then defining the information redundant necessity measure $N$ from $Pl_{\infty}$: $N(A) = 1 - Pl_{\infty}(\neg A)$. However, $\infty$-norm theory also proclaims $Bel_{\infty}(A)$, which intuitively captures the notion of the most potent argument which entails $A$. From an argument (i.e. path) centred perspective we can see that $Bel_{\infty}(\{A\})$ and $Pl_{\infty}(\{A\})$ indicate the strongest arguments that unambiguously support $A$ and the strongest argument that might possibly support $A$ respectively. Possibility theory does not have an equivalent $Bel_{\infty}$ concept. Is the $Bel_{\infty}$ concept useful? Interpreting $N$ as a degree of belief, Possibility theory requires that no more than one elementary proposition has non-zero necessity which may be too restrictive for many applications. The following example will, I hope, demonstrate the utility of the $Bel_{\infty}$ measure.

---

$N(A \cap B) = \min\{N(A), N(B)\}$ and let $A = \{XvY\}$ and $B = Y$ where $X \cap Y = \emptyset$. Thus, $N(Y) = \min\{N(X \cup Y), N(Y)\}$ and therefore, $N(Y) \leq N(X \cup Y)$. Generalising this result, $N(A \cup B) \geq \max\{N(A), N(B)\}$. Now, suppose by contradiction that two elementary propositions have two non-zero necessity values. Partition the frame into $\{A, \neg A\}$ so that $A$ and $\neg A$ each contain one of these elementary propositions. By our previous result we must have $N(A) > 0$ and $N(\neg A) > 0$ which contradicts a fundamental theorem of Possibility theory, namely $\min\{N(A), N(\neg A)\} = 0$. 

---

Footnote 7: $N(A \cap B) = \min\{N(A), N(B)\}$ and let $A = \{XvY\}$ and $B = Y$ where $X \cap Y = \emptyset$. Thus, $N(Y) = \min\{N(X \cup Y), N(Y)\}$ and therefore, $N(Y) \leq N(X \cup Y)$. Generalising this result, $N(A \cup B) \geq \max\{N(A), N(B)\}$. Now, suppose by contradiction that two elementary propositions have two non-zero necessity values. Partition the frame into $\{A, \neg A\}$ so that $A$ and $\neg A$ each contain one of these elementary propositions. By our previous result we must have $N(A) > 0$ and $N(\neg A) > 0$ which contradicts a fundamental theorem of Possibility theory, namely $\min\{N(A), N(\neg A)\} = 0$. 

---
3.10 Related Work

Example. An individual has a set of goals $G$ which are ranked in terms of their required completion time. \(^8\) $G_1 \succ G_2$ indicates that goal $G_1$ is required more urgently than goal $G_2$ and we will use an increasing scale of rank $R$ so that $R(G_1) > R(G_2) \iff G_1 \succ G_2$. A single action from a set of potential actions $A$ is sufficient to reach each goal and the problem is to construct a schedule which is an action sequence whose rank ordering satisfies the goal completion order. We assume that all action sequences are allowed.

When a goal $G_i$ is achievable by one and only one action $A_j$ then it is reasonable to assign the goal rank value to this particular action $R(A_j) = R(G_i)$. Further, if a number of goals $G$ are achievable by a single action then the rank value of the most urgent goal should be assigned to this action $R(A_j) = \max_{G_i \in G} R(G_i)$. However, some goals may be satisfied by more than one action $A_j \lor \ldots \lor A_{j'}$. In this case, the goal rank value should be assigned to the disjunction of the alternative actions $R(A_j \lor \ldots \lor A_{j'}) = R(G_i)$, and the individual must choose some other criterion to make the choice.

We can represent all these different ways of assigning rank values to actions using a frame of discernment $\Theta = A$. The mass value of a proposition is simply the normalised rank value of the actions represented by the proposition:

$$m(A) = \frac{R(A)}{\max_{A \subseteq A} R(A)}.$$

$Pl_\infty$ and $Bel_\infty$ then become:

- $Bel_\infty(A)$ is largest rank definitely assigned to some action in $A$,  
- $Pl_\infty(A)$ is the largest rank possibly assigned to some action in $A$.

Suppose that goals $G_1, G_2, G_3, G_4$ are achieved by the actions $A_1, A_2$ and $A_3$:

$$G_1 \rightarrow \{A_1, A_2\}, \ G_2 \rightarrow \{A_2\}, \ G_3 \rightarrow \{A_1\}, \ G_4 \rightarrow \{A_3\}$$

and an individual specifies the following preference ordering:

$$G_1 \succ G_2 \succ G_3 \succ G_4$$

with rankings $R(G_1) = 1.0, R(G_2) = 0.75, R(G_3) = 0.5$ and $R(G_4) = 0.25$, then:

$$m(\{A_1, A_2\}) = 1.0, \ m(\{A_2\}) = 0.75, \ m(\{A_1\}) = 0.5, \ m(\{A_3\}) = 0.25.$$  

and:

$$Bel_\infty(\{A_1\}) = 0.5, \ Pl_\infty(\{A_1\}) = 1.0,$$

$$Bel_\infty(\{A_2\}) = 0.75, \ Pl_\infty(\{A_2\}) = 1.0,$$

$$Bel_\infty(\{A_3\}) = 0.25, \ Pl_\infty(\{A_3\}) = 0.25.$$  

Thus, $A_1 \succ A_3$ and $A_2 \succ A_3$ but $A_1$ and $A_2$ cannot be ranked as their range of possible ranking overlaps. If we were to rewrite this example in terms of Possibility measures $\Pi$ and $\Pi'$ with $\Pi(\{A\}$ representing the most informative upper bound on the rank of $A$ then:

\(^8\)We will be interested in finding an ordering over crisp propositions given indifference as to which elements in a compound proposition are actually ranked. This differs from Fuzzy ranking problems (e.g., (González, 1990)) which are concerned with rank ordering fuzzy intervals whose boundaries are blurred.
and no ranking can be drawn between any pairs of atoms.

Now, consider constructing a social preference ordering of actions from a population of goal orderings. For readability, we define lower and upper rankings, \( L = \text{Bel}_\infty \) and \( U = \text{Pl}_\infty \), respectively. \( L_1 \) and \( U_1 \) refer to individual \( i \)'s ranking and \( L \) and \( U \) (without indices) refers to the social ranking. Let \( LU \) refer to either \( L \) or \( U \). The following are rational laws for the aggregation of individual rankings over sets of alternatives \( A, B \) and \( C \):

1. **Commutativity.** \( LU_1 \oplus LU_2 = LU_2 \oplus LU_1 \).

2. **Associativity.** \( (LU_1 \oplus LU_2) \oplus LU_3 = LU_1 \oplus (LU_2 \oplus LU_3) \).

3. **Monotonicity.** If \( LU_1(A) > LU_1(B) \) and \( LU_2(A) \geq LU_2(B) \) then \( LU(A) > LU(B) \).

4. **Equivalence.** If \( LU_1(A) = LU_1(B) \) and \( LU_2(A) = LU_2(B) \) then \( LU(A) = LU(B) \).

If an alternative \( \theta \in \Theta \) is ranked \( R_1 \) and \( R_2 \) by two individuals then the social ranking will sensibly reflect the average of this ranking:

\[
R(\theta) = \frac{R_1(\theta) + R_2(\theta)}{2}.
\] (3.32)

In general, rank values can be assigned to sets of alternatives. If individuals 1 and 2 assign rank \( R \) to sets of alternatives \( B \) and \( C \) respectively then there is some alternative in \( A = B \cap C \) which can sensibly be assigned the rank according to Equation 3.32. Thus:

If \( A = B \cap C \) then \( R(A) = \frac{R_1(B) + R_2(C)}{2} \).

There may be other sets, say \( D \) and \( E \), different from \( B \) and \( C \) such that \( A = D \cap E \). The social rank value formed using \( D \) and \( E \) may be different to that formed from \( B \) and \( C \). We will use an extended Borda ranking measure,\(^9\) which assigns the largest \( R \) value to the highest ranked alternative. There are many such measures, for example, when \( A_1, \ldots, A_n \) are ranked \( A_1 > \ldots > A_n \) then \( R(A_k) = \frac{k}{n} \) is a possibility. However, details of the Borda ranking are irrelevant for the following discussion. An action set may achieve more than one goal and each goal contributes an \( R \) value to the action set. Under these circumstances the largest possible \( R \) value is chosen to promote the most urgent goal. Of course, once the most urgent goal is achieved then all other goals satisfied by that action set will also be achieved. Thus, neglecting the normalisation constant in Equation 3.33:

\[
R(A) = \max_{B,C \subseteq A = B \cap C} \{ R_1(B) + R_2(C) \}.
\] (3.33)

Equation 3.33 conforms with our laws for rational aggregation presented above. Also, Equation 3.33 can be rewritten in a form similar to the Dempster Rule of Combination. Define \( \log w_i(A) = R_i(A) \), thus:

\[
w(A) = \exp \left[ \max_{B,C \subseteq A = B \cap C} w_1(B)w_2(C) \right].
\]

\(^9\)The classic Borda count assigns \( n - 1, n - 2, \ldots \) points for each first, second etc. rank (see, for example (Nurmi et al., 1990)).
3.10 Related Work

The numeric interval $[\log L(A), \log U(A)]$ indicates the range of possible ordering values that the highest ranked proposition in $A$ is allowed. The indifference (ignorance) encoded by the interval is an indication of the freedom any scheduler can enjoy when constructing a total order of the chosen actions.

To complete our investigation into the utility of the $p$-norm meta-theory, I will discuss Murphy’s adaptive form of the Dempster Rule of Combination (Murphy, 1996):

$$m(A) = \frac{\sum_{A=B\cap C} m(B) m(C)^p}{\sum_{B\cap C \neq \emptyset} m(B) m(C)^p}$$

for $0 \leq p \leq 1$. This rule is simply the $p$-norm Dempster rule of combination without the $p^{th}$ root. Murphy was motivated to propose this rule as the original Dempster rule was unable to adjust to sensing circumstances. For example, if a robot is moved to a better viewpoint and the landmark under attention is expected to persist, the new observation should be more informed and should count more than a new observation taken from the same vantage. Murphy thought that the rule of combination should adapt in response to the changing context and he proposed adjusting $p$ dynamically in response to environment conditions. As $p$ approaches 1 belief accrues quickly and decays slowly and the combination rule is optimistic. As $p$ approaches 0, the robot suspects that its new viewpoint is more informed and it is pessimistic about its previous observations. To compute $p$ Murphy offers four heuristics:

- If landmark persistence is high and clutter is high then $p$ is optimistic.
- If discriminability is improving then $p$ is pessimistic.
- If distinctiveness is good then $p$ is optimistic.
- If tracking error is increasing then $p$ is pessimistic.

Following Murphy, $p$-norm theories approaching $-\infty$ (i.e. $p \to -\infty$) prefer the minimal evidential path for each proposition and are, in a sense, pessimistic. However, theories approaching $\infty$ prefer the maximum path and are more optimistic. Theories close to $p = 0$ can be the most opportunistic and the most unstable as is demonstrated in Figure 3.9. So, although we will use only the $\infty$-norm theory in the remainder of this thesis, there is scope for application for all $p$-norm theories.

Various possibility measure combination rules have been investigated (Dubois and Prade, 1994) and two classes of rules were identified: conjunctive and disjunctive. Conjunctive aggregation makes sense if all the sources are considered as equally and fully reliable, while disjunctive aggregation corresponds to a weaker reliability hypothesis. Disjunctive aggregation uses the min operation (i.e. $\pi_1(x) \oplus \pi_2(x) = \min\{\pi_1(x), \pi_2(x)\}$) which corresponds to a purely logical view of the combination process: the source which assigns the least possibility degree to a given value is considered as the best-informed with respect to this value. When all sources perfectly agree there is no reinforcement effect. Alternatively, under the independent opinion pool conjunctive rule (i.e. $\pi(x) \oplus \pi_2(x) = \pi_1(x) \pi_2(x)$), if all sources agree that a value is possible then the aggregated possibility is strictly greater than each individual possibility (given some degree of independence of the sources). This dissertation assumes fault free sensors and investigates conjunctive aggregation only. It is perhaps important to note that the certainty weighted conjunction rule used in this thesis is not necessarily associative and is, therefore, not a T-norm aggregation operator (Bonissone, 1987) (However, the inference projection operation on $\infty$-norm mass is a T-conorm operation).
We mentioned earlier that consistent probability-possibility transformations may be available for ratios of probability and possibility values since problems due to normalisation would not arise. In this case, the probability value would not indicate the frequency of a proposition but, instead, it would indicate the relative probability of the likeliest path which reaches that proposition. A transformation of probability into possibility values would require a mapping which, for any two hypothesis $A$ and $B$, maps $\frac{P(A)}{P(B)}$ onto $\frac{\Pi(A)}{\Pi(B)}$ in such a way that necessity and possibility can be interpreted as lower and upper bounds on the probability distribution of the most preferred argument (path) to each proposition. It is straightforward to show, via an example, that such a transformation is not possible. Consider a binary system $\{Q, -Q\}$. If $N$ (necessity) and $\Pi$ (possibility) were lower and upper bounds on $Pr$ (probability) for a single path:

\[ N(Q) \leq Pr(Q) \leq \Pi(Q), \]
\[ N(-Q) \leq Pr(-Q) \leq \Pi(-Q). \]

We can construct an upper bound on the likelihood using the dual measure:

\[ \frac{Pr(Q)}{Pr(-Q)} \leq \frac{\Pi(Q)}{\Pi(-Q)} = \frac{\Pi(Q)}{1 - \Pi(Q)}. \]  \hspace{1cm} (3.34)

Suppose, two distributions, $D_1$ and $D_2$, are fused using probability and possibility theories:

\[ D_1 = \{Pr_1(Q) = 0.1, Pr_1(-Q) = 0.9\}, \]
\[ D_2 = \{Pr_2(Q) = 0.7, Pr_2(-Q) = 0.3\}. \]
Conservative bounds (and consistent necessity-possibility measures) are available for distributions \(D_1\) and \(D_2\) which are \(\{Pr_1(Q) \in [0.0, 0.1], Pr_1(\neg Q) \in [0.9, 1.0]\}\) and \(\{Pr_2(Q) \in [0.0, 0.8], Pr_2(\neg Q) \in [0.2, 1.0]\}\) respectively. Assuming that the sources of evidence are independent then using Bayes rule on the true probability distributions we obtain:

\[
\frac{Pr(Q)}{Pr(\neg Q)} = 0.429,
\]

Using the independent opinion pool conjunctive rule we obtain:

\[
\Pi(Q) = 0.08, \quad \Pi(\neg Q) = 1.0
\]

and \(\frac{\Pi(Q)}{1-\Pi(Q)} = 0.087\). Clearly, Inequality 3.34 does not hold and treating \(\Pi\) and \(N\) as linear functions of upper and lower probabilities can generate inconsistent bounds on fused estimates. As the above example demonstrates, this problem exists for pair-wise (compound) path comparisons.

Can we interpret \(Bel_\infty\) and \(Pl_\infty\) as lower and upper bounds on the probability distribution of the most preferred argument (path) for a proposition? This is a reasonable question since the dual measure in the \(\infty\) norm theory is weaker than that of possibility theory. The following example illustrates that, again, this is not a valid interpretation, in general.

**Example.** Suppose, again, that we are to fuse two distributions and assume that the true distributions are \(\{Pr_1(A) = 0.1, Pr_1(B) = 0.9, Pr_1(C) = 0.1\}\) and \(\{Pr_2(A) = 0.1, Pr_2(B) = 0.6, Pr_2(C) = 0.3\}\). Interpreting \(Bel_\infty\) and \(Pl_\infty\) as lower and upper bounds on these distributions we may consider two belief functions on the frame \(\{A, B, C\}\) with only two focal elements in each, \(m_1(\{A, B\}) = 1, m_1(\{C\}) = 0.1\) and \(m_2(\{A\}) = 0.1, m_2(\{B, C\}) = 1\):

\[
\begin{align*}
Bel_1(\{A\}) &= 0.0 \leq Pr_1(A) \leq 1.0 = Pl_1(\{A\}), \\
Bel_1(\{B\}) &= 0.0 \leq Pr_1(B) \leq 1.0 = Pl_1(\{B\}), \\
Bel_1(\{C\}) &= 0.1 = Pr_1(C) = 0.1 = Pl_1(\{C\})
\end{align*}
\]

and:

\[
\begin{align*}
Bel_2(\{A\}) &= 0.1 = Pr_2(A) = 0.1 = Pl_2(\{A\}), \\
Bel_2(\{B\}) &= 0.0 \leq Pr_2(B) \leq 1.0 = Pl_2(\{B\}), \\
Bel_2(\{C\}) &= 0.0 \leq Pr_2(C) \leq 1.0 = Pl_2(\{C\})
\end{align*}
\]

Combining the true probability distributions using Bayes rule yields \(\frac{Pr(B)}{Pr(A)} = 54\). However, combining the lower and upper conservative estimates using the \(\infty\) Norm Dempster Rule of Combination yields the following Bayesian BPA:

\[
m(\{A\}) = 0.1, \quad m(\{B\}) = 1.0, \quad m(\{C\}) = 0.1.
\]

Clearly, \(\frac{Pl_\infty(\{B\})}{Bel_\infty(\{A\})} < \frac{Pr(B)}{Pr(A)}\). The preceding discussion, I hope, demonstrates that a straight-forward interpretation of possibility and \(\infty\) norm plausibility ratio measures in terms of probability ratios is not available.

After developing the \(\infty\) norm theory independently, our attention was subsequently drawn to \(\varepsilon\)-belief function theory (Benferhat et al., 1995). The theory of \(\varepsilon\)-Belief Functions is based on the \(\varepsilon\)-semantics of Adams (Adams, 1975; Pearl, 1988) both of which were proposed as models for default reasoning. In these theories a conditional probability (mass) with a value infinitesimally close to

\footnote{Many thanks to Dr Nic Wilson at Oxford Brookes University for introducing me to this theory.}
0 or 1 is associated with the plausibility of the default rule. \( \varepsilon \)-semantics can be used to address the problem of whether Tweety can fly (f) given that he is a penguin (p) and given the background default knowledge that penguins are birds (b) and that birds can fly but also that penguins cannot fly (Pearl, 1988). The default rules \( \{ b \rightarrow f, p \rightarrow \neg f, p \rightarrow b \} \) are described by the following \( \varepsilon \) probability values:

\[
\begin{align*}
P(-f|p) &= 1 - \varepsilon \\
P(b|p) &= 1 - \varepsilon \\
P(f|b) &= 1 - \varepsilon
\end{align*}
\]

Tweety cannot fly as he is a penguin, Tweety is a bird as penguins are birds, Tweety can fly as he is a bird.

Since \( P(f|p) \geq P(f|p, b)P(b|p) \) then the belief \( P(f|p, b) \) that Tweety can fly given that he is a bird and a penguin is bounded above:

\[
P(f|p, b) \leq \frac{P(f|p)}{P(b|p)} = \frac{\varepsilon}{1 - \varepsilon} = O(\varepsilon).
\]

Thus, \( P(-f|p, b) = 1 - O(\varepsilon) \) and Tweety cannot fly.

Although the purpose of default reasoning is far removed from that of noise filtering, the mathematical properties of the \( \varepsilon \)-belief theory and \( \infty \)-norm theory applied to non-parametric models are similar. An \( \varepsilon \)-mass function on \( \Theta \) is a function \( m_{\varepsilon} : 2^{\Theta} \rightarrow [0, 1] \) such that, for each \( A \subseteq \Theta \), either \( m_{\varepsilon}(\emptyset) = 0 \), \( m_{\varepsilon}(A) = \varepsilon_A \), or \( m_{\varepsilon}(A) = 1 - \varepsilon_A \), where \( \varepsilon_A \) is an infinitesimal hyper-real. By introducing an extra term \( \varepsilon \rightarrow 0 \) into the mean and median estimator BPAs defined in this chapter we obtain an equally valid set of extreme and unbiased BPAs:

\[
m_{\varepsilon, \text{mean}}(Q|x) \triangleq \begin{cases} 
1 & \iff z \in Q, \\
\exp\left(\frac{-|z|}{\varepsilon}\right) & \iff z \in -Q.
\end{cases}
\]

\[
m_{\varepsilon, \text{med}}(Q|x) \triangleq \begin{cases} 
1 & \iff z \in Q, \\
n\varepsilon & \iff z \in -Q.
\end{cases}
\]

The \( \infty \)-norm theory is quite distinct from the \( 1 \)-norm theory applied to \( \varepsilon \)-belief functions. To illustrate this, consider a three region problem \( \{Q_0, Q_1, Q_2\} \) and the fusion and subsequent propagation of two observations \( z_1 \in Q_0 \) and \( z_2 \in Q_2 \) using median estimators. A binary decision approach to the \( \varepsilon \)-belief theory would yield the following BPAs for \( z_1 \) and \( z_2 \) respectively (neglecting the renormalisation constant): \( \{(1-\varepsilon)^2, \varepsilon(1-\varepsilon), \varepsilon^2\} \) and \( \{\varepsilon^2, \varepsilon(1-\varepsilon), (1-\varepsilon)^2\} \). Their subsequent fusion would yield the uninformative (and unnormalised) BPA: \( \{\varepsilon^2(1-\varepsilon)^2, \varepsilon^2(1-\varepsilon)^2, \varepsilon^2(1-\varepsilon)^2\} \). Renormalisation gives \( \{0.33, 0.33, 0.33\} \). Similarly, fusion using the \( \infty \)-norm theory yields \( \{(\varepsilon\varepsilon)^2, (\varepsilon\varepsilon)^2, (\varepsilon\varepsilon)^2\} \) which becomes \( \{1, 1, 1\} \) after renormalisation. So the plausibility ratio values are identical for both theoretical approaches. However, if the BPAs are now subject to propagation according to the rules \( Q_0 \rightarrow \{Q_0\}, Q_1 \rightarrow \{Q_1\} \) and \( Q_2 \rightarrow \{Q_1\} \) then we see that the plausibility ratios differ between the theories. The \( \varepsilon \)-belief theory yields \( \{0.33, 0.66, 0.0\} \) whereas the \( \infty \)-norm theory yields \( \{1, 1, 1\} \). Unlike the \( \infty \)-norm BPA which is uninformative about \( Q_0 \) and \( Q_1 \), the \( \varepsilon \)-belief theory prefers \( Q_1 \). Thus the \( \varepsilon \)-belief theory and the \( \infty \)-norm theory exhibit distinct behaviours.

### 3.11 Conclusions

This chapter has addressed the problem of sensor data-fusion when noisy data is inter-related by incomplete (qualitative) observation and (possibly stochastic) process models. A simple and computationally attractive independent opinion pool filtering framework based on path and observation weighted Dempster-Shafer theory of evidential reasoning was proposed.
3.11 Conclusions

The \( \infty \)-norm Dempster-Shafer theory can be viewed as a Model-Based Multiple Hypothesis Filter (MHT) over individual arguments (paths) (Cox and Leonard, 1994; Moran, 1994). Essentially, it is a Viterbi Maximum Likelihood Decoder (Bertsekas, 1987) (VMLD) for low grade propositional representations applied to plausibility measures. It uses a non-parametric generalised theory of evidential reasoning to identify the most potent path (i.e. decoded message) from observations (i.e. codes) from the environment physics. The MHT algorithm proposed by (Cox and Leonard, 1994) maintains clusters of hypotheses, each cluster containing those hypotheses which share observations. Thus, the QF hypotheses fall into a single cluster in this framework. The QF maintains a preference ordering over states and the corrigibility of the order dictates when confident decisions can be made. Non-parametric BPAs for combined mean and median estimation were derived from a simple requirement for unbiased filtering based on mass convergence. Standard error confidence measures maintain upper bounds on the quality of observation information and the framework can be extended to correlated observations using covariance measures. The quadratic complexity of the covariance representation over states and observations is an improvement over the necessary exponential complexity for representing a necessarily complete description of conditionals in a probabilistic framework (Neapolitan, 1990).

The probability-free interpretation of mass and use of mass covariance matrices overcomes the much cited problem of BPA independence problems for which Dempster’s rule of combination is not consistent with probability theory (Voorbraak, 1991; Liu and Bundy, 1994; Murphy, 1996; Wu et al., 1996). Both Murphy (Murphy, 1996) and Wu et al. (Wu et al., 1996) generalise the Dempster rule of combination by introducing an adaptive function which operates on the pooled mass to obtain the true mass assignment. Liu (Liu and Bundy, 1994) presents a comparison between Dempster-Shafer theory and Bundy’s Incidence calculus. Incidence calculus maintains evidence assigned to possible worlds and, as Liu argues, can overcome the dependent evidence constraint of the Dempster-Shafer theory. However, this approach seems to require a complete probabilistic model and it is unclear what its advantages are over a straight Bayesian implementation. In situations where no information of the form of the evidence is available, Parsons (Parsons, 1994) suggests that the range of possible evidence values should be represented as a numerical interval and manipulated using qualitative arithmetic. However, such methods can diverge rapidly. Our approach does not rely on an interpretation of the mass values in terms of probability theory but it operates with state preference orderings and a variance driven decision method. Dependencies between mass functions, which indicate uncertainty in the preference ordering, are represented explicitly in our system by their covariances.

The \( \infty \)-norm theory highlights the strong theoretical relationships between Dempster-Shafer theory and Possibility theory beyond consonance. It demonstrates how straightforward it is to extend the basic Dempster-Shafer mass interpretation of evidence to Possibility theory and, further, how straightforwardly Shafer’s treatise (Shafer, 1976) and proofs therein can be transferred to Possibility theory. We have justified our choice of the \( \infty \)-norm theory from the all \( p \)-norm theories for non-parametric noise filtering. The \( \infty \)-norm theory is a path centred (Lagrangian) view of evidence. Although the Lagrangian interpretation is practically pleasing it is theoretically less than ideal. In cases where the Lagrangian interpretation differs markedly from the Eulerian interpretation the QF can produce highly inconsistent estimates. In the next chapter we will classify situations when the Lagrangian and Eulerian interpretations are equivalent. The next chapter investigates a number of topics within the QF framework:

- efficiency (optimal mass convergence rates).
• sensitivity (conservative estimates of quality measures when information is incomplete).

• robustness (bias and quality when many arguments abound).
Chapter 4

Efficiency, Sensitivity and Robustness of the Qualitative Filter

I had worked hard for nearly two years, for the sole purpose of infusing life into an inanimate body. For this I had deprived myself of rest and health. I had desired it with an ardour that far exceeded moderation; but now that I had finished, the beauty of the dream vanished, and breathless horror and disgust filled my heart.

M. W. Shelley
Frankenstein or The Modern Prometheus
Chapter V

4.1 Introduction

The previous chapter presented the bare bones of the Qualitative filter which constructs a preference ordering of qualitative states using non-parametric estimation of $\infty$-norm Dempster-Shafer mass. Evidence is combined using a weighted opinion pool. In this chapter, we endeavor to find a description of the weights for optimal estimation convergence rates. The filter which uses optimal weights is called the information filter and, in the second half of this chapter, we explore its properties.

4.2 Efficient Estimators

It is important to maintain minimal uncertainty in our estimate if an early decision is to be made. However, we must first contrast the problem of maximising $I$ with the problem of maximising Fisher information in quantitative methods. Quantitative MMSE methods (e.g., the Kalman filter) generally estimate some parameter whose true value is independent of the filtering method used. The Qualitative filter, however, maintains an estimate of the mass. We adjust the weights to minimise the uncertainty in the estimate, but this also affects the expected mass values, which we are attempting to estimate. Thus, unlike quantitative methods where the variance in the estimate is minimised, our decision rule indicates that we must seek to minimise the ratio of the variance of the mass likelihood ratio to the value of the mass likelihood ratio:

$$\frac{V_i}{m_i^2}.$$
4.2 Efficient Estimators

We call the effect of the weights on the estimated value the *introspection property*. We consider the effects that the introspection property has on theories developed by naively comparing weighted log likelihood estimators with the Kalman filter (Pearl, 1988; Hummel and Manevitz, 1996).

We can use the flexibility of our observation weights $K_1$ and $K_2$ in Equation (3.1) (reproduced here):

$$\pi_1 = \left[ \sum_p \left( \prod_{i=1}^p \pi_{i} \right) \right]^{\frac{1}{p}}$$

for two observation sets to tune the decision rate of our filter. We seek an optimal filter.

**Definition 10** When $x_T$ is the true state of the system and a filter has weights $\{K_i\}$ then a filter with weights $\{K_i\}_0$ is optimal if for every region pair:

$$\forall \{K_i\}, x_T \mid E(I_{\{K_i\}_0}) \geq E(I_{\{K_i\}}).$$

Unfortunately, unlike the linear Kalman filter, there is no optimal Qualitative filter!

**Theorem 11** An optimal filter does not exist.

**Proof** The aggregate of two Biscay information values $I_1$ and $I_2$ is, in general form:

$$E(I_{12}) = \frac{E(I_1) + \rho E(I_2)}{\sqrt{1 + S^2 + 2\rho S}}$$

where $\rho$ is the correlation term and $S = \frac{K_{12} \sigma_1 \sigma_2}{K_{11} \sigma_1}$. The only free parameter. For the filter to be optimal we must have $\left. \frac{dE(I_{12})}{dS} \right|_{S=0} = 0$,

$$\frac{dE(I_{12})}{dS} = \frac{E(I_2) - E(I_1)\rho}{(1 + S^2 + 2\rho S)^{\frac{3}{2}}} = 0.$$ 

Thus, the optimal value of $S$ is:

$$S(E(I_1), E(I_2), \rho) = \frac{E(I_2) - E(I_1)\rho}{E(I_1) - E(I_2)\rho}.$$

Clearly, the optimal value for $S$ is not independent of either $E(I_1)$ nor $E(I_2)$.

Since an unbiased, optimal filter depends on knowledge of $S$, we conclude that, in the general filtering problem, no optimal filter can be found. Alternatively, we aim to find a filter which is optimal in a sub-space of $E(I_1)$ and $E(I_2)$, in the sense that the volume of this sub-space is larger than that for any other filter.

For the one-dimensional Kalman filter, the Fisher information formed by the fusion of two estimates is the sum of the Fisher information of each estimate. By analogy, we propose the following *information filter* in which the information of the fused estimate is proportional to the sum of the information assigned to each estimate \(^1\). Thus, assigning $K_1 = \frac{1}{\sqrt{V_1}}$ and $K_2 = \frac{1}{\sqrt{V_2}}$ we obtain:

$$\frac{M_1^t}{\sqrt{V_1^t}} = K \left( \frac{M_1^{t-1}}{\sqrt{V_1^{t-1}}} + \frac{M_t}{\sqrt{V_t}} \right). \tag{4.1}$$

\(^1\)The name *information filter* is used since it is the weighted sum over the information content of each observation set.
for some constant $K$. Unlike the Kalman filter where the information is not scaled, the value for $K$ in the Qualitative filter is constrained by the introspection property. Taking the variance of both sides we obtain:

$$K^2 = \frac{1}{2 \left( 1 + \frac{C}{\sqrt{V_{t}^{-1} V_{t}}/V_{t}} \right)}$$  \hspace{1cm} (4.2)

Without loss of generality we may assign:

$$M_t = \left( \frac{M_{t-1}^{M_1}}{\sqrt{V_{t}^{-1}}} + \frac{M_t}{\sqrt{V_{t}}} \right).$$ \hspace{1cm} (4.3)

and:

$$V_t = 2 \left( 1 + \rho_{t-1} \right)$$ \hspace{1cm} (4.4)

where $\rho_{t-1} = \frac{C}{\sqrt{V_{t}^{-1} V_{t}}}$ is the correlation between estimates $M_{t-1}$ and $M_t$. Equations 4.3 and 4.4 constitute the Qualitative Information filter. It is intuitively pleasing to see that the Information filter can be derived from the unweighted filter (i.e. $K_1 = K_2 = 1$) with $\sqrt{V_t^{-1}}$ and $M_t^{M_1}$ scaled so that $\sqrt{V_t^{-1}} = \sqrt{V_t}$. Scaling ensures that when the decision strengths of input masses are comparable, the mass values themselves are comparable. This is crucial as the unweighted filter may miss an almost certain decision. Consider for example combining the following independent mass ratios: $M_1 = 1.0$, $M_2 = 0.3$ with standard deviations $\sigma_1 = 1.0$ and $\sigma_2 = 0.1$ respectively. The unweighted filter gives $M_{1,2} = 1.3$ with $\sigma_{1,2} = 1$. Thus, $M_{1,2}$ is of the same order as $\sigma_{1,2}$ and no decision is drawn. However, the confidence in observation 2 and the lack of conflict with observation 1 indicates that, intuitively, a decision (with decision gate $G = 3$) should be drawn. This is the case for the weighted filter which yields $\frac{M_2}{\sigma_2^2} = 4$.

We will now compare the Information filter derived in this section with the unweighted filtering scheme and with a filter proposed by Hummel (Hummel and Manevitz, 1990) for aggregating estimates of belief functions. The mathematical form of the Hummel filter is analogous to the Kalman filter and is appropriate (and optimal) only when the mean estimate of the probability/belief function is identical between sources. In our case, the qualitative models will generate different mean estimate values and the QF cannot seek common ground between them. For this reason, the QF seeks to identify some consensus only in the ordering of states and in so doing attempts to obtain the largest information value ($I$) as opposed to the minimum variance of the combined evidence.

We compare unweighted and inverse variance weighted schemes with the information filter proposed in this thesis. We conclude that although neither weighting scheme is preferred for all scenarios, on the whole the information filter has better statistical convergence properties. We proceed by examining the mutual information obtained by fusing two independent observations (indexed 1 and 2) and consider the expected log ratio of the decision value $R = E \left( \frac{M_t}{\sigma_t} \right)$ obtained by the QF with $K_t = \frac{1}{\sigma_t}$ in Equations 3.26 and 3.27 in Chapter 3, the Hummel filter $H$ with $K_t = \frac{1}{\sigma_t}$ and the Unweighted filter $Un$ with $K_t = 1$. To reduce the dimensions of the comparison we define $\Sigma_t = \sigma_2/\sigma_1$ which is the ratio of the standard deviations for the two observations and $\Delta = E \left( \frac{M_t}{\sigma_2} \right)/E \left( \frac{M_t}{\sigma_1} \right)$
which is the ratio of the decision values between the two observations:

\[ R_{QF/H} \triangleq \log \frac{E(\delta)}{E(\delta)_H} = \log \frac{(1 + \Delta)\sqrt{1 + \Sigma^2}}{\sqrt{2}(\Sigma + \Delta)}, \]

\[ R_{QF/Un} \triangleq \log \frac{E(\delta)}{E(\delta)_{Un}} = \log \frac{(1 + \Delta)\sqrt{1 + \Sigma^2}}{\sqrt{2}(1 + \Sigma\Delta)}, \]

\[ R_{H/Un} \triangleq \log \frac{E(\delta)}{E(\delta)_{Un}} = \log \frac{(\Sigma + \Delta)}{(1 + \Sigma\Delta)}. \]

Figures 4.1, 4.2 and 4.3 show the value of \( R \) for various values of standard deviation ratios and decision ratios. By examining the contours for \( R_{QF/H} \) we can see that the Qualitative filter is significantly more successful when the magnitude of the uncertainty between in the observations has the same polarity and the relative decision strength of each observation. The Hummel filter is more successful when this polarity is reversed. However, as the spread of contour lines indicates, the relative effectiveness of the Hummel filter over the Qualitative filter when the Hummel is preferred is not as profound as the success of the Qualitative filter when the Hummel is not preferred. Similar arguments apply to the effectiveness of the Qualitative filter compared with the Unweighted filter (see graphs for \( R_{QF/Un} \)).

\[ \begin{align*}
\text{(a) } R_{QF/H} \text{ Surface} & \\
\text{(b) } R_{QF/H} \text{ Contours}
\end{align*} \]

**Figure 4.1:** Comparison of expected decision fused strength between Qualitative filter and Hummel filter for two observations: \( R = \log \frac{E(\delta)}{E(\delta)_H} \).

So, we can conclude that the most efficient QF estimator is the information form. The following definition generalises the information form of the qualitative filter to \( n \) observations.

**Definition 11** The general form of the Information filter for \( n \) observations is defined to be (using Biscay information \( I \)):

\[ I^n = K^{-1} \sum_{i=1}^{n} I_i \]
Figure 4.2: Comparison of expected decision fused strength between Qualitative filter and Unweighted filter for two observations: \( R = \log \frac{E(\Theta_{QF})}{E(\Theta_{Un})} \).

Figure 4.3: Comparison of expected decision fused strength between Hummel filter and Unweighted filter for two observations: \( R = \log \frac{E(\Theta_{HF})}{E(\Theta_{Un})} \).

where:

\[
K = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(I_i, I_j)}.
\]
Example: Combined Probabilistic and Statistical Fusion

The information filter’s ability to combine probabilistic and statistical information is demonstrated by a simple static problem.

An unlabelled urn is sensed by a visual sensor and a Geiger counter. This urn can be one of two $U_1$ and $U_2$ and the problem is to determine which one it is. Urn $U_1$ contains 80 red balls and 20 blue balls and the urn has been irradiated so that the balls emit a constant supply of x-rays of mean radiation 4 units and standard error 1 units. Urn $U_2$ contains 20 red balls and 80 blue balls each of which radiate 6 units on average with a standard error of 1 units. The visual sensor is colour sensitive and supplies and unbiased median estimate on the $\{0, 0.5, 1\}$ quantity-space depicting the fractional number of red balls observed. The Geiger counter is used to obtain a mean estimate on the quantity-space $\{-\infty, 5, \infty\}$ depicting the mean radiation strength of the observed balls. The probability of observing a red or blue ball from each urn is:

$$Pr(\text{red}|U_{rm}) = 0.8,$$
$$Pr(\text{red}|U_{r2}) = 0.2,$$
$$Pr(\text{blue}|U_{rm}) = 0.2,$$
$$Pr(\text{blue}|U_{r2}) = 0.8$$

and therefore, the variance of our median estimator (obtain using Equation 3.24) is:

$$Var\left(\log \frac{m_{\text{med}}(U_1|\text{red})}{m_{\text{med}}(U_2|\text{red})}\right) = 4(\log \alpha)^2 0.8(1 - 0.8) = 0.64(\log \alpha)^2.$$

Similarly, for blue ball observations:

$$Var\left(\log \frac{m_{\text{med}}(U_1|\text{blue})}{m_{\text{med}}(U_2|\text{blue})}\right) = 0.64(\log \alpha)^2.$$

The variance of the Geiger counter mean estimator is:

$$Var\left(\log \frac{m_{\text{mean}}(U_1|z)}{m_{\text{mean}}(U_2|z)}\right) = 16.$$

During an experiment, 6 balls were randomly selected in turn from an unknown urn (each ball was replaced after being sensed by both sensors):

<table>
<thead>
<tr>
<th>Obs</th>
<th>Vision sensor</th>
<th>Geiger counter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>red</td>
<td>4.3</td>
</tr>
<tr>
<td>2</td>
<td>blue</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>red</td>
<td>4.6</td>
</tr>
<tr>
<td>4</td>
<td>red</td>
<td>1.9</td>
</tr>
<tr>
<td>5</td>
<td>red</td>
<td>4.7</td>
</tr>
<tr>
<td>6</td>
<td>blue</td>
<td>2.3</td>
</tr>
</tbody>
</table>

The variance in the Geiger counter noise is conditionally independent of the visual probability
distribution given a particular urn and the information form of the QF yields:

\[ I_1^2 = \frac{1}{\sqrt{12}} \left( \sum_{i=1}^{6} (I_{\text{visual},i} + I_{\text{gigercounter},i}) \right) \]
\[ = \frac{1}{\sqrt{12}} \left( \frac{2}{\sqrt{0.64}} + \frac{8.4}{1} \right) \]
\[ = 3.32. \]

Thus, the unlabelled urn is actually Urn $U_1$.

In summary, although standard methods cannot be used to find an optimal filter, the information filter proposed in this section utilises, “on average”, more information than does the unweighted filter.

### 4.2.1 Symmetric Biscay Distributions: An Improved Filter

When the Biscay information is known to be symmetric about its expected value then the potency of the information filter can be improved. The Biscay information of symmetrically distributed information can be scaled by a factor $\sqrt{2}$ without causing the information filter to exceed the desired $\frac{1}{2}$ false-positive error rate. In this section we demonstrate this for independent observations.

Let $\Xi = \{z_1, \ldots, z_N\}$ be a set $N$ of observations and let $S$ and $A$ denote those observations obtained from symmetric and asymmetric distributions respectively: $\Xi = S \cup A$. For any two states each observation $z_i$ defines Biscay information $I_i$. Also define:

\[ (\forall z_i \in S) \; c_i = \sqrt{2}, \]
\[ (\forall z_i \in A) \; c_i = 1. \]

Consider the Biscay information for $N$ observations formed by assimilating the observations in $\Xi$ subject to the following weighted information filter:

\[ I_1^N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_i I_i \]  \hspace{1cm} (4.5)

where $\sqrt{N}$ is the usual information form normalisation constant for independent observations.

Let $\chi$ denote a set of $N_\chi$ trials and $C(j, V)$ represent the condition $\sum_{i=1}^{N} \frac{c_i (I_i - E(I_i))}{\sigma G} > V$ for trial $j \in \chi$. The probability of a false-positive is:

\[ Pr_\chi \left( \sum_{i=1}^{N} \frac{c_i (I_i - E(I_i))}{\sigma G} > 1 \right) \leq \frac{1}{N_\chi} \sum_{j \in \chi \cap C(j,1)} \left( \sum_{i=1}^{N} \frac{c_i (I_i(j) - E(I_i(j)))}{\sigma G} \right)^2 \]
\[ \leq \frac{1}{N_\chi} \sum_{j \in \chi \cap C(j,1)} \left[ \left( \sum_{i=1, i \in S} \sqrt{2} (I_i(j) - E(I_i(j))) \right)^2 + \left( \sum_{i=1, i \in A} \frac{I_i(j) - E(I_i(j))}{\sigma G} \right)^2 \right] + \text{cross terms.} \]
where the “cross-terms” involve pairs of observations from $S$ and $A$ and is a function of:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (I_i(j) - E(I_i(j))) \times (I_k(j) - E(I_k(j))).$$

These cross-terms are assumed to be zero for independent observations in the limit $N \to \infty$. Thus:

$$Pr_X \left( \frac{\sum_{i=1}^{N} c_i (I_i - E(I_i))}{\sigma G} > 1 \right) \leq \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_i(j) - E(I_i(j)))}{\sigma G} \right)^2 + \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_j(j) - E(I_j(j)))}{\sigma G} \right)^2 \leq \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_i(j) - E(I_i(j)))}{\sigma G} \right)^2 + \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_j(j) - E(I_j(j)))}{\sigma G} \right)^2 \leq \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_i(j) - E(I_i(j)))}{\sigma G} \right)^2 + \frac{1}{N_X} \sum_{j \in X, C(j, 0)} \left( \sum_{i=1}^{N} \frac{(I_j(j) - E(I_j(j)))}{\sigma G} \right)^2 \leq \frac{1}{G^2}.$$

since observations $i$ are assumed to be independent and $\sigma^2 = E_X (\sum_{i=1}^{N} (I_i(j) - E(I_i(j)))^2)$.

### 4.3 Conservative Estimators

It is often difficult to obtain, or even maintain, exact noise variance values (due to measurement uncertainties for example) and, in some circumstances, only a bound on the value of the variance is known (the median estimator variance for example). When observing a planar wall, the gyroscope standard deviation varied between 0.02 units for a smooth wall and 0.15 units for a rough wall. So although the mean gyroscope output is close to zero for both walls, its underlying noise statistics are significantly different for each wall and which type of wall is being examined may not be known apriori. So, we want to find a filter with which no irrational decisions are made using conservative values for the true variance measures. This leads us to the idea of a decision conservative estimator $I^*$ which conserves the desired upper bound on the false-positive error rate. In the following I will assume that the binary Biscay information measures $I$ and $I^*$ concern the true state $Q$ and some other state $Q'$ and are short for $I_{Q,Q'}$ and $I^*_{Q,Q'}$ respectively. 2

**Definition 12** An estimate $I^*$ is a decision conservative estimate of $I$ if the following conditions hold:

$$Pr(I^* > G \mid E(I) < 0) \leq \frac{1}{G^2}.$$

---

2Simply asserting that $|I^*_i| \leq |I_i|$ is an insufficient condition for decision conservation.
4.4 QF Problems

The following theorem sets out sufficient conditions for $I^*$ to be decision conservative.

**Theorem 12** $I^*$ is a decision conservative estimate of $I$ if:

1. $I^*$ is an unbiased estimate. That is, between the true state and any other state:

$$E(I^*) \geq 0 \iff E(I) \geq 0.$$ 

2. $\text{Var}(I^*) \leq 1$.

**Proof** By condition 1, $E(I) < 0 \Rightarrow E(I^*) < 0$ and:

$$Pr(I^* > G \mid E(I) < 0) = Pr(I^* > G \mid E(I^*) < 0)$$

and, provided condition 2 is true then, by the Chebyshev inequality:

$$Pr(I^* > G \mid E(I) < 0) \leq \frac{1}{G^2}.$$

$\Box$

So, a simple mean unbiased observation is decision conservative provided the standard error value is a conservative estimate of the true standard error (i.e. $\sigma^* \geq \sigma$).

We must now determine the conditions under which estimates inferred or produced from the assimilation of decision conservative estimates are themselves decision conservation. The Biscay information equations for inference and fusion are of the form:

$$I = k \sum_i I_i.$$

We want to find the conditions for $k$ which guarantees that $I$ is decision conservative. We assume that the input estimates $I_i$ obey the conditions in Theorem 12: namely, $E(I_i) \geq 0$ and $\text{Var}(I_i) \leq 1$.

Provided $k \geq 0$ then $I$ trivially obeys Theorem 12, Condition 1. As for Condition 2:

$$\text{Var}(I) = k^2 \text{Var} \left( \sum_i I_i \right).$$

So, $\text{Var}(I) \leq 1$ provided $k \geq \frac{1}{\sqrt{\text{Var}(\sum_i I_i)}}$. This inequality is satisfied provided the correlation estimate $\rho_{i,j}^*$ is no less than the true correlation $\rho_{i,j} = \text{Cov}(I_i, I_j)$.

**Theorem 13** The inference or fusion of a subset of decision conservative Biscay estimates $\{I_1, \ldots, I_n\}$ is itself decision conservative provided the estimated correlation value $\rho^*$ between pairs of Biscay values obeys:

$$(\forall 1 \leq i, j \leq n) \quad \rho_{i,j}^* \geq \rho_{i,j}.$$ 

4.4 QF Problems

Some undesirable properties of the QF framework are discussed in this section.
4.4 QF Problems

Consider a simple situation. A qualitative estimate of the true state in quantity-space $\Theta_2 = \{Q_3, Q_4, Q_5\}$ is made by aggregating in $\Theta_2$ estimates inferred from an observation quantity-space $\Theta_1 = \{Q_1, Q_2\}$. The inference model comprises the following transitions:

$$Q_1 \rightarrow \{Q_3, Q_4\}, \quad Q_2 \rightarrow \{Q_4, Q_5\}$$

All (compound) paths to $Q_3$ comprise the transitions $Q_1 \rightarrow Q_3$. Similarly, paths to $Q_5$ are made up from transitions $Q_2 \rightarrow Q_5$. Thus, for any number of observations there is only one (compound) path to each of $Q_3$ and $Q_5$. However, compound paths to $Q_4$ can comprise transitions $Q_1 \rightarrow Q_4$ or $Q_2 \rightarrow Q_4$ and for $N$ observations there are $2^N$ different compound paths that lead to $Q_4$. $Q_4$ is said to be fed by multiple-paths and is the source of difference between the Eulerian and the Lagrangian views of what it means for a filter to be unbiased. Naturally, as $N \rightarrow \infty$ then the number of paths to $Q_4$ also tends to $\infty$. Suppose that the true state is in $Q_5$ (so it is observed to be in $Q_2$) and a noise distribution generates observations in $Q_1$ and $Q_2$. Although each observation is unbiased so that $Q_2 \rightarrow Q_5$ is preferred to all other paths (including each path leading to $Q_4$), state $Q_5$ is never preferred by the QF!. This is our first paradox and it arises from the multiple-path problem.

The multiple-path problem is simply a consequence of our Lagrangian view of unbiased filtering. The multiple-path problem has degrees of severity in different situations and, in Section 4.5, we discuss non-trivial situations when multiple-path systems have identical Eulerian statistics to the Lagrangian view. The multiple-path problem is similar to many problems found in voting systems. Namely, that pair-wise voting can lead to unintuitive preference orderings.

\[\begin{array}{c|c|c|c}
1 & 2 & 3 \\
\hline
\ \ 1_1 & \ \ 2_2 & \ \ 3_z \\
\hline
A & B & C \\
\hline
\ \ 1_A & \ \ z_\prime & \ \ 1_B
\end{array}\]

Figure 4.4: Two observations, $z$ and $z'$, within quantity-spaces, $\{1, 2, 3\}$ and $\{A, B, C\}$, respectively.

Another paradox identified in the voting literature is the cyclic preference ordering problem. The cyclic preference ordering problem refers to when the aggregation of individual transitive preference orderings yields intransitive preference orderings. Simpson’s paradox is a famous example of this problem. For any two states $Q_1$ and $Q_2$, $Q_1 \succ Q_2$ indicates that $Q_1$ is strictly preferred to $Q_2$, $Q_1 \succeq Q_2$ that $Q_2$ is not preferred to $Q_1$ (i.e., $Q_1 \succeq Q_2 \iff \neg(Q_2 \succ Q_1)$) and $Q_1 \sim Q_2$ indicates that neither $Q_1$ nor $Q_2$ are preferred to each other (i.e., $Q_1 \sim Q_2 \iff [Q_1 \succeq Q_2] \wedge [Q_2 \succeq Q_1]$).

Pair-wise preferences between three alternatives $Q_1$, $Q_2$ and $Q_3$ with $Q_1 \succ Q_2$, $Q_2 \succ Q_3$ but $Q_3 \succ Q_1$ is an example of a cycle. If $I_{Q_1, Q_2}$ is the Biscay information between states $Q_1$ and $Q_2$ then $Q_1 \succeq Q_2 \iff I_{Q_1, Q_2} \geq 0$. Cycles can occur when Biscay information is aggregated using the information form of the QF. To demonstrate this, consider the aggregation of two three state frames, $\Theta_1 = \{1, 2, 3\}$ and $\Theta_2 = \{A, B, C\}$, with BPA arising from mean unbiased observations, $z$ and $z'$, respectively (see Figure 4.4):

\[\begin{align*}
I_{2,1} &= \frac{z - \mu_2}{\sigma}, & I_{3,2} &= \frac{z - \mu_3}{\sigma}; & I_{3,2} &= \frac{z - \mu_2}{\sigma}; \\
I_{B,A} &= \frac{z - \mu_B}{\sigma}, & I_{C,B} &= \frac{z - \mu_C}{\sigma}; & I_{C,A} &= \frac{z - \mu_C}{\sigma}.
\end{align*}\]
4.5 Lagrangian Decision Preserving Eulerian Estimators (SOPAP)

Suppose also that:

\[ z = 0, \ l_1 = -12, \ l_2 = -2, \ z' = 1, \ l_A = -10, \ l_B = 10, \ \sigma = 1. \]

Both frames define transitive preference orderings. A third quantity-space \( \Theta_3 = \{ X, Y, Z \} \) is formed by fusing \( \Theta_1 \) and \( \Theta_2 \) according to the \( \infty \)-norm Dempster Rule of Combination. Regions 2 and 4 are combined into region \( X \), regions \( B \) and 1 into region \( Y \) and regions \( C \) and 3 into region \( Z \). Assuming that the observation noise for \( z \) and \( z' \) are independent, the Biscay information within the new quantity-space is:

\[
I_{X,Y} = \frac{I_{A,B} + I_{A,C}}{\sqrt{2}} = 0.71, \\
I_{Y,Z} = \frac{I_{B,C} + I_{A,C}}{\sqrt{2}} = 1.41, \\
I_{X,Z} = \frac{I_{A,B} + I_{B,C}}{\sqrt{2}} = -2.12.
\]

Thus although \( I_{X,Y} > 0 \) and \( I_{Y,Z} > 0 \), \( I_{X,Z} < 0 \) and the preference ordering is intransitive.

Intransitivity of preference orderings is a common product of many pairwise voting algorithms and there is much literature on meta-theories for rational and democratic voting strategies which guarantee transitivity. In a famous monograph, Arrow (Arrow, 1976) considered the aggregate of individual preference orderings over alternatives. An individual may be seen in the social context to be an observation instance and the alternative is a qualitative state. Arrow proved that, under certain rationality and democratic conditions, the aggregate of individual preference orderings (called the social choice) for three or more alternatives must be either intransitive over the set of alternatives, irrational, imposed or dictatorial.

It is perhaps interesting to note that the unweighted QF (for which \( k_i = 1 \)) satisfies all of Arrow’s conditions simultaneously. Transitivity is preserved simply because inference and aggregation preserves the log mass ratio identity \( M_{a,b} = M_{a,c} + M_{c,b} \) for any states \( a, b \) and \( c \). The reason why the Impossibility Theorem does not hold for the unweighted form of the QF is that Arrow’s proof assumes that, for individuals \( x \) and \( y \), whenever \( x \succ_1 y \) and \( y \succ_2 x \) then either \( x \sim y \) or \( x \succ y \) or \( x \succ y \) at all times. In order to avoid dictatorship then it is necessary to assure that \( x \succ_1 y \land y \succ_2 x \supset x \sim y \). Strictly, the social welfare function is dictatorial whenever \( (3i,x,y) \succ x \supset x \succ y \). However, within the QF framework comparative mass assignments between individuals allows both \( x \succ_1 y \land y \succ_2 x \supset x \succ y \) for some of the time and \( x \succ_1 y \land y \succ_2 x \supset x \succ y \) for some of the time. This is a weakened form of dictatorship where one voter may dictate the outcome of a specific vote but not for each trial. Thus, by allowing a weakened form of dictatorship, Arrow’s Impossibility Theorem can be avoided.

In summary, intransitivity is not a problem of the framework, rather it arises only for specific instances (albeit desirable instances). When cycles are detected one or more input estimates are discarded so that the remainder are cycle free when fused. Which estimates to discard is determined randomly and independent of their values.

4.5 Lagrangian Decision Preserving Eulerian Estimators (SOPAP)

The KF is guaranteed to be unbiased for linear process and observation models. This guarantee does not hold for non-linear models. Similarly, the QF can be shown to be unbiased and exhibit Lagrangian bounded Eulerian false-positive error rates under certain domain modelling conditions, namely the condition of state order preservation along paths (SOPAP).
4.5 Lagrangian Decision Preserving Eulerian Estimators (SOPAP)

**Definition 13** A filter is stable when its Eulerian false-positive error-rate is bounded above by the Lagrangian false-positive error-rate.

Firstly, a filter operating on a single observation is shown to be stable and then we will define the SOPAP condition and show that a filter operating on a SOPAP system is also stable.

<table>
<thead>
<tr>
<th>......</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>......</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4</td>
<td>1.3</td>
<td>1.2</td>
<td>1.1</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

**Figure 4.5:** Quantity-space with landmarks ordered $l_{-4} < \ldots < l_3$ and regions ordered $-3 < \ldots < 2$.

Regions and landmarks in each quantity-space are labelled thus: the state containing the true state is always labelled 0; landmarks and regions to the right are labelled successively by increasing integers starting from 1 and those landmarks and regions to the left are labelled by decreasing integers starting at -1 (see Figure 4.5). We define $l_0 = 0$ and emphasize that $l_0$ need not be a landmark. Its inclusion improves the readability of the following analysis.

**Theorem 14** Let $Q, Q'$ and 0 be regions such that the true state is in 0. Then the mean and median estimator BPAs obey the following mass preference ordering relationships:

\[
Q \preceq Q' \prec 0 \quad \Rightarrow \quad I_{Q,0} \geq I_{Q',0},
\]

\[
0 \prec Q' \preceq Q \quad \Rightarrow \quad I_{Q',0} \geq I_{Q,0},
\]

\[
Q' \prec 0 \preceq Q \quad \Rightarrow \quad I_{Q',0} \leq -I_{Q,0}.
\]

**Proof** We prove these results for both mean estimator and the median estimator BPAs separately.

**Mean estimator**

Define $M_x = \frac{\sum_{i=0}^x l_i}{|x|}$ and assume $\text{sign}(x) = \text{sign}(n)$. Then:

\[
M_x - M_{x+n} = \frac{|n| \sum_{i=0}^x l_i - |x| \sum_{i=x+1}^{x+n} l_i}{|x(x+n)|}.
\]

Now, since:

\[
|x| \min(l_x, l_0) \leq \sum_{i=0}^x l_i \leq |x| \max(l_x, l_0)
\]

and:

\[
|n| \max(l_{x+1}, l_{x+n}) \geq \sum_{i=x+1}^{x+n} l_i \geq |n| \min(l_{x+1}, l_{x+n})
\]

then:

\[
M_x - M_{x+n} \begin{cases} 
\leq 0 & \text{if } x \geq 0 \text{ and } n \geq 0, \\
\geq 0 & \text{if } x \leq 0 \text{ and } n \leq 0.
\end{cases}
\]
When landmarks $x'$ and $x$ are ordered $x' < 0 < x$:

$$M_x - M_{x'} = \frac{\sum_{i=0}^{x'} l_i}{|x|} - \frac{\sum_{i=0}^{x} l_i}{|x|} \geq l_0 - l_0 = 0.$$ 

Now, for the mean estimator we have:

$$I_{x',0}(z) = \begin{cases} \frac{z - M_x}{\sigma} & \text{if } X > 0, \\ \frac{z - M_x}{\sigma} & \text{if } X < 0. \end{cases}$$

Thus:

$$Q \leq Q' < 0 \implies M_{Q'} - M_Q \geq 0 \implies I_{Q',0} \geq I_{Q,0},$$

$$0 < Q' \leq Q \implies M_{Q'} - M_Q \geq 0 \implies I_{Q',0} \geq I_{Q,0}.$$ 

When $Q' < 0$ and $Q > 0$ then $M_Q - M_{Q'} \geq 0$. Thus:

$$I_{Q',0} = \frac{M_{Q'} - z}{\sigma} = -\frac{z - M_Q}{\sigma} + \frac{M_{Q'} - M_Q}{\sigma} \leq -\frac{z - M_Q}{\sigma} = -I_{Q,0}.$$ 

Therefore:

$$Q' < 0 < Q \implies I_{Q',0} \leq -I_{Q,0}.$$ 

**median estimator**

When $z \in Q_z$ then by inspection:

$$I_{Q_z,0}(z) \begin{cases} = +1 & \text{if } (Q_z \leq Q < 0) \lor (0 < Q \leq Q_z), \\ = -1 & \text{if } (Q < 0 < Q_z) \lor (Q_z < 0 < Q), \\ \in \{1,0,-1\} & \text{if } (Q \leq Q_z < 0) \lor (0 < Q_z \leq Q) \end{cases} \tag{4.9}$$

Equations 4.6 and 4.7 follow immediately from Equation 4.9. Equation 4.8 follows by reasoning about the following cases. Either $(Q_z \leq Q < 0)$ or $(0 < Q \leq Q_z)$ in which case $I_{Q,0} = 1$ by Equation 4.9. Since $Q' < 0 < Q$ then $(Q' \times 0 < Q_z)$ and, by Equation 4.9, $I_{Q',0} = 1 = -I_{Q,0}$. Or $(Q \leq Q_z < 0)$ or $(0 < Q_z \leq Q)$ and, since $Q' \times 0 < Q$ then $Q' < 0 < Q_z$. Thus, by Equation 4.9, $I_{Q',0} = -1 \leq -I_{Q,0}$ for all $I_{Q,0} \in \{1,0,-1\}$. Or $Q_z$ and $Q$ are on opposite sides of 0 so that $(Q < 0 < Q_z)$ or $(Q_z < 0 < Q)$ in which case $I_{Q,0} = -1$. Thus, by Equation 4.9, $I_{Q',0} \in \{1,0,-1\}$ and $I_{Q',0} \leq 1 = -I_{Q,0}$. 

□

Using Theorem 14, we now show that the QF is stable.

**Corollary 1** For a single observation, the QF is stable.
4.5 Lagrangian DecisionPreserving Eulerian Estimators (SOPAP)

**Proof** The proof focuses on state $-1$ when the true state is in 0. The false-positive error rate is bounded $Pr(|I_{-1,0}| > G) \leq \frac{1}{G}$. Therefore, since for any state $Q$, either $Q \preceq -1 \preceq 0$ or $-1 \preceq 0 \preceq Q$, it follows by Theorem 14 for all $Q$, $I_{Q,0} > G \supset |I_{-1,0}| > G$. Thus:

$$Pr(\exists Q : I_{Q,0} > G) \leq Pr(|I_{-1,0}| > G) \leq \frac{1}{G^2}.$$ 

$\square$

Combinations of observations are not necessarily stable. We present a sufficient condition for filter stability: State Order Preservation Along Paths (SOPAP).

**Definition 14** The ordered binary predicate $SOPAP(\varphi, \psi')$ holds between two paths $\varphi$ and $\psi'$ when for states $Q \in \varphi$ and $Q' \in \psi'$ on these paths either:

$$(\forall Q \in \varphi, Q' \in \psi') (Q \preceq Q' \preceq 0 \lor 0 \preceq Q' \preceq Q).$$

(4.10)

or:

$$(\forall Q \in \varphi, Q' \in \psi') (Q' \preceq 0 \preceq Q).$$

(4.11)

A set of paths $\{\varphi\}$ is Entirely SOPAP when:

$$(\forall \varphi, \psi') (SOPAP(\varphi, \psi') \lor SOPAP(\psi', \varphi)).$$

Trivially, the quantity-space for an observation is entirely SOPAP. Two properties of SOPAP follow immediately: reflexivity and transitivity.

**Theorem 15**

- **SOPAP is Reflexive.** $SOPAP(\varphi, \varphi)$.
- **SOPAP is Transitive.** $SOPAP(\varphi, \psi') \land SOPAP(\psi', \psi'') \supset SOPAP(\varphi, \psi'')$.

**Proof** To prove reflexivity let $\psi' \equiv \varphi$. Then:

$$(\forall Q \in \varphi, Q' \in \psi') (Q \preceq Q' \preceq 0 \lor 0 \preceq Q' \preceq Q),$$

which satisfies Condition 4.10 in Definition 14.

To prove transitivity we proceed by constructing a table of consistent orderings of states $A \in \varphi$, $B \in \psi'$ and $C \in \psi''$ about 0 according to Definition 14. Writing $A \succ B \succ C$ as $ABC$, SOPAP($\varphi, \psi'$) implies that for each state $A \in \varphi$ and $B \in \psi'$ either $AB0$, $0BA$ or $B0A$. Similarly, SOPAP($\psi', \psi''$) implies that for each state $B \in \psi'$ and $C \in \psi''$ either $BC0$, $0CB$ or $C0B$. The following table shows the consistent state orderings of $A$, $B$ and $C$ about 0 when SOPAP($\varphi, \psi'$) and SOPAP($\psi', \psi''$) ("-" indicates no consistent ordering):

<table>
<thead>
<tr>
<th></th>
<th>AB0</th>
<th>0BA</th>
<th>B0A</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0CB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C0B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thus, either \((\forall A \in \varphi, C \in \varphi') (AC0 \lor OCA)\) or \((\forall A \in \varphi, C \in \varphi') C0A\). These are the conditions for SOPAP in Definition 14.

\(\square\)

**Theorem 16** If SOPAP is entire over a set of paths \(\mathcal{P}\) then:

\[
(\exists \varphi' \in \mathcal{P})(\forall \varphi' \in \mathcal{P}) \text{ SOPAP}(\varphi', \varphi).
\]

**Proof** The proof proceeds by induction over subsets of paths in \(\mathcal{P}\). For a single path \(\varphi\), reflexivity gives SOPAP(\(\varphi, \varphi\)). For the inductive part suppose for some \(\mathcal{W} \subseteq \mathcal{P}\) the theorem holds. Then:

\[
(\exists \varphi \in \mathcal{W})(\forall \varphi' \in \mathcal{W}) \text{ SOPAP}(\varphi', \varphi).
\]

Consider \(\varphi'' \in \mathcal{P} \setminus \mathcal{W}\) then, since SOPAP is entire over \(\mathcal{P}\) then, either SOPAP(\(\varphi'', \varphi\)) in which case it immediately follows that:

\[
(\forall \varphi' \in \mathcal{W} \cup \{\varphi''\}) \text{ SOPAP}(\varphi', \varphi)
\]

or, SOPAP(\(\varphi, \varphi''\)) in which case by transitivity \((\forall \varphi' \in \mathcal{W}) \text{ SOPAP}(\varphi', \varphi'')\) and:

\[
(\forall \varphi' \in \mathcal{W} \cup \{\varphi''\}) \text{ SOPAP}(\varphi', \varphi'').
\]

\(\square\)

So, when a set of paths is entirely SOPAP then, by Definition 14 and Theorem 14, either:

\[
(\forall \varphi' \in \varphi, C' \in \varphi') (I_{\varphi', 0}(z) \geq I_{\varphi, 0}(z))
\]

or:

\[
(\forall \varphi' \in \varphi, C' \in \varphi') (I_{\varphi', 0}(z) \leq -I_{\varphi, 0}(z))
\]

The Biscay information assigned to path \(\varphi'\) and any other path \(\varphi\) is:

\[
I_{\varphi', 0}(z_1) = R \sum_{i=1}^{t} I_{\varphi'(i), 0}(z_i)
\]

\[
I_{\varphi, 0}(z_1) = R \sum_{i=1}^{t} I_{\varphi(i), 0}(z_i).
\]

respectively and therefore, for our SOPAP system either \(I_{\varphi'(z_1), 0} \leq -I_{\varphi, 0}(z_1)\) or \(I_{\varphi'(z_1), 0} \geq I_{\varphi, 0}(z_1)\). Thus, \((\exists \varphi'')(\forall \varphi')(I_{\varphi', 0} \geq G \supset |I_{\varphi', 0}| > G)\). For any decision threshold \(G\), when an erroneous decision is drawn there is some path \(\varphi\) such that \(I_{\varphi(z_1), 0} > G\). However, no matter which path \(\varphi\) causes this decision one particular path (\(\varphi\)) will have Biscay information \(|I_{\varphi, 0}| > G\) each time. The Chebyshev Inequality tells us that the probability of \(|I_{\varphi, 0}| > G\) is less than \(\frac{1}{G^2}\). Thus, the Eulerian false-positive error-rate is less than the Lagrangian false-positive error-rate \(\frac{1}{G^2}\) for the individual path \(\varphi\).

In summary, mean and median observation BPAs are entirely SOPAP and the fusion of entirely SOPAP BPAs produces an entirely SOPAP BPA. Inference is also entirely SOPAP provided Condition 4.10 or 4.11 in Definition 14 is maintained. The Eulerian false-positive error-rate is bounded above by \(\frac{1}{G^2}\) when SOPAP is entire.
4.6 Conclusions

This chapter investigates a number of topics within the QF framework: efficiency, sensitivity and robustness. The information form, based on inverse standard error weighting, was found to be the most efficient form as it exhibits the best convergence rates "on average" and also preserves decisive measurements. Occasionally, the information form induces ordering cycles between states and, to prevent their formation, a method of discounting was proposed.

It was shown how the filtering framework can accommodate decision conservative estimates of covariance measures with a degradation of the rate of convergence but not of the accuracy of the estimate. It was also shown how, under certain conditions (i.e. SOPAP), the Lagrangian approximation to the underlying Eulerian filtering task is decision conservative. Provan (Pradhan et al., 1996) andDecoste (DeCoste, 1990) have observed that belief network structures are insensitive to imprecise probabilities. Preliminary experiments indicate that the QF may be insensitive to many non-SOPAP structures. For example, insignificant bias is exhibited by the information form of the QF applied to the three state multiplication problem $Q_{\text{mag}}(X) \otimes Q_{\text{mag}}(Y)$ for Gaussian noise where $X$ and $Y$ have identical quantity-spaces $\{(-\infty, -1), (-1, 1), (1, \infty)\}$. The Biscay information bias when $x = y = 1$ is only $E(I_{+,0}) = 0.03 \pm 0.01$ and $E(I_{+,0}) = 0.19 \pm 0.01$ for noise standard-errors 0.5 and 5 respectively.

To complete the theoretical development of the QF framework, the next chapter extends the QF framework to accommodate semi-quantitative domain information. Biscay information is used to generate a filter which approximates to the KF when uncertainty dominates imprecision and to the IIF when imprecision dominates.
Chapter 5

Qualitative Filtering and
Semi-Quantitative Representations

5.1 Introduction

Semi-quantitative reasoning involves combining incomplete quantitative and qualitative knowledge and three semi-quantitative methods, Q2, Q3 and NSIM, have been proposed for process simulation. All three methods use interval arithmetic; Q2 (Kuipers and Berleant, 1988) uses the mean value theorem and known interval endpoint landmark values to bound the values of intermediate landmarks. It also uses static envelopes to bound the range of possible monotonic functions; Q3 (Berleant, 1991) extends Q2 with step-size refinement, adaptively inserting landmarks into qualitative intervals to decrease the effective step-size and strengthen inference; NSIM (Kay and Kuipers, 1993) uses an extended Runge-Kutta method to simulate the numerical solutions to an extremal system of ODEs (the so called dynamic envelopes). NSIM generally produces tighter bounds than either Q2 or Q3 but can be numerically less stable. In this chapter, we extend the concept of the Biscay information developed in Chapter 3 to filtering using semi-quantitative information. The approach is applied to an NSIM formulation of the Ballistic Missile Problem.

5.2 The Dynamic Landmark Filter (DLQF)

So far, we have explored inference and fusion on quantity-spaces which are apriori partitioned. Often modelling information exists in the form of static envelopes which can be used to generate landmarks appropriate to the problem at hand. For example, when the parameter $Y$ is inferred from the observable $X$ via the imprecise mapping $y \in (x-1, x+1)$. In this case, landmarks at $x - 1 - 2\sigma$ and $x + 1 + 2\sigma$ define the tightest estimate of $Y$. This chapter shows show how the QF framework can be extended from mass estimation to parameter estimation using dynamically assigned landmarks. The Kalman filter is a parameter value estimator and so, we have come full circle in our development of qualitative model-based estimation techniques. Initially, the Kalman filter parameter estimation methodology inspired a mass estimator for quantity-space hypothesis testing. The DLQF extends the mass estimator methodology to dynamically generated landmarks and ultimately a Kalman filter-type parameter estimator for filtering when models are imprecise.
5.2.1 The Binary Biscay Distribution

So far the parameter space partition has been specified apriori in order to assign mass to partitions of possible parameter values. In cases when landmarks can be generated dynamically, corresponding values between initial quantity-spaces may not coincide. Further, if a decision is to be made, the estimate landmarks may not coincide with the query quantity-space. Ideally, the Biscay information for all quantity-spaces would be known and those quantity-spaces with appropriate corresponding values will be chosen for fusion. However, the cost of maintaining such a corpus of quantity-space Biscay measures is prohibitive. Alternatively, we introduce the idea of the binary Biscay Distribution which is the measure of Biscay information between states for all possible binary problems.

**Definition 15** For any landmark l dividing a parameter space into \( Q = (-\infty, l) \) and \( -Q = (l, \infty) \), the Biscay Distribution for the observation \( z \) is the function:

\[
I(l, z) \triangleq I_{Q, -Q}(z).
\]

Figure 5.1 shows the Biscay distribution \( I(l, z) \) for the mean estimator of \( X \) and, the inferred distribution for \( Y \) where \( Y = X + (-10, 10) \).

![Figure 5.1: Biscay information \( I(x, z) \) and \( I(y, z) \) for an observation of \( x = 0 \) and subsequent inference of \( y = x + (-10, 10) \).](image)

To build an approximate Biscay distribution from a given quantity-space we use a process analogous to “defuzzification” in Fuzzy Logic Theory (Kosko, 1992) and call this process *Quantification*. Given a BPA defined on a quantity-space \( \{l_1, \ldots, l_m\} \), at each landmark defined in the quantity-space construct a coarse binary state BPA and determine the binary Biscay information for each pair. The set of landmarks and their binary Biscay information are points on the Biscay distribution and thus, the given quantity-space yields a discrete snapshot of the Biscay distribution \( I(x, z) \). We can, from this information, find an approximation to the actual Biscay distribution which would have been obtained if we had used a differently partitioned quantity-space. We assume that \( I(x, z) \) is piece-wise analytic so that we can obtain a Taylor expansion for \( I(x, z) \) about an arbitrary point.
\[ I(x, z) = I(x_0, z) + I'(x_0, z)(x - x_0) + \cdots \circ (I''(x_0, z)). \] (5.1)

Since \( I(l_i, z) \) and \( I(l_{i+1}, z) \) are known then, solving Equation 5.1 for \( I(x_0, z) \) and \( I'(x_0, z) \) at \( l_i \) and \( l_{i+1} \) we obtain:

\[ (\forall l_i \leq x \leq l_{i+1}) \quad I(x, z) \approx \frac{(x - l_{i+1})I(l_{i+1}, z) + (l_i - x)I(l_i, z)}{l_i - l_{i+1}}. \] (5.2)

The Biscay distribution is a monotonically increasing function of the parameter value and Equation 5.2 is a piece-wise linear approximation of the Biscay distribution.

The DLQF uses the mean estimator Biscay distribution. A crucial property of the mean estimator BPA is that for at any landmark \( l \) the difference between the Biscay distribution at \( l \) and the expected Biscay distribution at \( l \) is independent of \( l \). It is a function of the observation \( z \), its standard error \( \sigma \) and the true state \( x_T \) only. By Theorems 5 and 9:

\[ I(l, z) - E(I(l, z)) = \frac{z - x_T}{\sigma}. \]

Since \( I(l, z) - E(I(l, z)) \) is independent of \( l \) then the parameter space may be partitioned dynamically either dependent or independent of the observation and still preserve the binary decision property:

\[ \Pr(I(l, z) > G \mid E(I(l, z)) < 0) \leq \frac{1}{G^2}. \]

**Definition 16** The \( v^\text{th} \) fractile \( J_v = (l, u) \) is an interval in the parameter-space defined so that the Biscay distribution values at \( l \) and \( u \) are \( I(l) = -v \) and \( I(u) = v \). The fractile width \( W = u - l \) and the fractile-value is \( v \).

The \( v^\text{th} \) fractile encloses the parameter true-state with a guaranteed reliability.

**Theorem 17**

\[ \Pr(x_T \in J_v) \geq 1 - \frac{1}{v^2}. \]

**Proof**

\[ \Pr(x_T \in J_v) = 1 - \Pr(x_T \notin J_v) = 1 - \Pr(\{I(x_T, z) > v\}). \]

By the Chebyshev Inequality:

\[ \Pr(\{I(x_T, z) > v\}) \leq \frac{E(I(x_T, z)^2)}{v^2}. \]

The DLQF BPA is unbiased and at the true state \( E(I(x_T, z)) = 0 \). Thus:

\[ \Pr(x_T \in J_v) \geq 1 - \frac{1}{v^2}. \]

\[ \square \]

The fractile is a state within a quantity-space containing landmarks \( l \) and \( u \). The fractile-value is the Biscay information between the fractile state and all states outside the fractile:

\[ I_{J_v, J_u}(z) = v. \]

Using quantification we can find the fractile for any desirable degree of estimate reliability. Further, the DLQF uses fractile information to produce optimal inferred and fused estimates.
5.2.2 Optimal Fusion

Two DLQF estimates are combined by fusing their corresponding Biscay distributions. This is to enable the filter to maintain a continuous estimate of desired decision intervals. For any two landmark values $x_1$ and $x_2$ we have, for a linear Biscay distribution and arbitrary point $x_0$:\footnote{Explicit reference to the observation $z$ is not made in the expression for Biscay information where ambiguity will not arise.}

\begin{align*}
I(x_1) &= I(x_0) + (x_1 - x_0)I'(x_0), \quad (5.3) \\
I(x_2) &= I(x_0) + (x_2 - x_0)I'(x_0). \quad (5.4)
\end{align*}

The $I^{th}$ fractile $J_I = (x_1, x_2)$ has an associated width $W = x_2 - x_1$ and $-I(x_1) = I(x_2) = I$ (see Figure 5.2). Subtracting Equations 5.4 from Equation 5.3 and substituting $-I(x_1) = I(x_2) = I$:

\[2I = I(x_2) - I(x_1) = (x_2 - x_1)I'(x_0) = W I'(x_0).\]

Therefore, the width of the $I^{th}$ fractile $J_I$ is:

\[W = \frac{2I}{I'(x_0)}. \quad (5.5)\]

The Biscay distribution $I(x)$ is the value of the binary state Biscay information between states separated by landmark $x$ and therefore two Biscay distributions may be combined using the weighted Dempster Rule of Combination:

\[I(x) = k_1 I_1(x) + k_2 I_2(x). \quad (5.6)\]

where:

\[k_1^2 + k_2^2 + 2\rho k_1 k_2 = 1 \quad (5.7)\]

and $\rho = \text{Cov}(I_1, I_2)$. Optimal values for $k_1$ and $k_2$ minimise the width $W$ of the fused estimate $J_I$. However, a simple relationship between $I$ and $W$ exists only under certain conditions:
5.2 The Dynamic Landmark Filter (DLQF) 126

- No imprecision.
- \( I \) is a linear function of \( x \). This constraint requires that inference is a linear process.

We will find values for \( k_1 \) and \( k_2 \) which are optimal under these specific conditions.

Applying Equation 5.6 to two landmarks \( x_1 \) and \( x_2 \):

\[
I(x_2) - I(x_1) = k_1[I_1(x_2) - I_1(x_1)] + k_2[I_2(x_2) - I_2(x_1)]
\]

\[
= [k_1I'_1 + k_2I'_2](x_2 - x_1)
\]

where \( I'_1 \) and \( I'_2 \) are the gradients of the input Biscay distributions. When \( x_1 \) and \( x_2 \) are chosen to bound the \( I^{th} \) fractile \( J_I \) of width \( W = x_2 - x_1 \), then:

\[
2I = I(x_2) - I(x_1)
\]

\[
= [k_1I'_1 + k_2I'_2]W. \tag{5.8}
\]

When \( W_1 \) and \( W_2 \) are the widths of the two input estimate fractiles \( J_I \) then by Equation 5.5:

\[
I'_1 = \frac{2I}{W_1}, \quad I'_2 = \frac{2I}{W_2}.
\]

Condition 5.2.2 implies that \( W_1, W_2 \) are independent of the observation \( z \) and, therefore, \( k_1 \) and \( k_2 \) can be functions of \( W_1 \) and \( W_2 \). Substituting for \( I'_1 \) and \( I'_2 \) in Equation 5.8 we obtain:

\[
W = \left( \frac{k_1}{W_1} + \frac{k_2}{W_2} \right)^{-1}. \tag{5.9}
\]

The optimal value of \( k_1 \) is that which minimises the width of the fused estimate. Substituting for \( k_2 \) from Equation 5.7 and differentiating by \( k_1 \):

\[
\frac{dW}{dk_1} \propto \left[\frac{1}{W_1} + \frac{1}{W_2} \left(-\rho \pm \frac{k_1[\rho^2 - 1]}{\sqrt{k_1^2[\rho^2 - 1] + 1}} \right) \right] = 0
\]

which is satisfied by:

\[
k_1^2[\rho^2 - 1] = \left( \frac{W_2}{W_1} - \rho \right)^2 \left( k_1^2[\rho^2 - 1] + 1 \right).
\]

Solving for \( k_1 \):

\[
k_1 = \frac{W_2 - \rho W_1}{\sqrt{(1 - \rho^2)X}} \tag{5.10}
\]

where \( X = W_1^2 + W_2^2 - 2\rho W_1 W_2 \). Substituting for \( k_1 \) in Equation 5.7:

\[
k_2 = \frac{W_1 - \rho W_2}{\sqrt{(1 - \rho^2)X}} \tag{5.11}
\]

The values for \( k_1 \) and \( k_2 \) are bounded below by 0 to ensure that \( I \) is an unbiased estimate if both \( I_1 \) and \( I_2 \) are unbiased estimates. Thus:

\[
2\text{Both } W_2 - \rho W_1 \text{ and } W_1 - \rho W_2 \text{ cannot be negative simultaneously. When } W_2 - \rho W_1 < 0 \text{ then } W_1 - \rho W_2 > W_1 - \rho^2 W_1 > 0.
\]

\[
\text{Thus: } \tag{5.10}
\]

\[
\text{Thus: } \tag{5.11}
\]

\[
\text{Thus: } \tag{5.10}
\]

\[
\text{Thus: } \tag{5.11}
\]
\[ k_1 = \begin{cases} 
1 & \iff W_1 - \rho W_2 < 0, \\
\frac{W_2 - \rho W_1}{\sqrt{1-\rho^2}X} & \iff (W_2 - \rho W_1)(W_1 - \rho W_2) \geq 0, \\
0 & \iff W_2 - \rho W_1 < 0.
\end{cases} \]

\[ k_2 = \begin{cases} 
1 & \iff W_2 - \rho W_1 < 0, \\
\frac{W_1 - \rho W_2}{\sqrt{1-\rho^2}X} & \iff (W_2 - \rho W_1)(W_1 - \rho W_2) \geq 0, \\
0 & \iff W_1 - \rho W_2 < 0.
\end{cases} \]

Substituting for \( k_1 \) and \( k_2 \) in Equation 5.6:

\[ I(x) = \frac{(W_2 - \rho W_1)I_{\alpha}(x) + (W_1 - \rho W_2)I_{\beta}(x)}{\sqrt{(1-\rho^2)X}}. \]  

(5.12)

Equation 5.12 is optimal in the sense that the width of the fused estimate is minimised when the uncertainty in the input estimates dominates imprecision. However, this does not inhibit its use when imprecision is not negligible. Also, \( k_1 \) and \( k_2 \) are functions of \( W_1 \) and \( W_2 \). In cases when \( I(x) \) is a non-linear function of \( x \), \( W_1 \) and \( W_2 \) will not be independent of \( z \) and the filter is not provably unbiased in such circumstances.

**Fusion Covariance Update**

Equation 5.12 is used recursively to update the correlation coefficient \( \rho \) when two parameters, \( \alpha \) and \( \beta \), say, are combined. Suppose, without loss of generality, that the \( \alpha \) estimate is constructed from two estimates, \( \alpha_1 \) and \( \alpha_2 \). Similarly for the \( \beta \) estimate:

\[ I_{\alpha}(x) = \frac{(W_{\alpha_1} - \rho_{\alpha_1,\alpha_2}W_{\alpha_2})I_{\alpha_1}(x) + (W_{\alpha_2} - \rho_{\alpha_1,\alpha_2}W_{\alpha_1})I_{\alpha_2}(x)}{\sqrt{(1-\rho_{\alpha_1,\alpha_2}^2)X_{\alpha}}}, \]

\[ I_{\beta}(x) = \frac{(W_{\beta_1} - \rho_{\beta_1,\beta_2}W_{\beta_2})I_{\beta_1}(x) + (W_{\beta_2} - \rho_{\beta_1,\beta_2}W_{\beta_1})I_{\beta_2}(x)}{\sqrt{(1-\rho_{\beta_1,\beta_2}^2)X_{\beta}}}, \]

where:

\[ X_{\alpha} = W_{\alpha_1}^2 + W_{\alpha_2}^2 - 2\rho_{\alpha_1,\alpha_2}W_{\alpha_1}W_{\alpha_2}, \]

\[ X_{\beta} = W_{\beta_1}^2 + W_{\beta_2}^2 - 2\rho_{\beta_1,\beta_2}W_{\beta_1}W_{\beta_2}. \]

If we assume, as we did for the derivation of the KF, that consecutive observations are independent, then the estimates indexed 1 and 2 are independent and \( \rho_{\alpha_1,\alpha_2} = 0 \) and \( \rho_{\beta_1,\beta_2} = 0 \). Taking the covariance of \( I_{\alpha}(x) \) with \( I_{\beta}(x) \) under conditions of independence:

\[ \rho_{\alpha,\beta} = \frac{W_{\alpha_1}\rho_{\alpha_1,\beta_1}W_{\beta_1} + W_{\alpha_2}\rho_{\alpha_2,\beta_2}W_{\beta_2}}{\sqrt{(W_{\alpha_1}^2 + W_{\alpha_2}^2)(W_{\beta_1}^2 + W_{\beta_2}^2)}}. \]  

(5.13)
5.2.3 Inference

Inference is all about constructing a Biscay distribution in the consequent parameter space given distributions in the antecedent parameter spaces. Inference may be non-monotonic and this necessitates drawing inferences from sets of fractiles defined on each antecedent parameter space. By definition, different fractiles are nested and drawing inferences from identically valued fractiles must produce nested fractiles in the consequent parameter space. From nested fractiles it is possible to reconstruct the Biscay distribution by quantification.

Suppose we wish to draw an inference from \( p \) antecedent parameter values. Firstly, \( n \) fractiles with fractile-values \( v_1, \ldots, v_n \) are defined on each antecedent parameter space. The fractiles for parameter \( i \) are denoted \( \beta_i = \{ J_i, v_1, \ldots, J_i, v_n \} \). Then, the interval arithmetic semi-quantitative model \( f \) is used to infer fractiles in the consequent parameter space. Define \( \text{Args} \) to be all possible combinations of fractile arguments to \( f \):

\[
\text{Args} = \beta_1 \times \beta_2 \times \cdots \times \beta_p.
\]

and, for any member \( A \in \text{Args} \), let \( J_i(A) \) denote the fractile for parameter \( i \) \( (1 \leq i \leq p) \). Then the \( i^{th} \) fractile in the consequent parameter space is the union of inferred intervals with Biscay information \( \nu \):

\[
J_i = \bigcup_{A \in \text{Args}, f(A) = \nu} f(A)
\]

where:

\[
I_{f(A) \rightarrow f(A)} = \frac{\sum_{i=1}^p I_{J_i(A)} - J_i(A)}{\sqrt{\sum_{i=1}^p \sum_{j=1}^p \rho_{i,j}}}
\]

**Inference Covariance Update**

Values for the inferred correlation values are obtained recursively using knowledge of the correlations between antecedent parameter BPAs. Suppose the value of parameters \( \alpha \) and \( \beta \) are inferred from \( n_\alpha \) parameters with Biscay distributions \( I_{\alpha,i} \) \( (1 \leq i \leq n_\alpha) \) and \( n_\beta \) parameters with Biscay distributions \( I_{\beta,i} \) \( (1 \leq i \leq n_\beta) \) respectively. The Biscay distributions for the inferred parameters are:

\[
I_{\alpha}(x) = \frac{\sum_{i=1}^{n_\alpha} I_{\alpha,i}(x_i)}{\sqrt{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\alpha} \rho_{\alpha,i,j}}} \quad \text{and} \quad I_{\beta}(x) = \frac{\sum_{i=1}^{n_\beta} I_{\beta,i}(x_i)}{\sqrt{\sum_{i=1}^{n_\beta} \sum_{j=1}^{n_\beta} \rho_{\beta,i,j}}}
\]

where \( \rho_{\alpha,i,j} = \text{Cov}(I_{\alpha,i}, I_{\alpha,j}) \) and \( \rho_{\beta,i,j} = \text{Cov}(I_{\beta,i}, I_{\beta,j}) \). The correlation coefficient between \( I_{\alpha} \) and \( I_{\beta} \) is therefore:

\[
\rho_{\alpha,\beta} = \text{Cov} \left( \frac{\sum_{i=1}^{n_\alpha} I_{\alpha,i}}{\sqrt{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\alpha} \rho_{\alpha,i,j}}}, \frac{\sum_{i=1}^{n_\beta} I_{\beta,i}}{\sqrt{\sum_{i=1}^{n_\beta} \sum_{j=1}^{n_\beta} \rho_{\beta,i,j}}} \right)
\]

and, therefore

\[3\]To reduce the computational cost of inference in practice, inferences are drawn from identically valued fractiles only:

\[(\forall m \in [1, n]) \ J_{\alpha,m} = f(J_{\alpha,1}, \ldots, J_{\alpha,m}).\]

Inferences limited to identically valued fractiles is an innate property of the IIF. A detailed investigation of the consequences of this decision must await further research.
\[ \rho_{\alpha \beta} = \frac{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{ij} \pm \rho_{kij}}{\sqrt{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{ij}^2 \cdot \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{kij}^2}} \quad (5.14) \]

where \( \rho_{kij} = \text{Cov}(I_{\alpha,i}(x), I_{\beta,j}(x)) \). The sign ambiguity \((\pm)\) arises from the fact that the inferred parameter may be either negatively or positively correlated with the antecedent. Negative correlations arise when \( \alpha \) is a decreasing function of the antecedent. The inferred parameter may be both positively and negatively correlated with the antecedent at different parts of the parameter space. When the sign ambiguity cannot be resolved a conservative (upper bound) for \( \rho_{\alpha \beta} \) can be obtained. Verification that conservative values for the correlation terms produce conservative estimates is presented next.

## 5.3 Conservative Estimation

In the Biscay distribution representation inter-parameter correlations are represented by a single (conservative) value. Correlation values must cater for both negative and positive correlations induced by monotonic increasing and decreasing segments of a non-monotonic inference. In this section, we show that conservative estimates of correlation values can be maintained which guarantee conservative estimates of Biscay information.

**Definition 17** A correlation \( \rho_{k_{ij}}^* \) is a conservative value for \( \rho_{ij} = \text{Cov}(I_i, I_j) \) if:

- Inference is a monotonic increasing function for all parameter values and \( \rho_{k_{ij}}^* \geq \rho_{ij} \).
- Inference is a monotonic decreasing function for all parameter values and \( \rho_{k_{ij}}^* \\leq -\rho_{ij} \).
- Inference is non-monotonic and \( \rho_{k_{ij}}^* > |\rho_{ij}| \).

We will now show that Equations 5.13 and 5.14 produce the necessary conservative correlation values after the fusion or inference of decision conservative estimates.

Assume \( p \) initial decision conservative estimates \( \{I_1, \ldots, I_p\} \) and also assume that conservative estimates \( \rho^* \) of their correlation values are known: \((\forall 1 \leq i, j \leq p)\ \rho_{k_{ij}}^* \geq \text{Cov}(I_i, I_j)\). In the case of inference, by Equation 5.14, the actual correlation between the inferred parameters \( \alpha \) and \( \beta \) is

\[ \rho_{\alpha \beta} = \frac{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}} \pm \rho_{k_{ij}}}{\sqrt{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}}^2 \cdot \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}}^2}} \]

and a conservative estimate \( \rho_{k_{ij}}^* \) is

\[ \rho_{k_{ij}}^* = \frac{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}}^*}{\sqrt{\sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}}^* \cdot \sum_{i=1}^{n_\alpha} \sum_{j=1}^{n_\beta} \rho_{k_{ij}}^*}}. \]

It follows immediately that the correlation \( \rho_{k_{ij}}^* \) computed using Equation 5.14 and conservative estimates \( \rho_{k_{ij}}^* \) instead of the actual correlations \( \rho_{k_{ij}} \) is itself conservative since \( \rho_{k_{ij}}^* \geq |\rho_{ij}| \).
Similarly, for the fusion case, by Equation 5.13:

$$
\rho_{\alpha_3 j}^* = \frac{W_{\alpha 1} \rho_{\alpha_2 j}^* W_{\beta 3} + W_{\alpha 2} \rho_{\alpha_3 j}^* W_{\beta 2}}{\sqrt{(W_{\alpha 1}^2 + W_{\beta 3}^2)(W_{\alpha 2}^2 + W_{\beta 2}^2)}}
$$

$$
\geq \frac{W_{\alpha 1} \rho_{\alpha_2 j} W_{\beta 3} + W_{\alpha 2} \rho_{\alpha_3 j} W_{\beta 2}}{\sqrt{(W_{\alpha 1}^2 + W_{\beta 3}^2)(W_{\alpha 2}^2 + W_{\beta 2}^2)}}
$$

$$
= \rho_{\alpha_3 j}^*.
$$

5.4 Some Properties of the DLQF

The fusion Equation 5.12 has two intuitively appealing properties for uncertainty dominated estimates: a monotonic decrease of uncertainty with accumulating data and a strong structural analogy to the 1D KF variance update equation (Equation 1.18, Chapter 1). Substituting for \(k_1\) and \(k_2\) in Equation 5.9 we obtain:

$$
W = W_1 W_2 \sqrt{\frac{1 - \rho^2}{X}}.
$$

(5.15)

In the special case when the observations \(z_1\) and \(z_2\) are uncorrelated (i.e., \(\rho = 0\)) and the estimate is dominated by uncertainty then:

$$
W = \frac{W_1 W_2}{\sqrt{W_1^2 + W_2^2}}.
$$

(5.16)

Suppose \(W_1\) and \(W_2\) represent the width of the \(n^{th}\) standard error (\(W_1 = 2n\sigma_1\) and \(W_2 = 2n\sigma_2\)) then we obtain the familiar Kalman filter update equation:

$$
\frac{1}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}.
$$

Thus, the KF variance update equation is a limiting case of the DLQF. However, the DLQF differs markedly from the KF in one important aspect. The QF never assumes that there is a single point estimate which is unbiased with respect to the true state of the system.

**Theorem 18** If two estimates of width \(W_1\) and \(W_2\) are combined using the DLQF then \(W \leq \min\{W_1, W_2\}\).

**Proof** By Equations 5.10 and 5.11 the result is immediate for cases when either \(W_1 - \rho W_2 < 0\) or \(W_2 - \rho W_1 < 0\) since either \(k_1 = 0\) or \(k_2 = 0\) and the most conservative input estimate is preferred as \(|\rho| \leq 1\). For cases when \((W_2 - \rho W_1)(W_1 - \rho W_2) \geq 0\) let \(W_i = \min\{W_1, W_2\}\) and \(W_j = \{W_1, W_2\} \setminus \{W_i\}\). Then by Equation 5.15 (and \(W_i^2 + W_j^2 + 2\rho W_i W_j \geq 0\)) we have:

$$
W = W_i \sqrt{\frac{W_j(1 - \rho^2)}{W_i^2 + W_j^2 + 2\rho W_i W_j}}
$$

$$
= W_i \sqrt{\frac{(W_j + \rho W_i)^2}{W_i^2 + W_j^2 + 2\rho W_i W_j}}
$$

$$
\leq W_i
$$

$$
\leq \min\{W_1, W_2\}.
$$
5.4 Some Properties of the DLQF

DLQF versus the IIF: Consensus and Discordance

Two scenarios of consensual and discordant observations demonstrate the superior aggregation properties of the DLQF over the IIF.

Consider first consensual observations. Suppose two independent observations yield the same interval estimate for a parameter $X$, $e_1 = (l, u)$ and $e_2 = (l, u)$. The IIF will not take advantage of the independence knowledge and conclude an aggregated estimate no more informative than the input estimates $e_2 = (l, u)$. However, the DLQF combines the Biscay distributions and returns a tighter estimate as is demonstrated in Figure 5.3, part (a).

![Figure 5.3: Illustration of DLQF fusion properties. In part (a), the fusion of two independent, identical estimates $(l, u)$ results in a tighter estimate $(l, u)$ than the IIF estimate. However, the fusion of two discordant estimates $(l_1, u_1)$ and $(l_2, u_2)$ produces an estimate $(l_3, u_3)$ which still covers the true state $x_0$, unlike the IIF estimate $(l_2, u_1)$, as is shown in part (b).](image)

When discordant observations are fused the IIF will not cater for this discord unless it is so severe that the input estimates do not overlap. Suppose two independent observations yield estimates for a parameter $X$, $e_1 = (l_1, u_1)$ and $e_2 = (l_2, u_2)$, with $u_2 > u_1$ and suppose that $e_2$ does not contain the true state $x_0$. The IIF aggregated estimate $(l_2, u_1)$ will also not contain the true state. However, when the DLQF combines the Biscay distributions it returns a tighter estimate and also adjusts the estimate to cater for the discordant observations as is demonstrated in Figure 5.3, part (b).

Figure 5.4 compares the IIF and DLQF for a static system observed under Gaussian noise with standard-error 1 unit. Each graph compares the coverage and interval widths for each filter. **Coverage** is the difference of the proportion of experiments for which the DLQF filter estimate necessarily contains the true state over the IIF. **Interval widths** are compared by displaying the ratio of the mean estimate width obtained by the DLQF against that obtained for the IIF.
for a thousand experiments are displayed in Figure 5.4 where each graph represents experiments under one of the following conditions: (a) zero imprecision (b) imprecision due to the inference \( f(x) = x + (-1, 1)\) and (c) imprecision due to an inference \( f(x) = x + (-1, 1)\) and an overestimation of noise standard error by a factor of 2.

![Graphs showing coverage and interval width](image)

(a) No Imprecision  
(b) Imprecision \( x \rightarrow x + (-1, 1)\)

(c) Imprecision and Noise Over-Estimate.

Figure 5.4: Comparison of the IIF and DLQF for a static system observed under Gaussian noise with standard-error 1 unit. Each graph compares the coverage and interval widths for each filter. Coverage is the difference of the proportion of experiments for which the DLQF filter estimate necessarily contains the true state over the IIF. Interval widths is the ratio of the mean estimate width obtained by the DLQF against that obtained for the IIF filter. Figures display statistics for a thousand experiments.

As the observation noise standard error increases multiple path problems can generate optimistic estimates. The affect of multiple paths is identical to that for the static landmark filter since we demonstrated at the beginning of this chapter that the Biscay statistics \( I - E(I)\) is independent of the choice of landmarks for the mean BPA. Divergence due to imprecision in the semi-quantitative model is the source of multiple paths for the DLQF. The affects of multiple paths is illustrated by two scenarios. Ten observations are made of \( x\) and \( y = f(x)\) is inferred using experiment dependent
5.5 Example Application: The Ballistic Father Christmas Problem

In the first scenario:
\[ f_1(x) = x + (-10, 10) \]

and in the second experiment 5 inferences are made each of the following:
\[ f_2(x) = x + (-1, 10), \]
\[ f_3(x) = x + (-10, 1). \]

Each experiment uses samples drawn from a Gaussian distribution of mean 0 and standard error 2 units. Intervals are chosen with landmarks at 0, \( \sqrt{2} \) and 2. The first and second experiments yield estimates of the second standard error of \( x \in (-10, 1, 11.6) \) and \( x \in (-1, 8, 3.0) \) respectively. Each estimate is consistent with the best possible informative estimate of \( (-10, 10) \) and \( (-1, 1) \) respectively. However, when a Gaussian distribution with standard error 10 units is used then the first experiment yields an estimate of \( (-5.7, 4.3) \) which is too optimistic! To date, no method exists to guarantee or predict the effects of multiple path for the DLQF although the problem is no different to the case of static landmarks.

5.5 Example Application: The Ballistic Father Christmas Problem

This example illustrates the advantage of both accuracy and precision offered by the DLQF over the IIF and KF in situations when imprecision is small but correlated over time. In the Ballistic Father Christmas problem Father Christmas travels from a Christian friendly country to a Muslim country. The aim is to track Father Christmas for one of two reasons:

- **Christian intentions.** To track and guide Father Christmas to a specific child.
- **Muslim intentions.** To track and intercept the western icon before any damage can be done.

Tracking is initialised when Father Christmas enters the atmosphere at high altitude and at a very high speed. His position is to be tracked by a radar which need not accurately measure range and bearing. Three types of force act on the saintly figure: aerodynamic drag, which is a function of his speed and has a substantial nonlinear variation in altitude, gravitational force which accelerates him towards the centre of the Earth and finally, random buffeting terms. The effect of these forces gives a trajectory which is initially ballistic but as the density of the atmosphere increases, drag effects become important and the sleigh rapidly decelerates until its motion is almost vertical.

The system model comprises the position of the sleigh \( (x_1 \text{ and } x_2) \), its velocity \( (x_3 \text{ and } x_4) \) and a parameter of its aerodynamic properties \( (x_5) \). The system dynamics are (Julier and Uhlmann, 1997):

\[
\begin{align*}
\dot{x}_1(k) &= x_3(k), \\
\dot{x}_2(k) &= x_4(k), \\
\dot{x}_3(k) &= D(k)x_3(k) + G(k)x_1(k) + \nu(k), \\
\dot{x}_4(k) &= D(k)x_4(k) + G(k)x_2(k) + \nu(k), \\
\dot{x}_5(k) &= 0.
\end{align*}
\]

\(^{4}\)The Father Christmas Problem is perhaps better known as the Ballistic Missile Problem, which is a bench mark problem in statistical estimation (Austin and Leondes, 1981; Lerro and Bar-Shalom, 1998; Julier and Uhlmann, 1997; Siouris et al., 1997). However, this thesis avoids military scenarios.
5.5 Example Application: The Ballistic Father Christmas Problem

$D(k)$ is the drag-related force term, $G(k)$ is the gravity-related force term and $\nu(k)$ is the process noise term: $\text{Var}(\nu(k)) = 2.4064 \times 10^{-5}$. Defining $R(k) = \sqrt{x_1^2(k) + x_2^2(k)}$ as the distance from the centre of the Earth and $V(k) = \sqrt{x_3^2(k) + x_4^2(k)}$ as the absolute sleigh speed then the drag and gravitational terms are:

$$D(k) = -0.59783 \exp(x_5) \exp \left( \frac{R_0 - R(k)}{H_0} \right) V(k),$$

$$G(k) = -\frac{Gm_0}{R^3(k)}.$$ 

For this example, $H_0$, $Gm_0$, $R_0$ and $x_5$ are known precisely ($H_0 = 13406$, $Gm_0 = 3.9860 \times 10^5$, $R_0 = 6374$, $x_5 = 2.69$). The motion of the sleigh is measured by radar which is located at $(x_{r}, y_{r}) = (0, 6374)$. It is able to measure range $r$ and bearing $\theta$ where:

$$r = \sqrt{(x_1(k) - x_{r})^2 + (x_2(k) - y_{r})^2} + \omega_1(k),$$

$$\theta = \arctan \left( \frac{x_2(k) - x_{r}}{x_1(k) - y_{r}} \right) + \omega_2(k).$$

The terms $\omega_1$ and $\omega_2$ are zero mean, uncorrelated noise with standard error of 1m and 17mmad respectively. The properties of the radar are very crudely known and the observed bearing is some systematic deviation $\Delta$ from the true bearing $\theta_{\text{obs}} = \theta_{\text{true}} - \Delta$ where $\Delta \in (0.0, 0.1)$.

How do standard numerical methods cope with this problem? We demonstrate the practical difficulties this problem poses for an Extended Kalman Filter (EKF) implementation.\footnote{The Extended Kalman Filter is a formalisation of the standard Kalman filter for non-linear systems. The EKF is accurate to third order in its mean and standard deviation estimates.} Firstly, let us consider the impact on the EKF of a systematic offset in radar bearing $\theta$. The systematic error is encoded, along with the random error, in the covariance matrix. A systematic offset means bearing observation noise is not independent of the observations at the previous time-steps. This violates a principal axiom of the EKF. Usually extra observation noise (stabilising noise) is added to the observation parameter in attempt to accommodate correlated observation noise. However, choosing the right amount of additional noise is a fine art as the following experiment demonstrates.

The following experiments were performed on simulated data. Figure 5.5 shows a typical trajectory and estimate sequence obtained using the EKF when the radar bearing imprecision is 0.09 radians and the systematic error is encoded in the observation covariance matrix: $\sigma_0 = 0.1 + 0.017$. Although $\sigma_0$ accommodates both the observation noise and systematic error, the EKF covariance matrix representation is unable to maintain a sufficient measure of systematic error through inference and, as the figure clearly shows, the EKF diverges as a result of over-confident polar to Cartesian inference from the radar observation.\footnote{Similar divergent behaviour was obtained for other non-linear approximate methods such as (Julliër and Uhlmann, 1997).} Of course, we could eventually produce a non-divergent EKF by increasing $\sigma_0$ further, but this process is rather unintuitive and ad hoc.

To investigate the behaviour of the DLQF and IIFs under various modelled situations the filters were implemented in Allegro Common Lisp (Version 4.1) on a SUN SPARC work station clone 4/75. The combination of interval arithmetic and multiple occurrence of variables in the system models caused excessive observation imprecision for $t < 50$ and excessive prediction imprecision for $t > 50$. To demonstrate some semblance of data-fusion point estimates were inferred using extremes of the input (antecedent) intervals. The inferred interval was then found from the extreme values of these point inferences. Of course, such an approach is sound only when the interval arithmetic expression
is monotonic. The DLQF used a two fractile representation with fractile values of 0 and 2 and, in order for the estimator to operate in real-time, inferences were drawn using only groups comprising identically valued input fractiles as outlined in Section 5.2.3. Typically inference and fusion over an entire trajectory (i.e. 200 steps) exhibited the following operation times: IIF inference 40.5 secs; IIF fusion less than 2 sec; DLQF inference 31.5 secs (correlation update, an extra 8.9 secs) and DLQF fusion 6.0 secs (correlation update, less than 2 secs extra).

Figure 5.6 compares the behaviour of the IIF and DLQF filters which use precise observation models where as Figures 5.8 and 5.9 compare the behaviour of the IIF and the DLQF which use imprecise observation models. A systematic radar bearing error is assumed (caused by motor stepping effects, for example). Figure 5.6 compares the trajectory estimates from the IIF and the DLQF when there is no imprecision in the radar bearing. In the figure, the estimate intervals represent (i) for the IIF, the intersection of interval estimates for parameters $x_1$ and $x_2$ separately, inferred from the second standard error intervals of the observations and (2) for the DLQF, the parameter value extremes for Biscay information $I(x_1) = \pm 2$ and $I(x_2) = \pm 2$. The IIF exhibits saw-tooth estimate uncertainty due to the fact that stochastic prediction uncertainty always accumulates but observation estimates impact on the overall estimate sporadically. The gradual accumulation of observation evidence by the DLQF eliminates the saw-tooth effect and produces more precise estimates. Figure 5.7 compares the coverage and estimate widths over 30 runs. The DLQF is, on average, 20% more precise than the IIF and, 5% and 16% more accurate for parameters $X_1$ and $X_2$ respectively.

In Figure 5.9 the random noise model is known and the DLQF is about twice as precise as the IIF. Over-estimating the standard error leads to increased occurrence of consensual estimates and the DLQF convergence rate far exceeds that of the IIF. When, for example, the noise model is over estimated two-fold, so that the range and bearing standard errors are assumed to be 0.002 and 0.034 respectively, the DLQF exhibits a three-fold precision over the IIF (see Figure 5.9).
5.6 Conclusions

This chapter has demonstrated the flexibility of the QF philosophy for problems involving dynamic landmarks. The concept of Blasckay information for discrete quantity-spaces has been extended to the notion of a Blasckay distribution for extrapolation of Blasckay information for arbitrary (not a priori defined) quantity-spaces. It is interesting to contrast the DLQF with a similar filter offered by Dubois (Dubois and Prade, 1994). Dubois develops a multi-interval estimation approach for which the set of intervals form a possibility distribution. The possibility membership function is interpreted in terms of probability measures but Dubois confines that, after inference and fusion, it is difficult to interpret the resulting degrees of possibility with a frequentist approach. Dubois and Prade's filter is also unable to accommodate correlated observations.

The DLQF has been demonstrated to be superior to the IIF in that (i) gradual accumulation of evidence in terms of mass leads to a more precise estimate and avoids the saw-tooth estimate uncertainty behaviour of the IIF filter and (ii) convergence of the DLQF is superior to the IIF when noise variance values are over-estimated. The DLQF is identical to the Kalman filter for precise estimates and yet is equivalent to the IIF when the estimate is dominated by imprecision. However, for intermediate conditions, when both imprecision and random uncertainty are present, the multiple path problem can lead to an optimistic representation of precision for fixed estimates.

The work presented in this chapter is preliminary and much remains to test the accuracy of the filter: a theoretical exploration of the effects of multiple path and the linearisation assumptions to name but a few. However, presented here is a compelling story which I believe demonstrates that the dynamic landmark filter has strong theoretical properties and strong application potential in situations where the IIF is insufficiently precise and the KF is unable to maintain systematic error over time.
Figure 5.6: Estimated sleigh trajectories obtained using IIF and DLQF for zero imprecision models ($\hat{\theta} = \theta_c$) using fractile values 0 and 2. Graphs (a) through to (f) comprise three solid lines. The inner line shows the true parameter values and the outer lines are the estimate bounds for either the IIF or the DLQF. In (a) and (b) the location of the Earth’s surface is shown. The IIF trajectory estimates exhibit the typical “saw-tooth” effect as observations contribute informatively to the estimate only occasionally.
Figure 5.7: Comparison of coverage and estimate widths over 30 runs of the zero-imprecision Ballistic Father Christmas scenario. Each ratio is the expected statistic of the DLQF divided by the expected statistic of the IIF. The DLQF is, on average, 20% more precise than the IIF and, 5% and 16% more accurate for parameters $X_1$ and $X_2$ respectively.
Figure 5.8: Estimated sleigh trajectories obtained using IIF and DLQF for $\delta \theta = 0.09$ (estimated $\delta \theta = 0.1$) using fractile values 0 and 2. Graphs (a) through to (f) comprise three solid lines. The inner line shows the true parameter values and the outer lines are the estimate bounds for either the IIF or the DLQF. In (a) and (b) the location of the Earth’s surface is shown.
Figure 5.9: Estimated sleigh trajectories obtained using IIF and DLQF for $\delta \theta = 0.09$ (estimated $\delta \theta = 0.1$) using fractile values 0 and 2 and over estimates in the observation noise standard errors (times 2). Graphs (a) through to (f) comprise three solid lines. The inner line shows the true parameter values and the outer lines are the estimate bounds for either the IIF or the DLQF. In (a) and (b) the location of the Earth’s surface is shown.
Chapter 6

Experiments

6.1 Introduction

In this chapter the utility of the interval algebra representation and the Qualitative filter is demonstrated for a number of problem types:

- Refinement of qualitative representations by non-redundant sensor fusion.
- Scale-space preprocessing of candidate qualitative model hypotheses.
- Non-linear filtering.

Each problem type is illustrated with real data drawn from appropriate scenarios in the robot sensing domain.

6.2 Experiment 1: Surface Discrimination (Sparse Representation)

This experiment demonstrates the utility of the qualitative Acoustic Flow model which uses ordinal domain information only. The qualitative Acoustic Flow model is applied to feature recognition using data from sparsely calibrated sensors. Real data is used from the robot sensing domain and is obtained as described in Section 2.2 using a tracking gyro mounted, sonar RCDD unit and a scanning infrared time-of-flight sensor.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>${0, 1.22, 1.50, 1.57, 1.64, 1.92, 3.14}$</td>
</tr>
<tr>
<td>$\delta R$</td>
<td>${-\infty, -0.1, -0.04, -0.01, 0, 0.01, 0.04, 0.1, \infty}$</td>
</tr>
<tr>
<td>$\delta \phi$</td>
<td>${-\infty, -0.05, -0.01, 0, 0.01, 0.05, \infty}$</td>
</tr>
<tr>
<td>$\tan(\theta)$</td>
<td>${-\infty, -14.1, -5.471, -2.73, 0, 2.73, 5.47, 14.10, \infty}$</td>
</tr>
<tr>
<td>$\delta R \tan(\theta)$</td>
<td>${-\infty, -0.14, -0.10, 0, 0.10, 0.14, \infty}$</td>
</tr>
<tr>
<td>$R + r$</td>
<td>${-\infty, 0, 5, \infty}$</td>
</tr>
</tbody>
</table>

Figure 6.1: Quantity-spaces for robot navigation domain.

A sonar estimate for $Q\text{mag}(R \tan \theta)$ is combined with the estimate for $Q\text{dir}(\phi)$ from the gyroscope thus:

$$m_{R+r}(A) = \max_{Q\text{mag}(A) \in Q\text{mag}(B) \in Q\text{mag}(C)} \{m_{\phi}(B)m_{R+r}(\theta)(C)\}.$$
The sonar+gyroscope estimate for Qmag\((R + r)\) and the laser curvature estimate Qmag\(\frac{d}{dr}R\) are then fused according to Table 2.4 in Section 2.1.

Calibration points (or corresponding values) and the quantity-spaces for this domain are shown in Figure 6.1. An infinite amount of information is required to converge to a landmark. Thus, each landmark is “blurred” and represented as a narrow interval on the real line in such a way that a finite number of observations will lead to convergence to a region surrounding the landmark. Blurring degree is context dependent. In this experiment we investigate a convex surface with a curvature no greater than 2\(m\). Consequently, we may choose to consider a plane to be any surface with curvature greater than 7\(m\) and a convex surface to have a positive curvature less than 3\(m\). Thus, if a robot is constrained to sense the surface at a range less than 2\(m\) then \(R + r > 5m\) uniquely defines a planar surface. A concave surface exhibits \(R + r < 0\) or is disambiguated from a convex surface using the laser sensor (as described in Section 2.1). Estimates are obtained using the QF mean estimator.

\[\begin{array}{c}
\text{Range (metres)} \\
-0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8
\end{array}\]

\[\begin{array}{c}
\text{Bearing (rads)} \\
2.6 2.4 2.2 2.0 1.8 1.6 1.4 1.2 1.0
\end{array}\]

Figure 6.2: Sonar, gyro and laser data for a convex surface. The sonar and gyro data is obtained from a gyro mounted tracking RCDD unit shown in Figure 2.1. The sonar and gyro data is gathered as the robot moves in front of the surface and the RCDD tracks the normal reflection of the surface. The laser data is obtained by a stationary robot which scans the entire surface from a single location. S1 and S2 are two subsets of laser data points used to determine the curvature of the surface.

Data for a convex surface is shown in Figure 6.2. The following standard deviations were assumed for the sonar sensor and gyroscope readings: 0.01 metres for \(R\), 0.02 rads for \(\theta\) and 0.1 rads for \(\delta\phi\). Biscay information for each observation set and for the aggregate is shown in Figure 6.3. A gating strategy is introduced to avoid outlier misclassification: an observation Biscay value, \(I_o\), is aggregated with the running estimate, \(I\), when either the Biscay values have the same sign or \(|I - I_o| \leq 3\).

Near-concave behaviours are rapidly filtered out. However, time steps 5 to 13 do not contribute to filtering out planar behaviour. The reason for this is that the model is ambiguous for these observations. Further refinement of the quantity-spaces is required to reduce this problem. The decision that the surface is convex or far-concave as opposed to planar can be made with 89% certainty.

\[1\text{Cross-correlation terms are assumed to be negligible.}\]
confidence only after 14 observations have been received from each sensor:

\[
I_{bg} (\{\text{convex, far-concave} \} , \{\text{planar}\} ) = 3.0, \\
I_{bg} (\{\text{convex, far-concave} \} , \{\text{close-concave}\} ) = 3.0.
\]

![Figure 6.3: Biscay information.](image)

Figure 6.3: Biscay information.

To determine surface curvature from the laser data we identify two subsets of data points S1 and S2, (shown as shaded regions in Figure 6.2) such that the data in S1 have greater range values than those in S2. We then evaluate the mean bearing distance \(l\) between points within the same subset and the variance of this distance. We can then determine the curvature from \(E(l_1 - l_2)\):

\[
Q \text{mag} (E(l_1 - l_2)) \begin{cases} 
> 0 & \iff Q \text{mag} (\frac{\partial R}{\partial s^2}) = +, \\
< 0 & \iff Q \text{mag} (\frac{\partial R}{\partial s^2}) = -.
\end{cases}
\]

We assume that the laser sensor range information is accurate to 0.05m and the bearing inaccuracy is comparatively negligible (Optik-Elektronik, 1995). From a single sweep, the following estimates for \(Q \text{mag} (\frac{\partial R}{\partial s^2})\) were obtained from the laser sensor data:

\[
I(\{\text{close-concave, convex, planar}\} , \{\text{far-concave}\} ) = I_a (\{\text{convex, far-concave}\} , \{\text{close-concave}\} ) + I_b (\{\text{close-concave, convex, planar}\} , \{\text{close-concave, convex, planar}\} ) = 7.0.
\]

Combining laser and sonar estimates using the \(\infty\)-norm Rule of Combination we obtain: 2

\[
\begin{align*}
I(\{\text{convex}\} , \{\text{close-concave}\} ) &= I_a (\{\text{convex, far-concave}\} , \{\text{close-concave}\} ) + I_b (\{\text{close-concave, convex, planar}\} , \{\text{close-concave, convex, planar}\} ) = 3.0, \\
I(\{\text{convex}\} , \{\text{far-concave}\} ) &= I_a (\{\text{convex, far-concave}\} , \{\text{convex, far-concave}\} ) + I_b (\{\text{close-concave, convex, planar}\} , \{\text{far-concave}\} ) = 7.0, \\
I(\{\text{convex}\} , \{\text{planar}\} ) &= I_a (\{\text{convex, far-concave}\} , \{\text{planar}\} ) + I_b (\{\text{close-concave, convex, planar}\} , \{\text{close-concave, convex, planar}\} ) = 3.0.
\end{align*}
\]

2 The normalisation constant is 1 since either \(I_a = 0\) or \(I_b = 0\) in each expression for \(I\).
Using a decision gate of 3 and that \( I(\text{convex}, \{\text{close-concave}\}) \geq 3, I(\text{convex}, \{\text{far-concave}\}) \geq 3 \) and \( I(\text{convex}, \text{planar}) \geq 3 \), we may safely conclude that the surface is *convex*.

The above experiment was repeated (but without the laser sensor) for planar surfaces which involved determining whether \( R + r > 5m \). Figure 6.4 shows the data and Biscay information for two recognition trials. However, the range differential \( \delta R \) values are of insufficient magnitude to disambiguate the planar surface type. This problem arises out of the discrete nature of the quantity-space representation. Reducing the sampling rate until \( |\delta R| \geq 0.04 \) rectifies this problem for trial 1 but the same sampling rate fails for trial 2.

In summary, the quantity-space representation is intended to simplify the modelling process when a fully calibrated system is not available. However, in practice, obtaining an informative qualitative model is hard. Even when the underlying theoretical qualitative model is informative in theory, the need to cater for noisy observations requires landmarks to be blurred. This introduces ambiguity into model-based inferences and to overcome this the quantity-spaces must be partitioned further. The problem is guaranteed to disappear eventually when the partition is sufficiently refined that a fully calibrated sensor can be extrapolated by linear regression between landmarks. At this point a Biscay distribution may be defined over the quantity-space. Finding a small set of adequate partitions is an extremely difficult problem even for the simple Acoustic Flow QDE and the quantity-space used in this experiment was designed initially by exploring the full quantitative Acoustic Flow Equations and then refined by trial and error. Automating the partitioning process using an adaptive algorithm is discussed in Chapter 7.

### 6.3 Experiment 2: Multi-hypothesis Scale-space Reasoning

The QF decision process requires pair-wise comparisons of process model hypothesis but in many circumstances the number of hypotheses consistent with the knowledge of the system can be prohibitively large. For example, to identify a complex feature structure comprising \( N \) convex and concave sub-features could require the enumeration of \( 2^N \) hypothetical structures and \( 2^{(N-1)}(2^N - 1) \) pair-wise comparisons. A more efficient approach would be to select a small set of exemplar segmentations of the data. It is often the case that the signal and noise processes operate at significantly different scales and model-free scale-space reasoning preprocessing methods can be used to suggest the most likely hypothesis prior to filtering. Scale-space methods have been used extensively for the signal processing of visual images and methods exist for mean and median unbiased filtering. Marr and Hildreth (Marr and Hildreth, 1980) introduced the (unique) linear Gaussian convolution operation for zero-crossing detection,

\[
f_\sigma(x) = N(x-a,\sigma) * f(x).
\]

Witkin (Witkin, 1983) defined scale in terms of \( \sigma \). As the size of \( \sigma \) increases the filtered image exhibits two effects: qualitative simplification (i.e. the removal of fine-scale features) and spatial distortion (i.e. dislocation, broadening and flattening of the features that survive). Although the locations of the zero-crossings are not preserved at coarser scales it is assumed that the true location of an event giving rise to a zero-crossing contour is the contour’s \( x \) location as \( \sigma \to 0 \). Thus, the distortion at coarse scales can be overcome by tracking coarse extrema to their fine-scale locations (Witkin, 1983). However, when a sparse number of scales is used it is often difficult to associate each coarse scale edge with the original image.
Figure 6.4: Two trials for specular plane sampling showing the original sensor data, Biscay information for each inference of $R + r$ and the aggregate Biscay information.

An alternative approach to the Marr and Hildreth filter is the morphological (set theoretic) non-linear filter such as the COMOC (median/mean of close-open and open-close). The morphological filter uses set union and set intersection to remove structure without deforming or scaling the underlying signal (Noble, 1989). The COMOC filter is described in Appendix C where we extend Noble’s proof that the filter is unbiased with respect to the mean for constant signals to that of rotationally symmetric signals. When noisy inferences are drawn using qualitative models we may use COMOC
in Biscay information space to suggest qualitative models for the underlying observation signal.

![Graphs of observation data over time for two runs.](image)

**Figure 6.5:** Sonar range and bearing data and gyroscope data for two runs in front of a rough convex-concave-convex surface.

We investigate the utility of both COMOC and Marr-Hildreth scale-space reasoning applied to Biscay information space for two runs of the robot in front of a predominantly concave surface with convex ends. Sensor data is shown in Figure 6.5. The Marr-Hildreth filter and the recursive COMOC filter scale-space representations are shown in Figures 6.6 and 6.7. For each ordinate scale value the data is classified as either belonging to convex (white space) or concave (black space) behaviour according to the value of \( \text{filter}(I) \). The dashed lines in this figure demarcate the scale values of the filtered images which exhibit identical qualitative behaviours. Although the Marr-Hildreth filter partitions the scale-space into more distinct qualitative behaviours the event locations of the crucial true convex-concave-convex behaviours for both spaces coincide. We shall use the COMOC filter to obtain a candidate set of qualitative model hypothesis for which the true structure (i.e. convex-concave-convex) appears between scales 5 and 15 for the first run and 5 and at least 20 for the second run.

![Image of scale-space representations for COMOC filter.](image)

**Figure 6.6:** 1D scale-space representations of inferred Biscay information for the curvature of a reflective feature obtained using the morphological COMOC filter. Black and white regions indicate concave and convex behaviour respectively.
6.3 Experiment 2: Multi-hypothesis Scale-space Reasoning

Figure 6.7: 1D scale-space representations of inferred Biscay information for the curvature of a reflective feature obtained using the linear Marr-Hildreth Gaussian convolution filter. Black and white regions indicate concave and convex behaviour respectively.

For Run 1, Table 6.1 suggests four candidate qualitative model hypothesis (‘+’ and ‘-’ indicate positive and negative curvature respectively): Hypothesis R4 corresponds to the scale of the observation noise and is a facsimile of the original noisy data. Hypothesis R3 indicates a seven part structure (convex-concave-convex-concave-convex-concave-convex) whereas hypothesis R2 indicates a three part structure (convex-concave-convex). Hypothesis R1 implies that the reflector is two part (convex-convex). Deriving a preference ordering and a confidence measure for each hypothesis using the COMOC structure is hard. However, we may derive a preference ordering for these hypotheses and a confidence measure using the QF. The following Biscay information was obtained for hypothesis pairs \((h_1, h_2)\) where \(h_1, h_2 \in \{1, 2, 3, 4\}\). An (non-conservative) observation standard deviation of 1.0 was assumed:

<table>
<thead>
<tr>
<th>(J_{h_1, h_2})</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>-1.43</td>
<td>-4.97</td>
<td>-5.18</td>
</tr>
<tr>
<td>2</td>
<td>1.43</td>
<td>0.0</td>
<td>-5.45</td>
<td>-4.95</td>
</tr>
<tr>
<td>3</td>
<td>4.97</td>
<td>5.45</td>
<td>0.0</td>
<td>-1.40</td>
</tr>
<tr>
<td>4</td>
<td>5.18</td>
<td>4.95</td>
<td>1.40</td>
<td>0.0</td>
</tr>
</tbody>
</table>

A hypothesis is discounted with an error rate no larger than \(\frac{1}{2}\) when the Biscay information between the original data and the hypothesis exceeds \(G\). The only hypothesis \(h\) which is not excluded by the Biscay information for a decision gate of 3 (i.e. so that \(J_{h, h} \leq 3\)) is hypothesis R3. Thus, we may conclude that the surface is three part convex-concave-convex.
6.4 Experiment 3: DLQF and EKF Comparison under Zero Imprecision

In the final experiment, we use the precise Acoustic Flow model to predict the radius of curvature of a uniform specular cylindrical surface of real radius of about 10m. The DLQF and EKF filters are compared using real data from the robot sensing domain (see Figure 6.8).

Figure 6.8: Gyroscope and RCDD sonar range and bearing data for traversal in front of a specular cylinder with regular radius of curvature of 0.1m.

As the robot moves in front of the surface it receives sets of corresponding readings of bearing, range and gyro change. The radius of curvature estimate for those individual observations obtained at time-step $k$ is calculated using:

$$r(k) = (R(k) + \frac{\Delta R(k)}{\Delta \phi(k)} \tan \theta(k))$$

(6.1)

where $\Delta R(k)$ and $\Delta \phi(k)$ denote the finite change of range and turn of the gyro respectively between time steps $k-1$ and $k$. Observation noise is assumed to be uncorrelated so that, for the EKF:\footnote{The range value $R(k)$ and the differential range $\Delta R(k) = R(k) - R(k-1)$ are obtained using different observations of $R(k)$.}

$$Var(r(k)) = Var(\Delta \phi) \left( \frac{\Delta R(k)}{(\Delta \phi(k))^2} \right)^2 + Var(\Delta R) \left( \frac{\tan \theta(k)}{\Delta \phi(k)} \right)^2 + Var(\theta) \left( \frac{\Delta R(k)}{(\sin \theta(k))^2 \Delta \phi(k)} \right)^2 + Var(R).$$

The standard errors for $\Delta \phi$, $\Delta R$, $\theta$ and $R$ were found to be 0.01 rad, 0.005m, 0.02m and 0.01rad respectively.

Figure 6.9 shows the target radius of curvature estimate and its standard error over time for the EKF and the DLQF filters. The figure shows individually inferred estimates (i.e. Figures (a) and (b)) for each time-step and also the result of fusing these estimates (i.e. Figures (c) and (d)). The EKF estimate is clearly too optimistic. The EKF fails to accommodate the significant non-linear error induced by the gyroscope value $\Delta \phi$ in Equation 6.1 which is close to zero. Figure (b) clearly shows that the inferred estimate error is too small for this reason. The DLQF, however,
Figure 6.9: Comparison of DLQF and EKF estimates for the radius of curvature of a specular convex surface of regular radius 0.2m. Figures (a) and (b) show individual estimates inferred from gyroscope and sonar range and bearing data. Figures (c) and (d) show the accumulated estimates for both filters. The EKF estimate of radius of curvature in graph (d), although a very precise estimate, is negative and the EKF concludes incorrectly that the surface is concave. The DLQF is indecisive. The EKF model is highly non-linear when the gyro turn-rate is small and, as Graph (b) demonstrates, excessive weight can be attached to negative curvature observations via their variance estimates.

maintains sufficient inference error. Although, many of the DLQF individual observation second fractile estimates for the radius of curvature include the value \( \Delta \phi = 0 \) so that these observations become non-informative estimates of the radius of curvature.

6.5 Conclusions

The experiments presented in this chapter illustrate various applications of interval algebra and the Qualitative filtering framework to various types of problem. Each problem type was represented by some example scenario in the robot sensing domain. They were:

- Refinement of qualitative representations by non-redundant sensor fusion.
6.5 Conclusions

- Scale-space preprocessing of candidate qualitative model hypotheses.
- Non-linear filtering.

Experiment 1 demonstrates the QF in operation on the feature recognition problem for robot navigation under noisy observations. This experiment illustrates the difficulties encountered when implementing qualitative models in real world situations. When observations are noisy, quantity-space partitions must contain reasonably wide intervals to allow mass to converge into them. However, wide intervals can lead to ambiguous inferences. Further, even in the absence of noise the quantity-space landmark values must lie within the range of parameter data values. We saw, in the case of planar recognition, that when $\delta R$ is too small the QF is unable to recognise that the surface is planar. The QF is uninformative in this case. Building appropriate parameter space partitions requires a degree of trial and error.

Experiment 2 addressed the problem of pair-wise filtering when the number of possible behaviours is prohibitively large. Scale-space reasoning can be used in such situations to reduce the number of behaviours. Experiment 3 showed how the DLQF can overcome problems of non-linear filtering. Problems which reduce the effectiveness of many familiar filtering techniques such as the EKF.
Chapter 7
Conclusions and Directions for Future Research

7.1 Conclusions and Summary of Contributions

The primary contribution of this thesis is a filtering framework appropriate for a broad range of information types describing many levels of imprecision. This work has been concerned with the qualitative representation of observations for various sensor modalities (specifically odometric, gyrometric, sonar and infra-red time-of-flight range sensors) and their fusion. This thesis has addressed the problem of multi-sensor data-fusion of noisy data given imprecise projection models and incomplete information of the underlying noise processes. The main contributions are:

- The refutation of Cox's assertion that additive evidence formalisms are necessary. Using Cox's requirements of first order logic and conditional evidence assignment, a general theory for evidential reasoning was developed which encompasses Dempster-Shafer Theory, Possibility Theory and the Kalman filter.

- Problems with the probabilistic interpretation of mass values and the empirical difficulty of obtaining and maintaining full probabilistic models from sparse data were overcome using a non-parametric, estimation-based statistical filtering framework based on the corrigibility of preference ordering over hypotheses. However, when required, an upper bound on the probability of false-positive error-rates can be obtained using the Chebyshev inequality.

- The Biscay measure of information was developed and the Biscay distribution was created for semi-quantitative reasoning with bounding envelopes. Conservative filters were developed for severely incomplete noise models.

- The Qualitative filtering framework was applied to fixed landmark hypothesis testing, dynamic landmark estimation and stochastic processes. Morphological scale-space reasoning techniques were shown to have appropriate properties for a priori filtering of large numbers of hypotheses.

- Single and multiple reflection robot centred sonar sensor models (the Acoustic Flow equations) were developed for qualitative reasoning. Qualitative models for continuous time frequency modulated sonar sensors were developed for surfaces of arbitrary curvature.

- The filtering framework was tested with real data in the robotics domain.

We have demonstrated that:
• Possibility and Dempster-Shafer theories are instances of a more general meta-theory: the P-norm theory.

• The non-probabilistic interpretation of evidential mass value, offered in this thesis, is self-consistent for inference and fusion under the Lagrangian approximation to Eulerian filtering.

• Mass filtering smooths the decision process of the IIF and accommodates systematic error which is difficult to encode in the KF. However, the QF acts optimally like the IIF when imprecision dominates and has an optimal variance update when uncertainty dominates.

• Pure qualitative models, based on QDE limit points, can be sufficient for feature curvature discrimination using complementary information from multiple sensors.

• The choice of landmarks constrains the range of informative robot behaviours. Thus, the robot’s behaviour is constrained by its internal representation of the environment and the degree of precision inherent in its model of the environment.

7.2 Future Work

In this section we restate two problems which remain to be solved before QR technology can be applied to sensor-data fusion and filtering applications confidently. These problems are the multiple-path problem and the partitioning problem. Detail solutions to these problems are proposed.

The multiple-path problem arises out of the Lagrangian approximation to Eulerian filtering. Analysis of the multiple-path problem is extremely complicated and requires structural reasoning as the section on SOPAP demonstrates. The partitioning problem is the problem of choosing informative landmarks for system modelling. The interval representation can be highly uninformative and inferences can be ambiguous. This was illustrated clearly by Experiment 1 in Chapter 6 when certain robot behaviours failed to disambiguate convex and planar surfaces. In practice, QR researchers assume that sufficient information is available to draw necessary conclusions. They assume that the information is sufficiently stable to define static bounding envelopes. An alternative approach to resolving inference ambiguity problems has been offered by the Fuzzy Logic community. An inference is ambiguous when an inferred interval is consistent with two or more hypothesis intervals. A fuzzy interval distance measure is defined between the inferred interval and each of the hypothesis intervals and is often a function of the distance between the centres of the inferred and hypothesis intervals and the widths of the intervals (see, for example, (Shen and Leitch, 1993)). However, these distance approaches assume extra information that is not available and not necessarily consistent with the actual system processes. Distance methods are fragile where the imprecision is severe.

The second problem is the quantity-space (re-)partitioning problem. Refinement of ambiguous inferences is a context specific process. Ideally, the QR system would be able to adapt its representation in accordance with the environment context. How do we recognise that a change of context has taken place and how do we define new context specific concepts? The need for recalibration is indicated by strong discordances between sensor values and predicted values. These disagreements are hard to explain away using noise only and may arise due to sensor failure or due to sensor competence violations. When observations are discordant the question arises as to whether information should be fused not just how it should be fused. Bower (Bower, 1974; Murphy, 1996) decomposes sensory integration into a four level taxonomy (integration modes):
7.2 Future Work

- **Level 1: Complete Sensor Unity.** All sensor observations are fused without any mechanism for detecting discordances.

- **Level 2: Unity with awareness of discordance and the possibility of recalibration.** Discordances between sensors can be detected and are reconciled by the recalibration of the offending sensors.

- **Level 3: Unity with awareness of discordance and tendency toward suppression.** Discordances are detected and the offending sensors are not recalibrated, instead those sensors are temporarily suppressed.

- **Level 4: No Unity at all.**

Level 1 has been the focus of this thesis for which we have assumed that models (although imprecise) are known a priori and the quantity-spaces are partitioned appropriately. Often sensor drift will introduce discordances between sensor observations. Level 2 in Bower’s taxonomy requires the detection and correction of this kind of discordance. Level 3, on the other hand, I believe is too strong. Although the offending sensors should be temporarily suppressed they should also be re-adapted. Model adaptation is also required when the quantity-spaces are insufficiently partitioned to supply an adequate control strategy, in which case dynamic quantity-spaces and models may be required. The aim here would be to automate the scientist: to automate the model building and conceptualisation processes. Thus, Bower’s Levels 3 and 4 point us towards six areas of potentially fruitful research:

- Automated learning of domain models.
- Adaptive domain models: the recalibration (i.e. reassociation) of quantity-spaces.
- Adaptive non-parametric statistical models.
- Combinations of the above.

Some of these areas are explored below. We propose a method for adaptive domain modelling based on Edelman’s Theory of Neural Group Selection (TNGS) which applies Hebbian learning of spatio-temporal correlations to neural group responses. Our method adapts the TNGS to symbolic (interval) representations and context specific models are constructed from sensor cue spatio-temporal correlations.

### 7.2.1 Domain Model Learning

This section outlines a simple algorithm which learns the qualitative functional relationships between sensor cues. This algorithm satisfies three criteria:

- **Occam’s Razor** “Entia non sunt multiplicanda praeter necessitatem” meaning “Entities should not be multiplied without/beyond necessity”. The model which involves fewest operators and variables should be preferred.

- **Dimensional analysis** (e.g. Speed V [m][s]^{-1}). Models should be dimensionally consistent.
• **Qualitative operator descriptions** (e.g. *, +, −, cos and sin). There should be a complete set of symbols from which consistent models can be constructed.

**Algorithm**

The following is a proposed algorithm outline for learning relationships between sensors:

1. Initially have \( N \) observation qualitative-value vectors. For example:

   \[
   < r, \theta, R, v, \phi, \dot{R} >= \lhd -Q3, +, - >. \tag{7.1}
   \]

   where \( Qmag(\theta) = Q3 \) refers to a bearing in the third quadrant.

2. Find the minimal representation of the \( N \) observation vectors in terms of disjunctions. For example:

   \[
   < - \lor +, 2 \lor 3, +, -, - \lor +, - >. \tag{7.2}
   \]

3. Identify relevant variables. These are variables which have different qualitative-values between observation vectors.

4. Identify possible functional relationships. For example, if \( \theta \) quadrants 1 and 4 yield similar observation vectors and similarly for quadrant 2 and 3 then infer a cosine relationship.

5. Starting with a single operator model find dimensionally consistent and observation consistent qualitative models. Repeat for increasing model sizes (in terms of the number of operators) until a model is found which explains all the disjunctive observation vectors.

**Experiment**

To illustrate the algorithm above we randomly generated 43 distinct consistent observation vectors \( < r, \theta, R, v, \phi, \dot{R} > \) subject to the constraint:

\[
Qdir(R) = \lhd Qmag(v) \lhd Qmag(cos \theta). \tag{7.3}
\]

The aim was to attempt to infer Equation 7.3 from these observation vectors. The 43 vectors reduced to 2 disjunctions:

\[
< - \lor +, 2 \lor 3, +, -, - \lor +, - > \quad \text{and} \quad < - \lor +, 1 \lor 4, +, +, - \lor +, - > \tag{7.4}
\]

from which the following variables were considered relevant:

\[
\{ R, v, r, \cos \theta \}. \tag{7.5}
\]

With a potential operator depth of 2, 318 equations were generated out of which 236 were considered dimensionally consistent. From these only 1 equation was consistent with all the observation vectors, namely Equation 7.3.

An interesting extension to the problem arises when we consider noisy data. In which case Occam’s razor is in conflict with the intuitively preferred most probable models. Further, maintaining and tabulating lists of correlations between observations can be expensive. An alternative approach would be to utilise on-line dynamic adaptation.
7.2.2 Adaptive Representations

It is, in general, undesirable to store instances of sensor cue occurrences (i.e. feature lists) for the purpose of model building. Representing all potential correlations between $P$ parameters, say, each with a quantity-space partition size $N$ would require $N^P$ instances. Further refinement of the representation would increase the representational burden. Inter-sensor correlations would be maintained more efficiently when only those qualitative distinctions salient to the task context are represented. This would be achieved in a single map when it is able to adapt to different concept repertoires generated by different sensor-cue correlations in different contexts.

Similar problems have been studied in the context of biological adaptive systems (Edelman, 1987; Edelman and Tononi, 1995; Stein and Meredith, 1986). Edelman addresses the problem of how organisms categorise the world just by interacting with their environments (Edelman, 1987):

*One of the fundamental tasks of the nervous system is to carry on adaptive perceptual categorisation in an “unlabelled” world - one in which the macroscopic order and arrangement of objects and events (and even their definition or discrimination) cannot be prefigured for an organism, despite the fact that such objects and events obey the laws of physics.*

Edelman recognises two basic phenomena a theory of categorisation must entail. Firstly, individual nervous systems show enormous structural and functional variability. Individual variation from brain to brain far exceeds that which could be tolerated for reliable performance in any machine constructed according to current engineering principles (Edelman and Tononi, 1995). Secondly, while the real world obeys the laws of physics, it is not uniquely partitioned into objects and events. To survive, an organism must be able construct taxonomies and their ontologies. It needs to partition the world into perceptual categories according to its adaptive needs. Even after partitioning occurs as a result of experience, the world remains to some extent an unlabelled place full of novelty (Edelman, 1987).

The answer, as described by Edelman’s Theory of Neural Groups Selection (TNGS), is that non-symbolic systems can develop their own ontology by means of *selective mechanisms* which continually and dynamically re-map the neural cortices. Different events in each cortex are correlated at different times and places by assimilating neural groups. Each map represents a sensory cue type and perception emerges from the competition between maps via communication channels called *re-entrants*. Each map is subject to continual recategorisation by adaptive, selective neural group competition. A group is a tightly correlated collection of neurons and existing groups of neurons compete to attract neighbouring neurons within the map. A group, once formed, represents an individual concept.

The TNGS proposes three mechanisms to account for the production of adaptive behaviour: *developmental selection*, *experimental selection* and *re-entrant signalling*. Developmental selection produces physically variant network structures in each individual (called *primary repertoires*). Experimental selection is the strengthening or weakening of synapses as a result of behaviour and leads to the formation of highly correlated groups known collectively as the *secondary repertoire*. Re-entry is the binding of physically segregated maps through spatio-temporal regularities so that distinct operations in different maps that are related to the same perceptual stimulus can be effectively integrated.\(^1\)

\(^1\)A minimal arrangement of two re-entrantly connected maps is called a *classification couple.*
Although Edelman is concerned with conceptual association it is possible to extend his ideas to
the interval arithmetic representation explored in this thesis and to inference and conceptual
unification. Edelman's TNGS may be applied to categorisation within Qualitative representations
if we are prepared to accept a semantic reduction into the \( \infty \)-norm theory of evidential reasoning.
This reduction articulates the meaning of a neural group. Such a reconciliation is useful as concepts
emerge out of inter-correlation of parameter quantity-spaces which leads to:

- An enlarged taxonomy of conceptual categories by splitting old categories or by correlating
  concepts between quantity-spaces. An example of the latter is Smith and Medin's concept of
  "pet-fish" which is formed by combining the concepts of "pet" and "fish" (Smith and Medin,
  1981). We will shortly demonstrate that conceptual combination can emerge from degenerate
  mapping of sensor cues onto the concept space in conjunction with Hebbian adaptation within
  the concept map. Competitive selection would ensure that under used associations would
  eventually achieve a small synaptic weight thus asserting a weak association.

- (Biologically plausible) hierarchical and abstract levels of representation which can be main-
  tained depending on the distinctions salient to categorisation.

Assume that \( P \) parameters, indexed \( i \in \{1, \ldots, P\} \), are partitioned into quantity-spaces of size
\( \mathcal{N}_i \). Suppose also that there is an explicit representation for correlations between these parameters
called the \textit{concept space}. \(^2\) The concept space comprises \( B \) \textit{correlates} and each correlate represents
the correlation between specific quantity-space regions. It is important to note that, unlike the
quantity-space, which implicitly represents a value ordering between regions, the concept space is
not topographic in this sense. \(^3\) The primary repertoire comprises a partly connected network: a
parameter quantity-space value is necessarily not connected to each correlate in the concept space.
The connection from value \( Q \) to correlate \( b \) indicates a possible mapping of the true state within \( Q \)
into a value represented by \( b \). The set of connections from \( Q \) is the \textit{projection} of \( Q \) onto the concept
space. To account for the multiple role of a parameter value the projection must be degenerate.
However, complete degeneracy would not allow the formation of multiple concept values. This is
the reason why the network must not be fully connected. However, within the concept space the
correlates are fully connected. A strong connection between two correlates indicates a conceptual
unification (or instances of the same category). Each correlate represents a concept, and categories
are represented explicitly as a group of strongly connected correlates.

Classification couples from parameter space to concept space can be interpreted in terms of \( \infty \)-
form Plausibility assignments. The BPA in the parameter space is mapped, degeneratively, onto
subsets of correlates in the concept space. Each parameter space excites correlates in the concept
space. The intersection of all projections indicate all possible correlates given the sensory inputs
and these correlates represent the concept value corresponding to these inputs.

Groups of correlates can be assimilated into larger groups which indicate an individual category.
In reality this can be performed by correlation using Hebbian learning. \textit{Intra-concept} connections
between correlated correlates are enhanced and connections between excited correlates and other

\(^2\)Edelman avoids an explicit correlate representation. Instead, he connects the cortical maps directly to one another.
The theory outlined in this chapter can be applied also to this kind of network for which the concept is represented
implicitly in the connections between the quantity-spaces.

\(^3\)Although direct mappings of sensor information into the superior colliculus has been shown to preserve topography
in cats and snakes (Stein and Meredith, 1986; Hartline and L Kess, 1978) mappings involving inference can often fail
to preserve topography. For example, an orderly representation of space for the FM Bat \textit{Eptesicus fuscus} was not
evident (Jen et al., 1984). The movement of the bat pinnae during echolocation was offered as a reason for this.
correlates are gradually suppressed. Thus, concept values and their properties (i.e., correlations with the sensor parameter values) emerge dynamically through a process of unsupervised, competitive selection between groups of correlates. For example, there are many qualitative behaviours that correspond to each of the values (i.e., convex and concave) of the surface curvature concept. To categorise with correlated concepts, connections between groups have to be strengthened. This may be achieved, for example, by temporally correlating stimuli which means that, within the Hebbian learning context, correlates remain excited for a (scale dependent) time interval. Of course, temporal correlation is only one rule for categorisation. However, only temporal correlation will be considered in the remainder of this preliminary sketch.

Based on the preceding discussion, we propose the following (primary and secondary) repertoire generation algorithm:

- $N$ sensor cue spaces are represented as $N$ quantity-spaces $ QS_n \ (1 \leq n \leq N)$.

- Each concept correlate is associated with at most one value from each quantity-space. This avoids situations of gross ambiguity when different stimuli corresponding to different categories excite the same correlate. Let $B_{nsd}$ and $B_{n,d'}$ be the set of correlates associated with parameter space $ QS_n $ and its quantity-values $Q_s$ and $Q_{d'}$ respectively. The unique assignment of correlate to parameter quantity-space means that $(\forall n, i, i') B_{nsd} \cap B_{n,d'} = \emptyset$. A specific instance of $N$ stimuli excites correlates $\bigcap_{n \in N} B_{n,i_n}$ and when two sets of stimuli differ by any individual stimuli then they must excite distinctly different correlates since:

$$ \left( \bigcap_{n \in N} B_{n,i_n} \right) \cap \left( \bigcap_{n \in N} B_{n,i'_n} \right) = \bigcap_{n \in N} (B_{n,i_n} \cap B_{n,i'_n}) = \emptyset. $$

The primary repertoire is constructed by randomly assigning to each correlate a single quantity-space value from each parameter space.

- The correlates are pair-wise connected by Hebbian links $\omega_{k,j}$ and all link values are initially set to zero. Suppose the current stimuli excite correlate $b$ then those correlates temporally related to $b$ are correlated by increasing their link value by 0.1 towards the saturation value 1. Links between $b$ and those correlates not temporally correlated are decreased by 0.05 towards 0. It is important to notice the positive undermining aspect of this revision rule. An alternative approach, adopted by Edelman, is to include a gradual time dependent decay of link value. This, however, can lead to highly unstable networks. For the remainder of this chapter it is assumed that the correlates associated with the current stimuli are correlated with those correlates associated with the previous two stimuli sets. A link is “live” and its correlates are assigned to the same group when its value exceeds 0.2.

The potential utility of this algorithm is demonstrated by the following experiment.

**Experiment**

The advantage of Edelman’s approach over more conventional psychological categorisation techniques (classical or probabilistic) is that, in Edelman’s approach, the categories or even the number of categories are not specified a priori. This experiment illustrates the emergence of categories by temporal correlations between recurring stimuli. Suppose that the observer is initially unaware of four concepts $+\otimes$, $-\otimes$, $+\odot$ and $-\odot$ operating in environments $\otimes$ or $\odot$. We assume that the observer
can focus on a single concept for an extended period of time so that randomly generated observations associated with each concept can be correlated.

```
+  +  -
+ - - +
- + - +

⊗ concepts  ⊗ concepts
```

Table 7.1: Qualitative sign relationships for ⊗ and ⊗ functions.

The categorisation algorithm developed above is illustrated on a non-trivial problem: learning the non-linear qualitative ⊗ table and then readapting to the qualitative ⊗ function. The observation pairs associated with each concept are given in Table 7.1. Two sensor modalities, A and B, each return a value of - or +. For the first period of time two distinct artifacts are encountered, +⊗ and -⊗, and each artifact has associated with it a set of possible sensor cues: +⊗ if and only if observations pairs are (+,+) or (-,-), and -⊗ if and only if observation pairs are (+,-) or (-,+). The desirable outcome of the algorithm would be to distinguish the artifacts and categorise {(+,+),(-,-)} together and {(+,-),(-,+)} together. Subsequently, the sensing process enters a new phase when the ⊗ process is monitored for which the sensors receive new stimuli corresponding to completely new concepts, +⊗ and -⊗: +⊗ if and only if observation pairs are (+,+) or (+,-), and -⊗ if and only if observation pairs are (-,-) or (-,+). Ideally, the algorithm would adapt to this new context and recategorise the correlate groups accordingly.

```
<table>
<thead>
<tr>
<th>Time-step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>+⊗</td>
<td>+⊗</td>
<td>+⊗</td>
<td>+⊗</td>
<td>-⊗</td>
<td>-⊗</td>
<td>-⊗</td>
<td>-⊗</td>
</tr>
<tr>
<td>Stimuli (A,B)</td>
<td>(1,1)</td>
<td>(0,0)</td>
<td>(0,0)</td>
<td>(1,1)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(0,1)</td>
</tr>
</tbody>
</table>
```

Figure 7.1: Sensor cue sequence and their causal concepts. Two stimuli pairs correlate temporally if they occur frequently in each others neighbourhood. For example, {(0,0),(1,1)} and {(0,1),(1,0)} are correlated.

The parameter quantity-spaces for A and B are both partitioned into the signed space {-, +} and a concept space C comprises twenty correlates (labelled 0 through to 19). Eighty pairs of observations (A,B) are taken. The first forty correspond to the ⊗ environment, for which Qmag(C) = +⊗ for the first twenty observation pairs and then Qmag(C) = -⊗ for the remainder. Subsequently, forty observation pairs are taken in the ⊗ environment, for which the first twenty pairs correspond to Qmag(C) = +⊗ and the final twenty to Qmag(C) = -⊗. Each observation pair is chosen randomly from a (context dependent) possible choice of two pairs.

We describe, in detail, one trial of the above scenario. In this particular case, the following primary repertoire was randomly generated:

Figure 7.2 shows the evolution of the secondary repertoire over time. As we can see from the figure, at time-step t = 10 all four groups are distinct and uncorrelated; no conceptual categories have yet been identified. However, at time t = 20 groups G1 and G2 are associated with the same category. This category corresponds to the +⊗ concept. This categorisation persists until time-step t = 40, a long time after +⊗ ceases to be observed. Within the ⊗ context groups G3 and G4 are associated at time-step t = 25 and while the ⊗ environment persists (until time-step t = 40)
the stimuli are correctly categorised. At time-step $t = 40$ the environment changes to $\oplus$ stimuli correlations. The algorithm recognises a contextual change relatively quickly and from $t = 42$ until $t = 58$ no groups are categorised. Eventually groups G2 and G3 are combined close to time-step $t = 60$. Group G2 is no longer associated with G1 and the new combined group corresponds to the $+\oplus$ concept. A transition period corresponding to the change of $Qmag(C)$ to $-$ eventually leads to a further combining of groups G2 and G3 corresponding to the $-\oplus$ concept.

<table>
<thead>
<tr>
<th>Group Identifier</th>
<th>Stimulus $(A, B)$</th>
<th>Group correlate components</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>$(-, -)$</td>
<td>(2 12 15 16 19)</td>
</tr>
<tr>
<td>G2</td>
<td>$(+, +)$</td>
<td>(4 7 10 11 14 17)</td>
</tr>
<tr>
<td>G3</td>
<td>$(+, -)$</td>
<td>(0 3 5 6 8 13)</td>
</tr>
<tr>
<td>G4</td>
<td>$(-, +)$</td>
<td>(1 9 18)</td>
</tr>
</tbody>
</table>

Figure 7.2: Secondary group formation over time. Each plot represents a group. The assimilation of groups occurs where plots converge. For example, at time-step $t = 10$ groups G1 and G2, representing observations $(0, 0)$ and $(1, 1)$, are combined and correspond to the higher level concept $+\oplus$. D is the degenerate group $G1 \cup G2 \cup G3 \cup G4$.

These results clearly show the potential application of Edelman’s Theory of Neural Group Selection to the formation of context dependent concepts within the qualitative framework. Further considerations would concern (i) the reassignment of under-used correlates to new parameter space values and (ii) upper and lower correlation values for primary concepts imprecisely identified.

So far we have considered the boot-strap case of conceptual formation: building new concepts from predefined concepts (Smith’s famous ‘pet-fish’). By accumulating corresponding values, the robot may refine its concepts. A limited memory space requires that only context relevant and salient concepts are learnt and maintained. Concurrent observations from the sensor suite offer a continuous supply of corresponding values between sensor quantity-spaces. Thus, there exists an opportunity to increase model precision on the fly. Simultaneous model building and model usage
was demonstrated in this thesis via the simultaneous classification of cylinder types based on their power-range profiles. It is obvious, however, that maintaining consistency between all observed corresponding values can become computationally infeasible as the number of observations accrue. Thus, one area of future research would be towards mechanisms which identify salient landmarks. These landmarks must be sufficient to capture all distinctions required for context sensitive data fusion tasks.

We mentioned at the beginning of this chapter that abstraction often leads to unacceptable inference ambiguity. How would the adapted-TNGS approach help to refine the QR representation and reduce the problem of ambiguity? The context determines the appropriate refinement of ambiguous inferences. When an ambiguous behaviour is indicative of a unique inferred hypothesis then TNGS will group together both ambiguous and unambiguous inferred correlates. Within a context, each group describes a single qualitative entity and the qualitative value is assumed be identical for all correlates within the group. Thus, ambiguous inference may be labelled (i.e. refined) using the qualitative labelling known to the unambiguous inferences within the same group.
Bibliography


Bibliography


Appendix A

Higher Order Sonar Reflection Model

Using the idea of images (Kuc and Siegel, 1987) we can construct a model for multiple reflections within an acute corner. A pressure wave released towards an acute corner may undergo numerous reflections alternating between walls before returning to the transducer. The aim is to calculate the distance travelled by the signal and the incident angle of the returning signal on the transducer. We can simplify our representation of the problem by noting the following property: rays traced from a point source after reflection in a plane are equivalent to rays emitted from a virtual image placed perpendicularly behind the plane (see Figure A.1). Thus, we can replace the planes of the system

![Diagram](image)

Figure A.1: The virtual image formed by reflection in a plane.

with images corresponding to the pulse reflections.

1. Reflect the transducer in each plane to form a set of images $I_0$.

2. Reflect each member of $I_k$ in all planes thus forming $I_{k+1}$, the set of new images.

3. Remove those members of $I_{k+1}$ for which the range to the transducer is equal to or less than the range to the transducer from a predecessor image.

4. Repeat from Step 2 until $I_n$ is empty.

This algorithm generates the complete set of images consistent with the differential wave equations and boundary conditions. In Figure A.2, $P_1$ is the image of the transducer $P$ reflected through wall OC and $P_2$ is the image of $P_1$ reflected through OD. Each reflection of the pressure wave on a wall corresponds to the reflection in that wall of the previously formed image. Image $P_n$ corresponds
to the last reflection before the wave is received by the transducer and has positional bearing $\mu_n$ relative to $OP$ and range $R_n$.

In what follows, the $n^{th}$ image formed by the $n^{th}$ reflection has bearing $\lambda_n$ relative to $\overline{OP}$. Variables are labelled according to whether they describe an image formed by an odd or an even number of reflections. We will derive expressions for those rays formed by an initial reflection of the transducer in the OC plane. The first image corresponds to an odd order reflection and has bearing $\lambda_{0}^{\text{odd}} = -2\epsilon$. Subsequent image bearings are determined incrementally:

\[
\begin{align*}
\lambda_n^{\text{even}} &= 2\lambda - \lambda_{n-1}^{\text{odd}} - 2\epsilon, \\
\lambda_n^{\text{odd}} &= -\lambda_{n-1}^{\text{even}} - 2\epsilon
\end{align*}
\]  

(A.1) 

and therefore:

\[
\begin{align*}
\lambda_n^{\text{even}} &= n\lambda, \\
\lambda_n^{\text{odd}} &= -(n-1)\lambda - 2\epsilon.
\end{align*}
\]  

(A.2)

Consequently, since $||OP_n|| = ||OP||$ for all $n$:

\[
\begin{align*}
\mu_n^{\text{even}} &= \pm \frac{\pi - n\lambda_2}{2}, \\
\mu_n^{\text{odd}} &= \pm \left( \frac{\pi - (n \mp 1)\lambda}{2} \mp \epsilon \right)
\end{align*}
\]  

(A.3) 

(A.4)

where the upper and lower signs correspond to initial reflection from OC and OD respectively.\(^1\)

Since $\triangle OPP_2$ and $\triangle OPP_3$ are isosceles, the range measurement is given by:

\[
R_n = ||OP|| \cos \mu_n.
\]  

(A.5) 

(A.6)

\(\text{\textsuperscript{1}}\text{The value of } \mu \text{ for initial reflection in OD is obtained by changing } \epsilon \text{ to } \lambda - \epsilon \text{ and changing the sign of the result.}\)
A.1 Superposition of Point Sources

The previous section developed single ray dynamics. However, since we are interested in finite transducers we develop in this section, a relationship between the orientation of the transducer and its images formed by reflection in the planes. We show that the angle subtended by the image on the transducer $\alpha_i$ and the angle subtended by the transducer on the image $\alpha_i$ are functions of the orientation $\mu$ of the image relative to the transducer and the orientation $\gamma$ of the corner relative to the transducer (see Figure A.3).

\[ \lambda_{n_{n''}}^{\text{even}} = n\lambda, \]
\[ \lambda_{n_{n'}^{\text{odd}}} = -(n - 1)\lambda - 2\epsilon, \]
\[ \lambda_{n_{n''}}^{\text{even}} = n\lambda - \beta, \]
\[ \lambda_{n_{n'}^{\text{odd}}} = -(n - 1)\lambda - 2(\epsilon - \beta) - \beta. \]

Thus:

\[ \lambda_{n_{n''}}^{\text{even}} - \lambda_{n_{n'}}^{\text{even}} = -\beta, \]
\[ \lambda_{n_{n'}^{\text{odd}}} - \lambda_{n_{n''}}^{\text{odd}} = \beta. \]

Therefore, as the transducer rotates, odd order images turn opposite to the transducer and even order images turn in the same direction as the transducer. The angle $\gamma$ is the orientation of the normal of the transducer to the corner centre and $\mu_n - \gamma$ is the orientation of the centre of the $n^{th}$ image relative to the normal direction of the transducer. The angular relationships for image and transducer for both even and odd images are (see Figure A.4):

- **Odd Order Reflections.** Orientation of image is $\alpha_i = \mu_n - \gamma$ relative to transducer. Transducer is orientated at $\alpha_i = \gamma - \mu_n$ relative to image.

- **Even Order Reflections.** Orientation of image is $\alpha_i = \mu_n - \gamma$ relative to transducer. Transducer is orientated at $\alpha_i = \mu_n - \gamma$ relative to image.

In the remaining sections, we demonstrate the validity of the model developed in the previous
sections by showing, firstly, its ability to predict the range information obtained from a sonar scan of an acute corner and then, secondly, its ability to predict the motion of RCDs for multiple reflections from non-parallel walls.

### A.2 Experimental Results

We verify the model developed in the previous sections for a circular Polaroid transducer by simulating the real-data profile in Figure 2.5. The transducer is subject to an oscillatory, exponential driving force $p(t)$ of frequency $f_R$ ($f_R \approx 50kHz$) (Kuc and Siegel, 1987):

$$p(t) = \exp \left( -\frac{t^2}{2\sigma^2} \right) \cos(2\pi f_R t).$$  \hspace{1cm} (A.14)

The radial pressure $p_R$ a distance $z$ from the transducer and angle $\alpha$ to the normal of the transducer is given by the following convolution equation:

$$p_R(t, z, \alpha) = \frac{p(t)}{z} * R \left( t - \frac{z}{c}, z, \alpha \right)$$  \hspace{1cm} (A.15)

where $R(t, z, \alpha)$ is the fourier transform of the transducer response function $R(f, z, \alpha)$ (Arfken, 1970):

$$R(f, z, \alpha) = \frac{ca}{f \sin \alpha} J_1 \left( \frac{2\pi f a}{c} \sin \alpha \right)$$  \hspace{1cm} (A.16)

c is the velocity of sound in air ($\approx 300ms^{-1}$) and $a$ is the radius of the transducer ($\approx 2cm$). In the Framhofer limit, the receiver intercepts a plane wave from the emitter. Since the receiver has the same geometry as the emitter then the response function for the receiver is similar to that of the emitter. The receiver pressure $P_R$ is given by:

$$P_R(t, z, \alpha_l, \alpha_r) = \text{Re} \int_{-\infty}^{\infty} df P_T(f, z, \alpha) r(f, z, \alpha) e^{i2\pi f(t-\xi)}. $$  \hspace{1cm} (A.17)
Thus
\[
p_R(t, z, \alpha_t, \alpha_r) = p_T \left( t - \frac{z}{c}, z, \alpha \right) * R \left( t - \frac{z}{c}, \alpha \right)
\]
\[
= \frac{p(t - \frac{z}{c})}{z} * R \left( t - \frac{z}{c}, \alpha_R \right) * R \left( t - \frac{z}{c}, \alpha_T \right)
\]
\[
= \frac{p(t - \frac{z}{c})}{z} * R_{T/R} \left( t - \frac{z}{c}, \alpha_R, \alpha_T \right)
\]
(A.18)

where \( R_{T/R}(t, \alpha_R, \alpha_T) = R_1(t, \alpha_R) * R_2(t, \alpha_T) \) is the transmitter/receiver response and \( \alpha_R \) and \( \alpha_T \) are defined in Section A.1.

The ultrasonic transducer detects a signal when the logarithm of the time integrated, gain controlled pulse power input exceeds a threshold value. Figure A.5 shows the receiver log, time integrated power distribution as a function of the transducer direction relative to the corner obtained by simulation for an acute corner of angle 0.87 rad. The horizontal dashed line in this figure indicates an empirically determined threshold value which yields the simulated RCDs in Figure A.5. Taking

![Figure A.5: Theoretical first, second and third order received log energy as function of transducer bearing relative to corner. RCDs (see Figure A.6) correspond to bearings where the power is greater than the transducer threshold displayed as dashed horizontal line.](image)

the occlusion effects of lower order reflections into account, we find a close correspondence between the simulated corner RCDs and those obtained physically (Figure 2.5).

### A.3 Limitations of Model: General Polygonal Surfaces

In this section, we show by a simple counter example that the qualitative description for the RCD behaviors described in Section 2.3.2 do not generalise to reflections from arbitrary polygonal walled structures. We will consider a third order reflection and show, using the theory of images, that the reflection moves opposite to that of the robot. This clearly contradicts the rule that odd order reflectors maintain constant bearing. In Figure A.7, the transducer moves from position \( X_1 \) to \( X_2 \). At each location the transducer fires at plane \( F_1 \). The signal which is eventually received by the
transducer bounces off \( F_1 \) close to point \( P_1 \), and bounces off planes \( F_2 \) and then \( F_3 \) before returning to the transducer. Thus, the reflection is third order. The images formed by the various specular reflections are denoted by primed variables in the figure. Thus, \( X'_1 \) is the image formed by the reflection of \( X_1 \) in plane \( F_1 \) and \( X''_1 \) is the image formed by the reflection of \( X'_1 \) in the corner formed by planes \( F_2 \) and \( F_3 \) (Kuc and Siegel, 1987). When the robot moves from \( X_1 \) to \( X_2 \) it sees its image move in a retrograde motion from \( X''_1 \) to \( X''_2 \). This is counter to the expected constant bearing behaviour of odd order reflections.
Appendix B
The QSim Acoustic Flow Model

This chapter details the qualitative model used in conjunction with QSIM version 2.0 (Kuipers, 1994).

(defun wrapphi1(state) (defun wrapphi2(state)
  (create-transition-state (create-transition-state
    :from-state state :from-state state
    :to-qle intra-rod :to-qle intra-rod
    :inherit-qmag :rest :inherit-qmag :rest
    :inherit-qlir :rest :inherit-qlir :rest
    :assert '(Phi (behind* inc))) )
  :assert '((Phi (behind dec))))
  
(defun wrapphi3(state) (defun wrapphi4(state)
  (create-transition-state (create-transition-state
    :from-state state :from-state state
    :to-qle intra-rod :to-qle intra-rod
    :inherit-qmag :rest :inherit-qmag :rest
    :inherit-qlir :rest :inherit-qlir :rest
    :assert '(Th (behind std))) )
  :assert '((Th (behind dec))))
  
(defun wrapphi5(state) (defun wrapphi6(state)
  (create-transition-state (create-transition-state
    :from-state state :from-state state
    :to-qle intra-rod :to-qle intra-rod
    :inherit-qmag :rest :inherit-qmag :rest
    :inherit-qlir :rest :inherit-qlir :rest
    :assert '(Th (behind std))) )
  :assert '((Th (behind dec))))
  
(define-qle intra-rod
  (text "Simple rod motion model")
  (quantity-spaces
    (R (inf 0)) "R:range")
    (TH (behind right 0 left behind) "Th:bearing")
    (D (inf min 0) "r:reflector radius")
    (Xdot (inf min 0) "Xdot:\(\frac{dR}{dt}\)")
    (Thdot (min 0 inf) "Thdot:\(\frac{dTh}{dt}\)"
    (X (inf min 0 inf) "X:R+D")
    (Y (inf min 0 inf) "Y:Speed")
    (Angy (inf min 0 inf) "Angv:Gyro")
    (Cosx (+ one 0 one))
    (Sinx (+ -one one)))
  (no-new-landmarks V D NCosth Sinth Xdot
    Thdot X Y Angv Tangy R Th)
  
  (constraints
    ((constant Angry))
    ((constant D))
    ((constant V)))
((d/dt X \dot{x}))
((d/dt T \dot{\theta}))
((\text{add} \ \text{Ang} \ \theta \dot{\theta} \ \text{X} \ \text{Y}))
((\text{mult} \ X \ \text{Tang} \ \theta))
((\text{mult} \ V \ \sin \theta \ \text{Y}))
((\text{mult} \ V \ \text{Ncos} \ \theta \ \dot{x})))
((\text{MI+} \ \text{T} \ \text{Sin} \ \text{X} \ \text{right} \ \text{left}) \ (\text{right} \ -1) \ (0 \ 0) \ (\text{left} \ 0))
((\text{MI-} \ \text{T} \ \text{Sin} \ \text{X} \ \text{left} \ \text{behind}) \ (\text{left} \ 0) \ (\text{behind} \ 0))
((\text{MI-} \ \text{T} \ \text{Sin} \ \text{X} \ \text{behind} \ \text{right}) \ (\text{behind} \ -1) \ (\text{right} \ 0))
((\text{MI+} \ \text{T} \ \text{Ncos} \ \text{X} \ \text{behind} \ \text{left}) \ (\text{left} \ 0) \ (\text{behind} \ 1))
((\text{MI-} \ \text{T} \ \text{Ncos} \ \text{X} \ \text{behind} \ \text{right}) \ (\text{behind} \ 0) \ (\text{right} \ 0))
((\text{transitions})
((\text{Sin} \ \text{X} \ \text{inc}) \ -> \ \text{endbeh})
((\text{Sin} \ \text{X} \ \text{dec}) \ -> \ \text{endbeh})
((\text{Ncos} \ \text{X} \ \text{inc}) \ -> \ \text{endbeh})
((\text{Ncos} \ \text{X} \ \text{dec}) \ -> \ \text{endbeh})
((\text{D} \ \text{in} \ \text{inc}) \ -> \ \text{endbeh})
((\text{D} \ \text{in} \ \text{dec}) \ -> \ \text{endbeh})
((\text{R} \ \text{in} \ \text{inc}) \ -> \ \text{endbeh})
((\text{R} \ \text{in} \ \text{dec}) \ -> \ \text{endbeh})
((\text{Th} \ \text{behind} \ \text{inc}) \ -> \ \text{wrapt})
((\text{Th} \ \text{behind} \ \text{dec}) \ -> \ \text{wrapt})
((\text{Th} \ \text{behind} \ \text{stn}) \ -> \ \text{wrapt})
((\text{unreachable values})
((\text{Y} \ \text{inf}) \ (\text{Y} \ \text{inf})
((\text{X} \ \text{inf}) \ (\text{X} \ \text{inf})
((\text{V} \ \text{inf}) \ (\text{V} \ \text{inf}))

\text{Layout} ((\text{D} \ \text{R} \ \text{T}))

\text{(defun rod-track ()}
\text{(qsim-clean-up)}
\text{(setq *initial+}
\text{ (make-new-state :from-ple intra-rod}
\text{ :text "Start rod track"}
\text{ :rsim (make-sim :Q2-constraints nil}
\text{ :state-limit 40})
\text{ :assert-values (1((\text{Th} ((0 \ left) \ nil))}
\text{ (\text{V} ((0 \ inf) \ nil))}
\text{ (\text{Angy} (0 \ std))))})

\text{(qsim *initial+}
\text{(qsim-bounded-display *initial+))}
Appendix C

Scale-Space Abstraction

C.1 Introduction

Mathematical morphology is a set theoretic filter which removes structure with dimensions below a set scale. Unlike Gaussian or mean filters the structure is removed cleanly and is not blurred across the remaining image. Such a filter is ideal for qualitative methods where the plant models may capture only some of the structure of the underlying image. Morphological filters are built from sequences of two basic operations: dilation and erosion. A continuous one dimensional image in \( \mathcal{R} \) is seen to divide \( \mathcal{R} \) into two sets each side of the image. A filter structure \( V \) is defined in \( \mathcal{R} \) which operates on any one of these sets. The filtering operation transforms a point \( x \) in the original image into:

\[
x_v = x + v
\]

where:

\[
v \in V.
\]

The basic filtering operations are defined thus:

- **Dilation.** \( A \oplus V = \bigcup_{v \in V} x_v \).
- **Erosion.** \( A \ominus V = \bigcap_{v \in V} x_v \).
- **Opening.** \( A \circ V = (A \ominus V) \oplus V \).
- **Closing.** \( A \bullet V = (A \oplus V) \ominus V \).

We can now define two further operations which maintain the underlying scale of the feature but remove structure which are smaller than the basic filtering element.

- **Close-opening.** \( A CO V = A \bullet V \circ V \).
- **Open-closing.** \( A OC V = A \circ V \bullet V \).

Close-opening introduces positive bias and open-closing negative bias with respect to the mean. A compromise morphological filter explored by (Noble, 1989) for symmetric noise distributions is the COMOC filter which calculates the mean of the close-open and open-close operations on the original noisy signal (for additive noise). It was shown in (Noble, 1989) that, for a constant signal, COMOC is unbiased in the sense that the mean of the filtered signal \( \hat{x} \) approximates arbitrarily closely to the true signal \( x_T \).

\[
E(\hat{x}) = x_T.
\]
We extend this proof and show that, for a signal subject to additive, symmetric noise, the COMOC filter is actually unbiased where the true signal is locally rotationally symmetric. However, we then present examples which show that the filter is actually biased in the general case and propose that local rotation symmetry is a necessary requirement for COMOC to be unbiased.

C.2 Proof of COMOC Unbiased Property around Points of Symmetry

In (Noble, 1989), a proof of the unbiased noise filtering property of COMOC for an underlying constant process was presented. Here, we present a more general proof which shows that COMOC is unbiased at symmetric points for single dimension process functions subject to symmetric noise. The proof uses the properties of close-open and open-close morphological filtering derived in (Stevenson and Arce, 1987) for the constant set $\Omega_n$ and the impulse set $\Upsilon_n$ but the output statistics described in Section C therein are generalised to non-constant functions.

A one-dimensional function $y_i = y(x_i)$ is subject to additive noise and is sampled at discrete points $i$. An observation of state $y_i$ is $z_i$. The noisy discrete signal is filtered using open-close (i.e. OC) and close-open (i.e. CO) filters of filter width order $n$. The probability density distributions at $i$ for the OC and CO filters are $F^{oc}$ and $F^{co}$ respectively. A threshold $T$ is chosen on $z_i$ in the noisy image and constant and impulse sets are identified in a similar manner to (Stevenson and Arce, 1987). However, unlike (Stevenson and Arce, 1987), the underlying process function is not constant. Hence, we define $t_i$ to be the residual between $y_i$ and the threshold $T$ (see Figure C.1). From

![Figure C.1: Illustration for proof of comoc unbiasedness.](image)

(Stevenson and Arce, 1987) (Equations 3.20 and 3.35, p1297) the cumulative distribution functions for the close-open and open-close filtered signals are:

$$F^{co}(y_i + t_i) = Pr\{z_i \in \Omega_n^0\} + Pr\{z_i \in \Upsilon_n^1\}. \quad (C.4)$$

Similarly:

$$F^{oc}(y_i + t_i) = 1 - Pr\{z_i \in \Omega_n^1\} + Pr\{z_i \in \Upsilon_n^0\}. \quad (C.5)$$

We will now derive expressions for $Pr\{z_i \in \Omega_n\}$ and $Pr\{z_i \in \Upsilon_n\}$. In the following we define:

$$\alpha_k = F(y_k + t_k) \quad (C.6)$$

where $t_k$ is defined so that $y_k + t_k = y_i + t_i = T$. The probability that $z$ belongs to the negative
constant set $\Omega_n^0$ is given by (see Figure C.2):

$$Pr\{z_i \in \Omega_n^0\} = \sum_{m=n+1}^{\infty} \sum_{j=0}^{m} \prod_{k=i-j}^{i-j+m} \alpha_k [1 - \alpha_{i-j-1}] [1 - \alpha_{i-j-m+1}]$$  \hfill (C.7)

and the probability that $z$ belongs to the positive constant set $\Omega_n^1$ is given by:

$$Pr\{z_i \in \Omega_n^1\} = \sum_{m=n+1}^{\infty} \sum_{j=0}^{m} \prod_{k=i-j}^{i-j+m} [1 - \alpha_k] \alpha_{i-j-1} \alpha_{i-j-m+1}.$$  \hfill (C.8)

The probability that $z$ belongs to the positive impulse set $\Upsilon_n^1$ is given by (Stevenson and Arce, 1987) Equation 3.26, p1297 (see Figure C.3):

$$Pr\{z_i \in \Upsilon_n^1\} = Pr\{z_i \in v_1, \Upsilon_n^1\} + \sum_{m=2}^{n} Pr\{z_i \in v_m, \Upsilon_n^1\}.$$  \hfill (C.9)

Now:

$$Pr\{z_i \in v_1, \Upsilon_n^1\} = [1 - \alpha_{i-1}] \prod_{k=i-n-2}^{i-1} \alpha_k \prod_{k=i+1}^{i+n+2} \alpha_k$$  \hfill (C.10)
and, for $m \geq 2$:

$$Pr\{z_i \in v_m, \gamma_n^1\} = \sum_{j=0}^{m} [1 - \alpha_{i-j}] [1 - \alpha_{i-j+m}] \left[ \sum_{k=0}^{i-j-1} \alpha_k \right] \left[ \sum_{k=i-j+m+2}^{i+n+2} \alpha_k \right].$$

(C.11)

Similarly, the probability that $z$ belongs to the negative impulse set is:

$$Pr\{z_i \in v_1, \gamma_n^0\} = \alpha_i \left[ \prod_{k=0}^{i-1} (1 - \alpha_k) \right] \left[ \prod_{k=i+1}^{i+n+2} (1 - \alpha_k) \right].$$

(C.12)

and, for $m \geq 2$:

$$Pr\{z_i \in v_m, \gamma_n^0\} = \sum_{j=0}^{m} \alpha_{i-j} \alpha_{i-j+m} \left[ \sum_{k=0}^{i-j-1} (1 - \alpha_k) \right] \left[ \sum_{k=i-j+m+2}^{i+n+2} (1 - \alpha_k) \right].$$

(C.13)

From Equations C.4 and C.5 we note that:

$$F^{\cos}(y_i + t_i) = 1 - F^{\cos}(y_i + t_i)$$

(C.14)

where star (i.e. *) is used to indicate the replacement of $\alpha_k$ with $1 - \alpha_k$.

If the signal is symmetric under $\pi$ rotation about $i$ then:

$$t_{i+j} = -t_{i-j}.$$  

(C.15)

Consequently, if $y(*)$ is corrupted by symmetric, uncorrelated noise then at each point, the cumulative distribution function $F$ for the original image satisfies:

$$1 - F(y_i + t_{i+j}) = F(y_i + t_{i-j}) = F(y_i - t_{i+j}).$$

(C.16)

If we replace $t_k$ by $-t_k$ in Equation C.7 we obtain Equation C.8. Similarly for Equations C.10 and C.12 and Equations C.11 and C.13. Thus, by Equations C.4 and C.5:

$$F^{\cos}(z_i < y_i + t_i) = F^{\cos}(z_i > y_i - t_i).$$

(C.17)

Hence:

$$F^{\cos}(z_i - y_i < t_i) = F^{\cos}(z_i - y_i > -t_i).$$

(C.18)

### C.3 COMOC as an Unbiased Mean Filter

We shall show that the condition in Equation C.18 is sufficient for the COMOC filter to be unbiased. The COMOC filtered signal at index position $i$ is the average of the close-open and open-close filtered signals:

$$z_i^{\text{comoc}} = \frac{z_i^{\text{co}} + z_i^{\text{oc}}}{2}.$$  

(C.19)
Therefore:
\[ z_{i}^{\text{comoc}} - y_i = \frac{(z_i^{\infty} - y_i) + (z_i^{\infty} - y_i)}{2}. \]  
(C.20)

From (Mood et al., 1982) the expected value \( E(X) \) for an arbitrary random variable \( X \) is:
\[ E(X) = \int_{0}^{\infty} [1 - F_X(x)]dx - \int_{-\infty}^{0} F_X(x)dx. \]  
(C.21)

Hence:
\[ E^{\infty}(Z - Y) = \int_{0}^{\infty} [1 - F^{\infty}(t_i)]dt_i - \int_{-\infty}^{0} F^{\infty}(t_i)dt_i. \]  
(C.22)

Also:
\[ E^{\infty}(Z - Y) = \int_{0}^{\infty} [1 - F^{\infty}(-t_i)]dt_i - \int_{-\infty}^{0} F^{\infty}(t_i)dt_i. \]  
(C.23)

Using Equation C.18 and the tautology \( P(a > b) + P(a < b) = 1 \) gives:
\[ E^{\infty}(Z - Y) = \int_{0}^{\infty} F^{\infty}(-t_i)dt_i - \int_{-\infty}^{0} [1 - F^{\infty}(-t_i)]dt_i. \]  
(C.24)

Transforming \( t_i \) to \( -t_i \) in Equation C.24 gives:
\[ E^{\infty}(Z - Y) = -E^{\infty}(Z - Y). \]  
(C.25)

Hence, by Equation C.20, \( E^{\text{comoc}}(Z - Y) = 0 \) and therefore, the COMOC filter is an unbiased mean estimator.

### C.4 COMOC as an Unbiased Median Filter

From Equation C.18 we obtain, when \( t_i = 0 \):
\[ F^{\infty}(z_i < y_i) = F^{\infty}(z_i > y_i). \]

Therefore:
\[ F^{\text{COMOC}}(z_i < y_i) = \frac{F^{\infty}(z_i < y_i) + F^{\infty}(z_i < y_i)}{2} = \frac{F^{\infty}(z_i > y_i) + F^{\infty}(z_i > y_i)}{2} = F^{\text{COMOC}}(z_i > y_i). \]

Thus, the COMOC filter is also an unbiased median estimator.

### C.5 Limitations of the COMOC filter

In the previous section we showed that the COMOC filter is unbiased when the true signal is locally rotationally symmetric. In this section, we show with the aid of a number of examples that the filter is biased in the general case. Figures C.4 and C.5 show a quadruple-step function and a sigmoid
function respectively superimposed with additive noise of various variance values. The upper dashed line shows the result of close-opening the original noisy signal, the lower dashed line is the result of open-closing this signal whereas the central dashed line shows the result of COMOC filtering. These figures clearly show the rounding effect of the filter at points of asymmetry.

![Diagram](image)

(a) Element width=4, standard error=1.0  
(b) Element width=10, standard error=1.0  
(c) Element width=4, standard error=10.0  
(d) Element width=10, standard error=10.0

**Figure C.4:** Quadruple-step function showing COMOC filter bias. Averaged over 200 trials, noise standard errors of 1.0 and 10.0 and filtering elements of width 4 and 10.
Figure C.5: Sigmoid and wedge functions showing COMOC filter bias. Averaged over 200 trials, noise standard error 1.0 and filter element width 10.
Appendix D

A Critique of R. T. Cox’s Paper “Probability, Frequency and Reasonable Expectation”

D.1 Introduction

In 1946, R. T. Cox (Cox, 1946) unified two ideas, the idea of frequency in an ensemble and the idea of reasonable expectation (i.e., subjective probabilities) and showed that the rules of probable inference are credited by common sense with a wider validity than can be established by deducing them from the frequency definition of probability. By specifying a set of plausible assumptions about possible reasoning Cox showed that the axioms of probability theory follow necessarily. Jaynes (Jaynes, 1979) used this result to conclude that:

“any method of inference in which we represent degrees of plausibility by real numbers, is necessarily either equivalent to Laplace’s, or inconsistent.”

Here Jaynes is referring to Laplace’s discovery of the famous “Bayes Theorem” (in his memoir of 1774 on the “Probabilities of Causes” (Jaynes, 1979)) and to the additive nature of probability theory. Jaynes claims that any method using any other form (such as max operations) must be inconsistent. Of course Cox’s theorem is a significant argument against all calculi other than probability theory whether applied to subjective information or frequencies. Justifiably, Cox’s argument still causes concern in AI (Pearl, 1988; Kosko, 1992; Krause and Clark, 1993). In defence of the use of max and min operations in Fuzzy logic theory and Possibility theory, Kosko (Kosko, 1992) attempted to dismiss Cox’s argument by noting that Cox’s proof assumes that the uncertainty combination operators are continuously twice differentiable (see (Kosko, 1992), page 203). Since max and min are not twice differentiable then technically, Cox’s theorem does not apply. Thus Kosko asserts that Cox’s theorem is indifferent to reasoning theories which employ max operators instead of the usual summation operators of probability theory. In this chapter we will see that twice differentiable functions can be found which are arbitrarily close to the max and min operators, thus negating Kosko’s indiffference argument.

D.2 The Critique

Our assertion is that Cox’s analysis is less constraining than originally thought and that Cox’s theorem is consistent with theories which use max operations. To demonstrate this we first show that Cox proves the necessity of a meta-theory (the p-norm meta-theory). We will refer to a generic notion of evidence E which needn’t have a statistical nor a subjective probabilistic interpretation. We will use equivalence relationships from first order logic to demonstrate that, if \( E(A_j) \) is the evidence
assigned to proposition $A_i$, then all algebras of evidential reasoning must satisfy a $p$-norm theory:

$$E(a \lor b) = \left[ E(b)^p + E(c)^p - E(ab)^p \right]^{\frac{1}{p}}$$  \hspace{1cm} (D.1)

for some positive integer $p \geq 1$. Probability theory corresponds to $p = 1$. The Dempster-Shafer theory of evidential reasoning and Possibility theory are defined over a finite and exhaustive set of mutually exclusive propositions $\Theta$. For a mutually exclusive frame of discernment $\Theta$ such that no two propositions may be true simultaneously then Equation D.1 generalises thus:

$$\left( \forall \Theta \subseteq \Theta \right) E \left( \bigcup_i A_i \right) = \left[ \sum_i E(A_i)^p \right]^{\frac{1}{p}}.$$

We subsequently show that there is a $p$-norm theory (which is, of course, twice differentiable) which produces belief values arbitrarily close to theories based on max operations and thus vindicate non-additive uncertainty calculi. Finally, with reference to different but logically equivalent propositional variable bases, we derive expressions for upper and lower evidence for imprecisely assigned evidence.

We begin, however, by investigating Cox’s proof for evidence unambiguously assigned to singleton propositional variables. For clarity of comparison with Cox’s original paper we adopt Cox’s nomenclature for the logical operations.

**Assumption 1.** Cox assumes that familiar first order logic identities hold (page 5):  \(^1\)

1. $\sim a = a$  \hspace{1cm} (D.2)
2. $a \cdot b = b \cdot a$  \hspace{1cm} (D.3)
3. $a \lor b = b \lor a$  \hspace{1cm} (D.4)
4. $a = a$  \hspace{1cm} (D.5)
5. $a \lor a = a$  \hspace{1cm} (D.6)
6. $a \cdot (bc) = (a \cdot b)c$  \hspace{1cm} (D.7)
7. $a \lor (b \lor c) = (a \lor b) \lor c = a \lor b \lor c$  \hspace{1cm} (D.8)
8. $\sim (ab) = \sim a \lor \sim b$  \hspace{1cm} (D.9)
9. $\sim (a \lor b) = \sim a. \sim b$  \hspace{1cm} (D.10)
10. $a \cdot (a \lor b) = a$  \hspace{1cm} (D.11)
11. $a \lor (ab) = a$  \hspace{1cm} (D.12)

**Assumption 2.** Cox also asserts a number of desirable properties for a theory of uncertainty:

- **Clarity.** Propositions must be defined precisely enough so that it would be possible to determine whether a proposition is indeed true or false.

- **Completeness.** It is possible to assign a degree of belief to any proposition which is precisely defined.

- **Scalar Continuity.** A measure of belief can vary continuously between values of absolute truth and falsehood and that the continuum of belief can be represented by a single real number.

- **Context Dependency.** Current degree of belief for a proposition depends on the current state of information of the individual assessing the belief.

\(^1\)Page numbers refer to the original publication in the *American Journal of Physics*. For reference to the Readings in *Uncertainty* add 352.
• **Hypothetical Conditioning.** The belief in the conjunction \( c \land b \) given that \( a \) is true should be related to the belief in \( b \) alone given \( a \) and to the belief in \( c \) given that both \( a \) and \( b \) are true. Formally, there exists a function \( F \) such that:

\[
\mathcal{E}(c \land b \mid a) = F(\mathcal{E}(c \mid b \land a), \mathcal{E}(b \mid a)).
\]

\( F \) is arbitrary and constrained only by the fact that it must be continuously twice differentiable.

• **Complementarity.** The belief in \( \sim b \) (not \( b \)) given \( a \) should be related to the belief in \( b \) given \( a \):

\[
\mathcal{E}(\sim b \mid a) = S(\mathcal{E}(b \mid a)).
\]

In his appendix, Cox derives the following necessary condition from these assumptions:

\[
S(a)^p + u^p = 1.
\]

That is:

\[
\mathcal{E}(\sim b \mid a)^p + \mathcal{E}(b \mid a)^p = 1. \tag{D.13}
\]

Using logical identities Cox then demonstrates that hypothetical conditioning must have the form of his Equation (C9): \(^2\)

\[
C\mathcal{E}(c \land b \mid a) = \mathcal{E}(c \mid b \land a)\mathcal{E}(b \mid a).
\]

Cox then argues that \( \mathcal{E}(\cdot)^p \) is a probability distribution.

We follow Cox's reasoning as presented in his paper. However, we will endeavour to find calculi for \( \mathcal{E}(\cdot) \) consistent with Cox's theorem. Throughout our exposition we retain explicit reference to the power relationship between evidence and the operators operating on the evidence. We begin with Equation (C10) from Cox's paper raised to the \( p^{th} \) power as described by Cox:

\[
\mathcal{E}(c \land b \mid a)^p = \mathcal{E}(c \mid b \land a)^p \mathcal{E}(b \mid a)^p. \tag{D.14}
\]

But, in place of Cox's Equation (C16) we adopt his earlier equation:

\[
\mathcal{E}(b \mid a)^p + \mathcal{E}(\sim b \mid a)^p = 1. \tag{D.15}
\]

By Equation D.14:

\[
\mathcal{E}(c \land b \mid a)^p + \mathcal{E}(\sim c \land b \mid a)^p = [\mathcal{E}(c \mid b \land a)^p + \mathcal{E}(\sim c \mid b \land a)^p] \mathcal{E}(b \mid a)^p. \tag{D.16}
\]

By Equation D.15:

\[
\mathcal{E}(c \mid b \land a)^p + \mathcal{E}(\sim c \mid b \land a)^p = 1.
\]

Therefore, substituting into Equation D.16:

\[
\mathcal{E}(c \land b \mid a)^p + \mathcal{E}(\sim c \land b \mid a)^p = \mathcal{E}(b \mid a)^p. \tag{D.17}
\]

\(^2\)An equation identifier prefixed with \( C \) refers to an equation in Cox's paper. Thus, C9 refers to Equation 9 in Cox's paper.
Using Equation D.15 and a rule of logic (Equation D.10) and then, using Equation D.17:

\[
\mathcal{E}(c \lor b \mid a)^P = 1 - \mathcal{E}(\sim (c \lor b) \mid a)^P \\
= 1 - \mathcal{E}(\sim c \land \sim b \mid a)^P \\
= 1 - \mathcal{E}(\sim b \mid a)^P + \mathcal{E}(c \land \sim b \mid a)^P.
\]  \hspace{1cm} (D.18)

Using a rule of logic (Equation D.3), Equation D.17 and then Equation D.3 again:

\[
\mathcal{E}(c \land \sim b \mid a)^P = \mathcal{E}(b \mid a)^P - \mathcal{E}(c \land b \mid a)^P \\
= \mathcal{E}(c \mid a)^P - \mathcal{E}(b \mid a)^P \\
= \mathcal{E}(c \mid a)^P - \mathcal{E}(c \land b \mid a)^P.
\]

Substituting this expression into Equation D.18:

\[
\mathcal{E}(c \lor b \mid a)^P = \mathcal{E}(c \mid a)^P + \mathcal{E}(b \mid a)^P - \mathcal{E}(c \land b \mid a)^P
\]  \hspace{1cm} (D.19)

Compare Equation D.19 with Cox’s concluding (probabilistic) identity:

\[
\mathcal{E}(c \lor b \mid a) = \mathcal{E}(c \mid a) + \mathcal{E}(b \mid a) - \mathcal{E}(c \land b \mid a)
\]

When the evidence operates over a frame of discernment such that \( c \) and \( b \) are mutually exclusive then, given a piece of evidence \( a \), the degree to which this evidence supports the event \( \sim b \) and \( c \) is zero (i.e. \( \mathcal{E}(c \land b \mid a) = 0 \)) and:

\[
\mathcal{E}(c \lor b \mid a)^P = \mathcal{E}(c \mid a)^P + \mathcal{E}(b \mid a)^P.
\]  \hspace{1cm} (D.20)

By recursion and asserting that \( a \) is true:

\[
\mathcal{E}(c_1 \lor \ldots \lor c_n)^P = \sum_{i=1}^{n} \mathcal{E}(c_i)^P.
\]  \hspace{1cm} (D.21)

When \( \{c_0, c_1, \ldots, c_n\} \) is exhaustive then \( \sim c_0 = c_1 \lor c_2 \lor \ldots \lor c_n \). Thus, by Equations D.15 and D.21 we obtain:

\[
\left[ \sum_{i} \mathcal{E}(c_i)^P \right]^{\frac{1}{P}} = \left[ \mathcal{E}(c_0)^P + \sum_{i=1}^{n} \mathcal{E}(c_i)^P \right]^{\frac{1}{P}} = \left[ \mathcal{E}(c_0)^P + \mathcal{E}(\sim c_0)^P \right]^{\frac{1}{P}} = 1.
\]  \hspace{1cm} (D.22)

We will now see that, given any acceptable numeric error \( \epsilon \) in the corresponding values of evidence assigned to propositions (whether elementary or disjunctive) there is a \( p \)-norm theory which is less than \( \epsilon \) close to the max theory. The max theory aggregates disjunctive evidence thus:

\[
(\forall A_i \subseteq \Theta) \quad \mathcal{E}(\bigcup_{i} A_i) = \max_{i} \{ \mathcal{E}(A_i) \}.
\]

We use an important inequality (derivable by induction on positive integers \( P \)):

\[
(\forall P \geq 1) \quad \left( 1 + \frac{A}{p} \right)^P \geq 1 + A.
\]

Taking the \( p^{th} \) root and constraining \( A \in [0, 1] \) so that \( A^P \leq A \) for \( p \geq 1 \) we obtain:

\[
(\forall P \geq 1, 0 \leq A \leq 1) \quad \left( 1 + \frac{A}{p} \right) \geq (1 + A^p)^{\frac{1}{p}}.
\]  \hspace{1cm} (D.23)
Let $Y_p$ be the result of aggregating disjunctive evidence $E_i$ using the $p$-norm theory:

$$Y_p = \left[ \sum_i E_i^p \right]^\frac{1}{p} = \max_i \{E_i\} \left[ \sum_i \left( \frac{E_i}{\max_i \{E_i\}} \right)^p \right]^\frac{1}{p}.$$  

The value of $Y_p$ is bounded thus:

$$(\forall p \geq 1) \quad \max_i \{E_i\} \leq Y_p \leq \max \left\{ \frac{\sum_i \{E_i\}}{\max \{E_i\}} , \frac{\sum_i \{E_i\}}{\max \{E_i\}} \right\}^\frac{1}{p} \left[ 1 + X^p \right]^\frac{1}{p}$$  

where $N_{E_i=\max \{E_i\}}$ is the number of $E_i$ for which $E_i = \max \{E_i\}$, $N_{E_i<\max \{E_i\}}$ is the number of $E_i$ which are strictly less than $\max \{E_i\}$ and $0 \leq X < 1$. Using Equation D.23, for all theories $p$:

$$\max_i \{E_i\} \leq Y_p \leq \max \left\{ \frac{\sum_i \{E_i\}}{\max \{E_i\}} , \frac{\sum_i \{E_i\}}{\max \{E_i\}} \right\}^\frac{1}{p} \left[ 1 + X^p \right]^\frac{1}{p} \leq \max \left\{ \frac{\sum_i \{E_i\}}{\max \{E_i\}} , \frac{\sum_i \{E_i\}}{\max \{E_i\}} \right\}^\frac{1}{p} \left[ 1 + \frac{X}{p} \right] \leq \max \left\{ \frac{\sum_i \{E_i\}}{\max \{E_i\}} , \frac{\sum_i \{E_i\}}{\max \{E_i\}} \right\}^\frac{1}{p} \left[ 1 + \frac{X}{p} \right] \leq \max \left\{ \frac{\sum_i \{E_i\}}{\max \{E_i\}} , \frac{\sum_i \{E_i\}}{\max \{E_i\}} \right\}^\frac{1}{p} \left[ 1 + \frac{X}{p} \right] \leq \max_i \{E_i\} + \frac{N}{p} \max_i \{E_i\}.$$  

where $N$ is the number of elementary propositions and is independent of $p$. Thus, there exists a $p$-norm theory $p$ whose summation operation over disjunctive aggregation is an arbitrarily close approximation to the theory based on disjunctive aggregation using the max operation. We derive the disjunctive aggregation operation for the $\infty$-norm theory. By Equation D.17:

$$E(c \cdot b \mid a)^p \leq E(b \mid a)^p.$$  

Also, by switching the variables and using Equation D.3:

$$E(c \cdot b \mid a)^p = E(b \cdot c \mid a)^p \leq E(c \mid a)^p.$$  

Thus:

$$E(c \cdot b \mid a)^p \leq \min \{E(c \mid a)^p, E(b \mid a)^p\}.$$  

Hence, we obtain our first $\infty$-norm theorem using Equation D.19. In the limit $p \to \infty$:

$$E(c \lor b \mid a) = \max \{E(c \mid a), E(b \mid a)\}.$$  

In summary, between Laplace’s death and the 1950s the statistics community held that (Jaynes, 1979) it is not legitimate to use probability in the sense of degree of plausibility rather than frequency. However, Cox’s argument demonstrated the opposite, that Laplace’s result (i.e. Bayes theorem) was necessarily true for all kinds of reasoning (subject to a monotonic transformation of evidence values). Thus, as it stood, Cox’s theorem seemed to prove that evidential theories with operations based on the max operations were inconsistent. However, Kosko stated that this was not valid as the conditions necessary for Cox’s theorem to hold do not apply to non-differentiable max operations.
Kosko thus asserted that Cox’s theorem was indifferent as far as fuzzy logic and Possibility theory were concerned. To turn the tide further, by showing that Cox’s proof is less constraining than originally thought and by asserting that max operations are approximated arbitrarily closely by a $p$-norm function consistent with Cox’s theorem (and proof), we have demonstrated that Cox’s theorem is actually consistent with theories of reasoning which apply max operations.

### D.3 Upper and Lower Evidence

So far we have considered precise assignments of evidence. To extend this result to imprecise statements of evidence consider the imprecise logical relationship between two representations, $\Theta_1 = \{a, b, c\}$ and $\Theta_2 = \{e, d\}$, given by $a \supset d$, $b \supset e$ and $c \supset d \vee e$. Now, “$d \vee e$” is true if $a$, $b$ or $c$ are true and, by Equation D.20:

$$E(d \vee e) = [E(a)^p + E(b)^p + E(c)^p]^\frac{1}{p}.$$  \hspace{1cm} (D.24)

That is, evidence assigned to $d \vee e$ is a combination of evidence directly assigned to $d$ and $e$ and evidence directly assigned to the proposition “$d \vee e$”. Constructing a frame $\{d, e, "d \vee e"\}$ and writing the evidence assigned directly to a member of the frame $m = E$ then Equation D.24 becomes:

$$E(d \vee e) = [m(d)^p + m(e)^p + m("d \vee e")^p]^\frac{1}{p}.$$ 

The function $m$ is called the basic mass assignment (or basic probability assignment BPA) and is an evidence function $E$ operating over a frame $2^\Theta = \{S \mid S \subseteq \Theta\}$. By Equation D.21 we have the fundamental axiom of the $p$-norm theories:

$$\left[ \sum_{A \in \Theta} m(A)^p \right]^\frac{1}{p} = E(\Theta) = 1.$$
Appendix E

The Generalised Dempster-Shafer Theory

Here we present proofs for theorems for p-norm theory of evidence. We refer to the proofs for the 1-norm method in (Shafer, 1976).

Lemma 1 (Equivalent to (Shafer, 1976), Lemma 2.1). If $A$ is a finite set, then:

$$\sum_{B \subseteq A} (-1)^{|B|} = \begin{cases} 1 & \text{if } A = \emptyset \\ 0 & \text{otherwise.} \end{cases} \quad (E.1)$$

Proof See (Shafer, 1976), p47.

Lemma 2 (Equivalent to (Shafer, 1976), Lemma 2.2). If $A$ is a finite set and $B \subseteq A$, then:

$$\sum_{C \subseteq B \subseteq A} (-1)^{|C|} = \begin{cases} (-1)^{|A|} & \text{if } A = B \\ 0 & \text{otherwise.} \end{cases} \quad (E.2)$$

Proof See (Shafer, 1976), p48.
Theorem 19 (Extension of (Shafer, 1976), Theorem 2.1) If \( \Theta \) is a frame of discernment and \( p \geq 1 \) then a function \( \text{Bel} : \mathcal{P} \rightarrow [0,1] \) given by:

\[
\begin{align*}
\text{Bel}(\emptyset) &= 0 \\
\text{Bel}(\Theta) &= 1 \\
\text{Bel}(A) &= \left( \sum_{B \subseteq A} m(B)^p \right)^{\frac{1}{p}}
\end{align*}
\]

satisfies the following condition for every positive integer \( m \) and every collection \( A_1, \ldots, A_m \) of subsets of \( \Theta \):

\[
\text{Bel}(A_1 \cup \ldots \cup A_m) \geq \left[ \sum_{I \subseteq \{1, \ldots, m\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel}(\bigcap_{i \in I} A_i)^p \right]^{\frac{1}{p}}.
\]

Proof. Fix a collection \( A_1, \ldots, A_m \) of subsets of \( \Theta \), and set:

\[ I(B) = \{ i | 1 \leq i \leq m; B \subseteq A_i \} \quad \text{(E.7)} \]

for each \( B \subseteq \Theta \). Using Lemma 1:

\[
\begin{align*}
\left[ \sum_{I \subseteq \{1, \ldots, m\}, I \neq \emptyset} (-1)^{|I|+1} \text{Bel}(\bigcap_{i \in I} A_i)^p \right]^{\frac{1}{p}} &= \left[ \sum_{I \subseteq \{1, \ldots, m\}, I \neq \emptyset} (-1)^{|I|+1} \sum_{B \subseteq \bigcap_{i \in I} A_i} m(B)^p \right]^{\frac{1}{p}} \\
&= \left[ \sum_{B \subseteq \Theta, I(B) \neq \emptyset} m(B)^p \sum_{I \subseteq I(B), I \neq \emptyset} (-1)^{|I|+1} \right]^{\frac{1}{p}} \\
&= \left[ \sum_{B \subseteq \Theta, I(B) \neq \emptyset} m(B)^p (1 - \sum_{I \subseteq I(B)} (-1)^{|I|}) \right]^{\frac{1}{p}} \\
&= \left[ \sum_{B \subseteq \Theta, I(B) \neq \emptyset} m(B)^p \right]^{\frac{1}{p}} \\
&= \left[ \sum_{B \subseteq A_i \cup \ldots \cup A_m, \text{for some } i} m(B)^p \right]^{\frac{1}{p}} \quad \text{(E.8)}
\end{align*}
\]

and this is certainly less than or equal to:

\[
\text{Bel}(A_1 \cup \ldots \cup A_m) = \left[ \sum_{B \subseteq A_1 \cup \ldots \cup A_m} m(B)^p \right]^{\frac{1}{p}}. \quad \text{(E.9)}
\]

\( \square \)
Theorem 20 (Extension of (Shafer, 1976), Theorem 2.2) Suppose Bel : \( \mathcal{P} \rightarrow [0,1] \) is the belief function given by the basic probability assignment \( m : \mathcal{P} \rightarrow [0,1] \) and \( p \geq 1 \). Then:

\[
Bel(A) = \left[ \sum_{B \subseteq A} m(B)^p \right]^{\frac{1}{p}} \quad (E.10)
\]

for all \( A \subseteq \Theta \) if and only if:

\[
m(A) = \left[ \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B)^p \right]^{\frac{1}{p}} \quad (E.11)
\]

for all \( A \subseteq \Theta \).

Proof. Both implications follow by simple calculations using Lemma 2.

1. If Equation E.10 holds for all \( A \subseteq \Theta \) then:

\[
\left[ \sum_{B \subseteq A} (-1)^{|A-B|} Bel(B)^p \right]^{\frac{1}{p}} = \left[ (-1)^{|A|} \sum_{B \subseteq A} (-1)^{|B|} Bel(B)^p \right]^{\frac{1}{p}}
\]

\[
= \left[ (-1)^{|A|} \sum_{B \subseteq A} (-1)^{|B|} \sum_{C \subseteq B} m(C)^p \right]^{\frac{1}{p}}
\]

\[
= \left[ (-1)^{|A|} \sum_{C \subseteq B} m(C)^p \sum_{B \subseteq C \subseteq A} (-1)^{|B|} \right]^{\frac{1}{p}}
\]

\[
= \left[ (-1)^{|A|} m(B)^p (-1)^{|A|} \right]^{\frac{1}{p}}
\]

\[= m(A). \quad (E.12)\]

2. If Equation E.11 holds for all \( A \subseteq \Theta \) then:

\[
\left[ \sum_{B \subseteq A} m(B)^p \right]^{\frac{1}{p}} = \left[ \sum_{B \subseteq A} \sum_{C \subseteq B} (-1)^{|B-C|} Bel(C)^p \right]^{\frac{1}{p}}
\]

\[
= \left[ \sum_{C \subseteq A} (-1)^{|C|} Bel(C)^p \sum_{B \subseteq C \subseteq A} (-1)^{|B|} \right]^{\frac{1}{p}}
\]

\[
= \left[ (-1)^{|A|} Bel(A)^p (-1)^{|A|} \right]^{\frac{1}{p}}
\]

\[= Bel(A). \quad (E.13)\]

\[\square\]

Theorem 21 (Extension of (Shafer, 1976), Theorem 3.3) Let Bel_{p1} and Bel_{p2} be two belief functions over \( \Theta \), and suppose Bel_{p1} \oplus Bel_{p2} exists and \( p \geq 1 \). Denote by \( Q_{p1} \), \( Q_{p2} \) and \( Q_p \) the
commonality functions for \( \text{Bel}_{p,1}, \text{Bel}_{p,2} \) and \( \text{Bel}_{p,1} \oplus \text{Bel}_{p,2} \) respectively. Then:

\[
Q(A)_p = KQ_{p,1}(A)Q_{p,2}(A)
\]

where \( K \) is the mass normalisation value:

\[
K = \left[ \sum_{A \subseteq \Theta} \sum_{i,j: A \cap B_j = B} m_{p,1}(A_i)^p m_{p,2}(B_j)^p \right]^{\frac{1}{p}}.
\]

Proof

\[
Q_p(A) = \left[ \sum_{A \subseteq B} m(B)^p \right]^{\frac{1}{p}}
= K \left[ \sum_{A \subseteq B} \sum_{i,j: A \cap B_j = B} m_{p,1}(A_i)^p m_{p,2}(B_j)^p \right]^{\frac{1}{p}}
= K \left[ \sum_{i,j: A \subseteq A_i \cap B_j = B} m_{p,1}(A_i)^p m_{p,2}(B_j)^p \right]^{\frac{1}{p}}
= K \left[ \left( \sum_{i: A \subseteq A_i} m_{p,1}(A_i)^p \right) \left( \sum_{j: A \subseteq B_j} m_{p,2}(B_j)^p \right) \right]^{\frac{1}{p}}
= K \left[ \left( \sum_{B \subseteq \Theta: A \subseteq B} m_{p,1}(B)^p \right) \left( \sum_{B \subseteq \Theta: A \subseteq B} m_{p,2}(B)^p \right) \right]^{\frac{1}{p}}
= KQ_{p,1}(A)Q_{p,2}(A).
\]

\[\square\]

**Definition 18** The outer reduction \( \bar{\theta} \) for the refinement \( \omega: 2^\Theta \to 2^\Omega \) is the mapping \( \bar{\theta}: 2^\Theta \to 2^\Theta \):

\[
\bar{\theta}(A) = \{ \theta \in \Theta | \omega(\theta) \cap A \neq \emptyset \}.
\]

**Theorem 22** (Equivalent to (Shafer, 1976), Theorem 6.2). Suppose \( \omega: 2^\Theta \to 2^\Omega \) is a refining, \( A \subseteq \Omega \), and \( B \subseteq \Theta \). Let \( \bar{\theta} \) be the inner reduction for \( \omega \). Then \( A \subseteq \omega(B) \) if and only if \( \bar{\theta}(A) \subseteq B \).

**Proof** See (Shafer, 1976), p130.
Appendix F

The Non-existence of a Generic Eulerian Approach To Qualitative Filtering

We prove that no non-parametric BPA $m(Q|z)$ consistent for all network structures exists:

**Theorem 23 (The Unavailability of the State Centred Approach)**

The following assumptions are mutually incompatible:

1. $m(Q|z)$ is a non-parametric mass assignment to proposition $Q$ by observation $z$.
2. Decide $x \in Q$ when $P(Q|z'_1) > P(Q'|z'_1)$ for all $Q' \neq Q$.
3. At least three states receive evidence: $(\exists Q_0, Q_b, Q_c) P((Q_0) P((Q_b) P((Q_c)) > 0$ (otherwise we would have the trivial two state system which is a Lagrangian system).
4. The decision is unbiased. That is, the true state $Q$ is almost surely chosen.
5. The $p$-norm Dempster-Shafer theory holds.
6. Multivalued network transitions are possible.

**Proof** If Assumption 5 holds then by Theorem 21 in Appendix E and the fact that the plausibility values and commonality numbers coincide for singleton propositions, then for singleton elements $Q_s$ and $Q_{s'}$ and (possibly dependent) observations $z_i$:

$$
\frac{P(Q_s|z_i)}{P(Q_{s'}|z_i)} = \prod_{i=1}^{t} \frac{P(Q_s|z_i)}{P(Q_{s'}|z_i)}.
$$

If assumption 4 holds then:

$$
x^* \in Q_s \Rightarrow E \left( \log \frac{P(Q_s|z_i)}{P(Q_{s'}|z_i)} \right) > 0 
\tag{F.1}
$$

and when $x$ coincides with the landmark separating $Q_s$ and $Q_{s'}$:

$$
E \left( \log \frac{P(Q_s|z_i)}{P(Q_{s'}|z_i)} \right) = 0. 
\tag{F.2}
$$

Assumptions 2, 4 and 5 entail Equation F.1. If Assumption 1 is true as well then Equation F.1 must hold for any pdf. We choose the following pdf:

$$
f(z|x) = \delta(z - x_T) 
\tag{F.3}$$
for which, when $x_T \in Q_s$:

$$\log Pl(Q_s | z) = E(\log Pl(Q_s | z)) > E(\log Pl(Q_s' | z)) = \log Pl(Q_s' | z).$$

Thus, if our non-parametric mass function $m$ is to accommodate Equation F.3, then for any $x_T$ and $Q_s$ with $x_T \in Q_s$, and $Q_s') \neq Q_s$ there exists a proposition $W$ with $Q_s \in W$ and $Q_s' \not\in W$ and $(\forall W', Q_s' \in W')(\exists z) m(W | z) > m(W' | z)$. This is called the rogue state lemma.

We derive necessary conditions for unbiased filtering in two (static) sub-network structures. By joining these into a single network, we show that Assumptions 1 to 6 are mutually inconsistent.

**Structure 1**

A three state frame $\{a, b, c\}$ and the true state coincides with a landmark separating adjacent regions $Q_b$ and $Q_c$. By Equation F.2 we have:

$$E(\log Pl(\{Q_b\} | z)) = E(\log Pl(\{Q_c\} | z))$$

and, by the rogue state lemma:

$$(\exists z, W)(\forall W')(Q_a \in W) \wedge (Q_b \not\in W) \wedge (Q_b \in W') \wedge [m(W | z) > m(W' | z)].$$

**Structure 2**

A two state frame $\{Q_x, Q_y\}$ and the true state coincides with a landmark separating regions $Q_x$ and $Q_y$. By Equation F.2 we have:

$$E(\log Pl(\{Q_x\} | z)) = E(\log Pl(\{Q_y\} | z)).$$

**Structure 3: The Combined Network**

We invoke Assumption 6 and construct a network comprising the following transitions:

$$Q_a \rightarrow Q_z, \ Q_b \rightarrow Q_x, \ Q_c \rightarrow Q_y$$

and the true state which coincides with the landmark separating $Q_b$ and $Q_c$ also coincides with the landmark separating $Q_x$ and $Q_y$. Thus, by Equation F.5, for all $p \geq 1$:

$$E(\log Pl(\{Q_z\} | z)) = E\left(\log \left[\sum_{W:Q_b \in W} m(W | z)^p + \sum_{W:Q_c \in W} m(W | z)^p \right]^{\frac{1}{p}}\right)$$

$$> E\left(\log \left[\sum_{W:Q_b \in W} m(W | z)^p \right]^{\frac{1}{p}}\right)$$

$$= E(\log Pl(\{Q_b\} | z)).$$

Also:

$$E(\log Pl(\{Q_y\} | z)) = E(\log Pl(\{Q_c\} | z)).$$
Combining Equations F.4, F.7 and F.8 we obtain:

\[ E(\log P_l(\{Q_x\}|z)) > E(\log P_l(\{Q_y\}|z)) = E(\log P_l(\{Q_x\}|z)) = E(\log P_l(\{Q_y\}|z)) \]

which clearly conflicts with Equation F.6.

\(\Box\)