FUZZY LOGIC WITH ENGINEERING APPLICATIONS

Fourth Edition

Timothy J. Ross
University of New Mexico, USA

WILEY

This edition first published 2017

First edition published 1995
Second edition published 2004
Third edition published in 2010

Library of Congress Cataloging-in-Publication Data

Names: Ross, Timothy J., author.
Title: Fuzzy logic with engineering applications / Timothy J. Ross.
Description: Fourth edition. | Southern Gate, Chichester, West Sussex, United Kingdom : John Wiley & Sons Inc., 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016021917 | ISBN 9781119235866 (pbk.) | ISBN 9781119235859 (Adobe PDF) | ISBN 9781119235842 (epub)
Subjects: LCSH: Engineering–Statistical methods. | Fuzzy logic.
Classification: LCC TA340 .R673 2017 | DDC 620.001/511313–dc23 LC record available at https://lccn.loc.gov/2016021917

A catalogue record for this book is available from the British Library.

Set in 10/12pt Times by SPi Global, Pondicherry, India

1 2017

Contents

Preface to the Fourth Edition

My primary motivations for writing the fourth edition of this text have been to (1) reduce the length of the text, (2) correct the errata discovered since the publication of the third edition, and (3) introduce limited new material for the readers. The first motivation has been accomplished by eliminating some sections that are rarely taught in the classroom by various faculty using this text and by eliminating some sections that do not add to the utility of the text as a tool to learn basic fundamentals of the subject.

Since the first edition was published, in 1995, the technology of fuzzy set theory and its application to systems, using fuzzy logic, has moved rapidly. Developments in other theories such as possibility theory and evidence theory (both being elements of a larger collection of methods under the rubric "generalized information theories") have shed more light on the real virtues of fuzzy logic applications, and some developments in machine computation have made certain features of fuzzy logic much more useful than in the past. In fact, it would be fair to state that some developments in fuzzy systems are quite competitive with other, linear algebra-based methods in terms of computational speed and associated accuracy.

There is some new material which is included in the fourth edition to try to capture some of the newer developments; the keyword here is *some* because it would be impossible to summarize or illustrate even a small fraction of the new developments of the last six years since the third edition was published. As with any book containing technical material, the third edition contained errata that have been corrected in this fourth edition. As with the first three editions, a solutions manual for all problems in the fourth edition and software can be obtained by qualified instructors who visit www.wiley.com/go/ross/fuzzy4e and provide official documentation of their teaching status. In addition to the solutions manual, a directory of software will be made available to all student users and faculty of the text on this same website. Most of the software routines make use of the MATLAB platform, and most of the routines have been written by my students, except for the standard routines that exist as MATLAB functions.

As I discussed in the preface of the third edition, the axioms of a probability theory referred to as the *excluded middle* are again referred to in this edition as axioms—never as laws.

The operations due to De Morgan are also not referred to as laws, but as *principles* because these principles do apply to some (*not all*) uncertainty theories (e.g., probability and fuzzy). The *excluded middle axiom* (and its dual, the *axiom of contradiction*) are not *laws*; Newton produced *laws*, Kepler produced *laws*, Darcy, Boyle, Ohm, Kirchhoff, Bernoulli, and many others too numerous to list here all developed *laws*. *Laws* are mathematical expressions describing the immutable realizations of nature. Definitions, theorems, and axioms collectively can describe a certain axiomatic foundation describing a particular kind of theory, and nothing more; in this case, the *excluded middle* and other axioms can be used to describe a probability theory. Hence, if a fuzzy set theory does not happen to be *constrained* by an *excluded middle axiom*, it is not a *violation* of some immutable law of nature like Newton's laws; fuzzy set theory simply does not happen to have an axiom of the excluded middle; it does not need, nor is *constrained by*, such an axiom. In fact, as early as 1905 the famous mathematician L. E. J. Brouwer defined this excluded middle axiom as a *principle* in his writings; he showed that the *principle of the excluded middle* was inappropriate in some logics, including his own, which he termed *intuitionism*. Brouwer observed that Aristotelian logic is only a part of mathematics, the special kind of mathematical thought obtained if one restricts oneself to relations of the whole and part. Brouwer had to specify in which sense the principles of logic could be considered "laws" because within his intuitionistic framework thought did not follow any rules, and, hence, "law" could no longer mean "rule" (see the detailed discussions on this in Chapters 5 and 13). In this regard, I continue to take on the cause advocated by Brouwer more than a century ago.

Also in this fourth edition, as in previous editions, I do not refer to "fuzzy measure theory" but instead describe it as "monotone measure theory"; the reader will see this in the title of Chapter 13. The former phrase still causes confusion when referring to fuzzy set theory; we hope to help in ending this confusion. In Chapter 13, in describing the monotone measure, m, I use the phrase describing this measure as a "basic evidence assignment (bea)," as opposed to the early use of the phrase "basic probability assignment (bpa)." Again, we attempt to avoid confusion with any of the terms typically used in probability theory.

As with the first three editions, this fourth edition is designed for the professional and academic audience interested primarily in applications of fuzzy logic in engineering and technology. I have found that the majority of students and practicing professionals are interested in the applications of fuzzy logic to their particular fields. Hence, the text is written for an audience primarily at the senior undergraduate and first-year graduate levels. With numerous examples throughout, this text is written to assist the learning process of a broad cross section of technical disciplines. It is primarily focused on applications, but each of the chapters begin with the rudimentary structure of the underlying mathematics required for a fundamental understanding of the methods illustrated.

Chapter 1 introduces the basic concept of fuzziness and distinguishes fuzzy uncertainty from other forms of uncertainty. It also introduces the fundamental idea of set membership, thereby laying the foundation for all material that follows, and presents membership functions as the format used for expressing set membership. The chapter summarizes a historical review of uncertainty theories and reviews the idea of "sets as points" in an n-dimensional Euclidean space as a graphical analog in understanding the relationship between classical (crisp) and fuzzy sets. A new section in the chapter addresses the intuition of propagating uncertainty by showing an example that compares the results of propagating probabilities on the one hand, or membership values on the other, through a simple nonlinear function. In this example there are some counterintuitive findings that readers will find both interesting and instructive.

Chapter 2 reviews classical set theory and develops the basic ideas of fuzzy sets. Operations, axioms, and properties of fuzzy sets are introduced by way of comparisons with the same entities for classical sets. Various normative measures to model fuzzy intersections (t-norms) and fuzzy unions (t-conorms) are summarized.

Chapter 3 develops the ideas of fuzzy relations as a means of mapping fuzziness from one universe to another. Various forms of the composition operation for relations are presented. Again, the epistemological approach in this chapter uses comparisons with classical relations in developing and illustrating fuzzy relations. Chapter 3 also illustrates methods to determine the numerical values contained within a specific class of fuzzy relations, called *similarity relations*. The section on a three-dimensional physical analogy of equivalence relations has been deleted.

Chapter 4 discusses the fuzzification of scalar variables and the defuzzification of membership functions. It introduces the basic features of a membership function and it discusses, briefly, the notion of interval-valued fuzzy sets. Defuzzification is necessary in dealing with the ubiquitous crisp (binary) world around us. The chapter details defuzzification of fuzzy sets and fuzzy relations into crisp sets and crisp relations, respectively, using lambda-cuts, and it describes a variety of methods to defuzzify membership functions into scalar values. Some of the defuzzification methods in the third edition have been deleted because they are seldom used in practice and because they are covered elsewhere in the literature. Examples of all methods are given in the chapter.

Chapter 5 introduces the precepts of fuzzy logic, again through a review of the relevant features of classical, or a propositional, logic. Various logical connectives and operations are illustrated. There is a thorough discussion of the various forms of the implication operation and the composition operation provided in this chapter. Three different inference methods, popular in the literature, are illustrated. Approximate reasoning, or reasoning under imprecise (fuzzy) information, is also introduced in this chapter. Basic IF–THEN rule structures are introduced and three graphical methods of inference are presented. The section on Natural Language has been shortened. A few more examples of the difficulties of using the axiom of the excluded middle are given in the summary of the chapter.

Chapter 6 provides several classical methods of developing membership functions, including methods that make use of the technologies of neural networks, genetic algorithms, and inductive reasoning.

Chapter 7 presents six automated methods that can be used to generate rules and membership functions from observed or measured input–output data. The procedures are essentially computational methods of learning. Examples are provided to illustrate each method. Many of the problems at the end of the chapter will require software; this software can be downloaded from www.wiley.com/go/ross/fuzzy4e.

Beginning the second category of chapters in the book highlighting applications, Chapter 8 continues with the rule-based format to introduce fuzzy nonlinear simulation and complex system modeling. In this context, nonlinear functions are seen as mappings of information "patches" from the input space to information "patches" of the output space, instead of the "point-to-point" idea taught in classical engineering courses. Fidelity of the simulation is illustrated with standard functions, but the power of the idea can be seen in systems too complex for an algorithmic description. This chapter formalizes fuzzy associative memories (FAMs) as generalized mappings.

Chapter 9 develops fuzzy decision making by introducing some simple concepts in ordering, preference and consensus, and multiobjective decisions. It introduces the powerful concept of Bayesian decision methods by fuzzifying this classic probabilistic approach. This chapter illustrates the power of combining fuzzy set theory with probability to handle random and nonrandom uncertainty in the decision-making process.

Chapter 10 discusses a few fuzzy classification methods by contrasting them with classical methods of classification and develops a simple metric to assess the goodness of the classification, or misclassification. This chapter also summarizes classification using equivalence relations. It now has a section on pattern recognition, gleaned from the third edition. This section introduces a useful metric in pattern recognition using the algebra of fuzzy vectors. A single-feature and a multiple-feature procedure are summarized in the chapter. The section on image processing has been deleted because other books have extensive coverage of this area.

The chapter in the third edition on fuzzy arithmetic and fuzzy numbers has been deleted. A summary of Zadeh's extension principle and a few simple examples of fuzzy arithmetic are included in Chapter 12.

Chapter 11 introduces the field of fuzzy control systems. A brief review of control system design and control surfaces is provided. Simple example problems in control are provided. Two sections in this chapter are worth noting: fuzzy engineering process control and fuzzy statistical process control, with examples on both provided. A discussion of the comparison of fuzzy and classical control is contained in the chapter summary, and a few more examples of fuzzy control in industrial systems and applications are also included.

Chapter 12 has been extensively changed by including more information on genetically evolved fuzzy cognitive maps, and new sections on the extension principle and fuzzy arithmetic and on fuzzy data fusion are also detailed. Previous sections on fuzzy optimization and fuzzy agent-based methods are still contained in this chapter.

Finally, Chapter 13 enlarges the reader's understanding of the relationship between fuzzy uncertainty and random uncertainty (and other general forms of uncertainty, for that matter) by illustrating the foundations of monotone measures. The chapter discusses monotone measures in the context of evidence theory, possibility theory, and probability theory. The chapter has a section on methods to develop approximate possibility distribution functions derived from both data intervals and scalar point data.

Most of the text can be covered in a one-semester course at the senior undergraduate level. In fact, most science disciplines and virtually all math and engineering disciplines contain the basic ideas of set theory, mathematics, and deductive logic, which form the only knowledge necessary for a complete understanding of the text. For an introductory class, instructors may want to exclude some or all of the material covered in the last section of Chapter 6 (neural networks, genetic algorithms, and inductive reasoning), Chapter 7 (automated methods of generation), and any of the final three chapters: Chapter 11 (fuzzy control), Chapter 12 (miscellaneous fuzzy applications), and Chapter 13 on alternative measures of uncertainty. I consider the applications in Chapter 8 on simulations, Chapter 9 on decision making, and Chapter 10 on classification to be important in the first course on this subject. The other topics could be used either as introductory material for a graduate-level course or for additional coverage for graduate students taking the undergraduate course for graduate credit.

In terms of organization, the first eight chapters of the text develop the foundational material necessary to get students in a position where they can generate their own fuzzy systems. The last five chapters use the foundation material from the first eight chapters to present specific applications.

Most of the problems at the end of each chapter have been revised with different numbers, and there are many new problems that have been added to the text and some problems from the third edition deleted. In reducing the length of the book, some old problems have been deleted from many chapters. The problems in this text are typically based on current and potential applications, case studies, and education in intelligent and fuzzy systems in engineering and related technical fields. The problems address the disciplines of computer science, electrical engineering, manufacturing engineering, industrial engineering, chemical engineering, petroleum engineering, mechanical engineering, civil engineering, environmental engineering, and engineering management, and a few related fields such as mathematics, medicine, operations research, technology management, the hard and soft sciences, and some technical business issues. The references cited in the chapters are listed toward the end of each chapter. These references provide sufficient detail for those readers interested in learning more about particular applications using fuzzy sets or fuzzy logic. The large number of problems provided in the text at the end of each chapter allows instructors a sizable problem base to afford instruction using this text on a multisemester or multiyear basis, without having to assign the same problems term after term.

Again I wish to give credit to some of the individuals who have shaped my thinking about this subject since the first edition of 1995, and to others who by their simple association with me have caused me to be more circumspect about the use of the material contained in the text. Three colleagues at Los Alamos National Laboratory have continued to work with me on applications of fuzzy set theory, fuzzy logic, and generalized uncertainty theory: Dr. Greg Chavez (who wrote much of Chapter 7) and Drs. Jamie Langenbrunner and Jane Booker (retired), who both have worked extensively in an area known as *quantification of margins and uncertainty (QMU)* in assessing reliability of man-made systems; in this regard these three individuals have all explored the use of fuzzy logic and possibility theory in their work. I wish to acknowledge the organizational support of two individuals in the Brazilian Institute, Instituto de Pesquisas Engergéticas e Nucleares (IPEN), in São Paulo, Brazil. These two researchers, Drs. Francisco Lemos and Antônio Barroso, through their invitations and travel support, have enabled me to train numerous Brazilian scientists and engineers in fuzzy logic applications in their own fields of work, most notably nuclear waste management, knowledge management, and risk assessment. My discussions with them have given me ideas about where fuzzy logic can impact new fields of inquiry.

I wish to thank two of my recent graduate students who have undertaken MS theses or PhD dissertations related to fuzzy logic and whose diligent work has assisted me in writing this new edition: Clay Phillips, Sandia National Laboratory, and Donald Lincoln, NStone Corporation. These former students have helped me with additional material that I have added in Chapter 12 and have helped discover some errata in this text. There have been numerous students over the past five years who have found much of the errata I have corrected; unfortunately, too numerous to mention in this brief preface. I want to thank them all for their contributions.

Four individuals need specific mention because they have contributed some sections to this text. I would like to thank specifically Dr. Jerry Parkinson for his contributions to Chapter 11, in the areas of chemical process control and fuzzy statistical process control; Dr. Greg Chavez for his contributions in automated methods; Dr. Sunil Donald for his early work in possibility distributions in Chapter 13; and Dr. Jung Kim for his contribution in Chapter 13 of a new procedure to combine disparate interval data. I would like to acknowledge the work of two graduate students at the University of New Mexico: Rashid Ahmad who has helped in developing some equations and figures for one chapter in the text, and to Pradeep Paudel for the development of the solutions manual for the text.

One individual deserves my special thanks and praise, and that is Professor Mahmoud Taha, my colleague in Civil Engineering at the University of New Mexico. In the last five years he has continued with his work in fuzzy logic applications, pattern recognition, and applications using possibility theory; I am proud and grateful to have been his mentor. I am indebted to his collaborations, his quick adaptation in the application of these tools, and in being a proficient research colleague of mine.

I am grateful for support in the past from IPEN in Brazil to teach this subject in their facility in late 2012 and to work in the area of fuzzy cognitive maps and to the Fulbright Foundation of Brazil to support me for two summers in 2013–2014 to continue my work at the Pontifícia Universidade Católica (PUC) do Rio de Janeiro, Brazil. Although most of my research at PUC was in bamboo engineering, I taught a graduate course there in fuzzy logic. Three individuals at PUC deserve special note: Professors Khosrow Ghavami and Raul Rosas da Silva, who both made my visit to Rio de Janeiro a most personally rewarding visit, and graduate student, Marco Antônio da Cunha, whose energy allowed me to advance my understanding of the use of cognitive maps in modeling the creation of a new field in bamboo engineering and in the use of agent-based models in a new field of enzymatic catalysis. Most recently (2015) on my sabbatical to the Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, New Mexico, I want to thank Messrs. Paul Zetocha and Apoorva Bhopale for their interest, and collaboration in, a new thrust area of research in fuzzy data fusion in the area of satellite orbital control.

With so many texts covering specific niches of fuzzy logic it is not possible to summarize all the important facets of fuzzy set theory and fuzzy logic in a single text. The hundreds of edited works and tens of thousands of archival papers show clearly that this is a rapidly growing technology, where new discoveries are being published every month. It remains my fervent hope that this introductory text will assist students and practicing professionals to learn, to apply, and to be comfortable with fuzzy set theory and fuzzy logic. I welcome comments from all readers to improve this text as a useful guide for the community of engineers and technologists who will become knowledgeable about the potential of fuzzy system tools for their use in solving the complex problems that challenge us each day.

Timothy J. Ross

University of New Mexico
Albuquerque, New Mexico

1

Introduction

It is the mark of an instructed mind to rest satisfied with that degree of precision which the nature of the subject admits, and not to seek exactness where only an approximation of the truth is possible.

Aristotle, 384–322 B.C.
Ancient Greek philosopher

Precision is not truth.

Henri E. B. Matisse, 1869–1954
Impressionist painter

All traditional logic habitually assumes that precise symbols are being employed. It is therefore not applicable to this terrestrial life but only to an imagined celestial existence.

Bertrand Russell, 1923
British philosopher and Nobel Laureate

We must exploit our tolerance for imprecision.

Lotfi Zadeh, 1973
Professor, Systems Engineering, UC–Berkeley

The preceding quotes, all of them legendary, have a common thread. That thread represents the relationship between precision and uncertainty. The more uncertainty in a problem, the less precise we can be in our understanding of that problem. It is ironic that the oldest quote is attributed to the philosopher who is credited with the establishment of Western logic—a binary logic that admits only the opposites of true and false, a logic that does not admit degrees of truth in between these two extremes. In other words, Aristotelian logic does not admit imprecision in truth. However, Aristotle's quote is so appropriate today; it is a quote that admits uncertainty.

It is an admonishment that we should heed; we should balance the precision we seek with the uncertainty that exists. Most engineering texts do not address the uncertainty in the information, models, and solutions that are conveyed within the problems addressed therein. This text is dedicated to the characterization and quantification of uncertainty within engineering problems such that an appropriate level of precision can be expressed. When we ask ourselves why we should engage in this pursuit, one reason should be obvious: achieving high levels of precision costs time or money or both. Are we solving problems that require precision? The more complex a system is, the more imprecise or inexact is the information that we have to characterize that system. It seems, then, that precision and information and complexity are inextricably related in the problems we pose for eventual solution. However, for most of the problems that we face, the quote credited to Professor Zadeh suggests that we can do a better job in accepting some level of imprecision.

It seems intuitive that we should balance the degree of precision in a problem with the associated uncertainty in that problem. Hence, this text recognizes that uncertainty of various forms permeates all scientific endeavors, and it exists as an integral feature of all abstractions, models, and solutions. Hence, the intent of this book is to introduce methods to handle one of these forms of uncertainty in our technical problems, the form we have come to call *fuzziness*.

The Case for Imprecision

Our understanding of most physical processes is based largely on imprecise human reasoning. This imprecision (when compared to the precise quantities required by computers) is nonetheless a form of information that can be quite useful to humans. The ability to embed such reasoning in hitherto intractable and complex problems is the criterion by which the efficacy of fuzzy logic is judged. Undoubtedly, this ability cannot solve problems that require precision, problems such as shooting precision laser beams more than tens of kilometers in space; milling machine components to accuracies of parts per billion; or focusing a microscopic electron beam on a specimen the size of a nanometer. The impact of fuzzy logic in these areas might be years away, if ever. But not many human problems require such precision, problems such as parking a car, backing up a trailer, navigating a car among others on a freeway, washing clothes, controlling traffic at intersections, judging beauty contestants, and a preliminary understanding of a complex system.

There are many simple examples in our culture that illustrate the lack of necessity for precision in much of what we do. There is a joke that is a good illustration about the lack of information contained in a precise number (Paulos, 1995). "A natural history museum guard told a visitor that the dinosaur on exhibit was 90,000,006 years old. Upon questioning about the specific number he used, the guard explained that he was told the dinosaur was 90,000,000 years old when he was hired, six years before! One can easily see the folly in adding a precise number to an imprecise number.

Another example follows us on a daily basis (Rocha, Massad, and Pereira, 2005). In food preparation the older manuals provide recipes that are appropriate enough for cooking delectable foods. A typical recipe calls for "*about* a cup" of this, a "*few* tablespoons" of that, a "*smidgen*" of something, "four *or* five" slices of something else, a "*couple* of medium-sized" other things and "seasoning *to taste*." The recipe goes on to state that this will produce "*about* four servings." This vagueness and ambiguity is not objectionable, but the arithmetic that comes from it is. In italicized print at the end of this older recipe, it's affirmed that the content of the ingredients contains 761 calories, 428 milligrams of sodium, and 22.6 grams of fat per serving. It is inconceivable that these numbers are so precise, and that they could be consistent, in any way, with the recipe the way it is written.

As another example in literature, Marcel Proust was a famous French neuroscientist who wrote about our memories. The style of his sentences was often verbose. In his book about Proust, Jonah Lehrer (2007), states, “Proust covers vast distances within the space of periods (one sentence is 356 words long), and often begins with the obscure detail (the texture of a napkin or the noise of water in the pipes) and ends with an inductive meditation on all things.” The reader would be reasonable to inquire, of what value in this statement is the precision of 356 words?

Requiring precision in engineering models and products translates to requiring high cost and long lead times in production and development. For other than simple systems, expense is proportional to precision: more precision entails higher cost. When considering the use of fuzzy logic for a given problem, an engineer or scientist should ponder the need for *exploiting the tolerance for imprecision*. Not only does high precision dictate high costs, but it also entails low tractability in a problem. Articles in the popular media illustrate the need to exploit imprecision. Take the “traveling sales rep” problem, for example. In this classic optimization problem, a sales representative wants to minimize total distance traveled by considering various itineraries and schedules between a series of cities on a particular trip. For a small number of cities, the problem is a trivial exercise in enumerating all the possibilities and choosing the shortest route. As the number of cities continues to grow, the problem quickly approaches a combinatorial explosion impossible to solve through an exhaustive search, even with a computer. For example, for 100 cities there are $100 \times 99 \times 98 \times 97 \times \ldots \times 2 \times 1$, or about 10^{200}, possible routes to consider! No computers exist today that can solve this problem through a brute-force enumeration of all the possible routes. There are real, practical problems analogous to the traveling sales rep problem. For example, such problems arise in the fabrication of circuit boards, in which precise lasers drill hundreds of thousands of holes in the board. Deciding in which order to drill the holes (where the board moves under a stationary laser) so as to minimize drilling time is a traveling sales rep problem (Kolata, 1991).

Thus, algorithms have been developed to solve the traveling sales rep problem in an optimal sense; that is, the exact answer is not guaranteed but an optimum answer is achievable; the optimality is measured as a percent accuracy, with 0% representing the exact answer and accuracies larger than zero representing answers of lesser accuracy. Suppose we consider a signal routing problem analogous to the traveling sales rep problem in which we want to find the optimum path (i.e., minimum travel time) between 100,000 nodes in a network to an accuracy within 1% of the exact solution; this requires significant CPU time on a supercomputer. If we take the same problem and increase the precision requirement a modest amount to an accuracy of 0.75%, the computing time approaches a few months! Now suppose we can live with an accuracy of 3.5% (quite a bit more accurate than most problems we deal with), and we want to consider an order-of-magnitude more nodes in the network, say 1,000,000; the computing time for this problem is on the order of several minutes (Kolata, 1991). This remarkable reduction in cost (translating time to dollars) is solely the result of the acceptance of a lesser degree of precision in the optimum solution. Can humans live with a little less precision? The answer to this question depends on the situation, but for the vast majority of problems we deal with every day the answer is a resounding yes.

A Historical Perspective

From a historical point of view, the issue of uncertainty has not always been embraced within the scientific community (Klir and Yuan, 1995). In the traditional view of science, uncertainty represents an undesirable state—a state that must be avoided at all costs. This was the state of science until the late nineteenth century when physicists realized that Newtonian mechanics did not address problems at the molecular level. Newer methods, associated with statistical mechanics, were developed, which recognized that statistical averages could replace the specific manifestations of microscopic entities. These statistical quantities, which summarized the activity of large numbers of microscopic entities, could then be connected in a model with appropriate macroscopic variables (Klir and Yuan, 1995). Now, the role of Newtonian mechanics and its underlying calculus, which considered no uncertainty, was replaced with statistical mechanics, which could be described by a probability theory—a theory that could capture a form of uncertainty, the type generally referred to as *random uncertainty*. After the development of statistical mechanics there has been a gradual trend in science during the past century to consider the influence of uncertainty on problems and to do so in an attempt to make our models more robust in the sense that we achieve credible solutions and at the same time quantify the amount of uncertainty.

Of course, the leading theory in quantifying uncertainty in scientific models from the late nineteenth century until the late twentieth century had been the probability theory. However, the gradual evolution of the expression of uncertainty using probability theory was challenged, first in 1937 by Max Black with his studies in vagueness and then with the introduction of fuzzy sets by Zadeh (1965). Zadeh's paper had a profound influence on the thinking about uncertainty because it challenged not only probability theory as the sole representation for uncertainty but also the foundations on which probability theory was based: classical binary (two-valued) logic (Klir and Yuan, 1995).

Probability theory dominated mathematics of uncertainty for more than five centuries. Probability concepts date back to the 1500s, to the time of Cardano when gamblers recognized the rules of probability in games of chance. The concepts were still much in the limelight in 1685, when the Bishop of Wells wrote a paper that discussed a problem in determining the truth of statements made by two witnesses who were both known to be unreliable to the extent that they tell the truth only with probabilities, p_1 and p_2, respectively. The Bishop's answer to this was based on his assumption that the two witnesses were independent sources of information (Lindley, 1987).

Probability theory was initially developed in the eighteenth century in such landmark treatises as Jacob Bernoulli's *Ars Conjectandi* (1713) and Abraham de Moivre's *Doctrine of Chances* (1738). Later in that century, a small number of articles appeared in the periodical literature that would have a profound effect on the field. Most notable of these were Thomas Bayes's "An essay towards solving a problem in the doctrine of chances" (1763) and Pierre Simon Laplace's formulation of the axioms relating to games of chance, "Memoire sur la probabilite des causes par les evenemens" (1774/1986). Laplace, only 25 years old at the time he began his work in 1772, wrote the first substantial article in mathematical statistics before the nineteenth century. Despite the fact that Laplace, at the same time, was heavily engaged in mathematical astronomy, his memoir was an explosion of ideas that provided the roots for modern decision theory, Bayesian inference with nuisance parameters (historians claim that Laplace did not know of Bayes's earlier work), and the asymptotic approximations of posterior distributions (Stigler, 1986).

By the time of Newton, physicists and mathematicians were formulating different theories of probability. The most popular ones remaining today are the relative frequency theory and the subjectivist or personalistic theory. The latter development was initiated by Thomas Bayes (1763), who articulated his powerful theorem for the assessment of subjective probabilities. The theorem specified that a human's degree of belief could be subjected to an objective, coherent, and measurable mathematical framework within subjective probability theory. In the early days of the twentieth century a formal framework for a conditional probability theory was developed.

The twentieth century saw the first developments of alternatives to probability theory and to classical Aristotelian logic as paradigms to address more kinds of uncertainty than just the random kind. Jan Lukasiewicz developed a multivalued, discrete logic (*circa* 1930). In the 1960s, Arthur Dempster (1967) developed a theory of evidence, which, for the first time, included an assessment of ignorance, or the absence of information. In 1965, Lotfi Zadeh introduced his seminal idea in a continuous-valued logic that he called *fuzzy set theory*. Glenn Shafer (1976) extended Dempster's work to produce a complete theory of evidence dealing with information from more than one source, and Lotfi Zadeh illustrated a possibility theory resulting from special cases of fuzzy sets. Later, in the 1980s, other investigators showed a strong relationship between evidence theory, probability theory, and possibility theory with the use of what was called *fuzzy measures* (Klir and Wierman, 1996), and what is now being termed *monotone measures*.

Uncertainty can be thought of in an epistemological sense as being the inverse of information. Information about a particular engineering or scientific problem may be incomplete, imprecise, fragmentary, unreliable, vague, contradictory, or deficient in some other way (Klir and Yuan, 1995). When we acquire more and more information about a problem, we become less and less uncertain about its formulation and solution. Problems that are characterized by little information are said to be ill-posed, complex, or not sufficiently known. These problems are imbued with a high degree of uncertainty. Uncertainty can be manifested in many forms; it can be fuzzy (not sharp, unclear, imprecise, approximate), it can be vague (not specific, amorphous), it can be ambiguous (too many choices, contradictory), it can be of the form of ignorance (dissonant, not knowing something), or it can be a form resulting from natural variability (conflicting, random, chaotic, unpredictable). Many other linguistic labels have been applied to these various forms, but for now these shall suffice. Zadeh (in Ross, Booker, and Parkinson, 2002) posed some simple examples of these forms in terms of a person's statements about when they shall return to a current place in time. The statement "I shall return soon" is vague, whereas the statement "I shall return in a few minutes" is fuzzy; the former is not known to be associated with any unit of time (seconds, hours, days), and the latter is associated with an uncertainty that is at least known to be on the order of minutes. The phrase, "I shall return within 2 minutes of 6 PM" involves an uncertainty that has a quantifiable imprecision; probability theory could address this form.

Vagueness can be used to describe certain kinds of uncertainty associated with linguistic information or intuitive information. Examples of vague information are that the data quality is "good" or that the transparency of an optical element is "acceptable." Moreover, in terms of semantics, even the terms *vague* and *fuzzy* cannot be generally considered synonyms, as explained by Zadeh (1995, p. 275): "usually a vague proposition is fuzzy, but the converse is not generally true."

Discussions about vagueness started with a famous work by the philosopher Max Black. Black (1937) defined a vague proposition as a proposition where the possible states (of the proposition) are not clearly defined with regard to inclusion. For example, consider the proposition that a person is young. Because the term *young* has different interpretations to different individuals, we cannot decisively determine the age(s) at which an individual is young compared with the age(s) at which an individual is not considered to be young. Thus, the proposition is vaguely defined. Classical (binary) logic does not hold under these circumstances, therefore we must establish a different method of interpretation.

Max Black, in writing his 1937 essay "Vagueness: An exercise in logical analysis," first cites remarks made by the ancient philosopher Plato about uncertainty in geometry and then embellishes on the writings of Bertrand Russell (1923) who emphasized that "all traditional logic habitually assumes that precise symbols are being employed." With these great thoughts as a prelude to his own arguments, he proceeded to produce his own, now-famous quote:

> It is a paradox, whose importance familiarity fails to diminish, that the most highly developed and useful scientific theories are ostensibly expressed in terms of objects never encountered in experience. The line traced by a draftsman, no matter how accurate, is seen beneath the microscope as a kind of corrugated trench, far removed from the ideal line of pure geometry. And the "point-planet" of astronomy, the "perfect gas" of thermodynamics, or the "pure-species" of genetics are equally remote from exact realization. Indeed the unintelligibility at the atomic or subatomic level of the notion of a rigidly demarcated boundary shows that such objects not merely are not but could not be encountered. While the mathematician constructs a theory in terms of "perfect" objects, the experimental scientist observes objects of which the properties demanded by theory are and can, in the very nature of measurement, be only approximately true.

Much later, in support of Black's work, Quine (1981) states:

> Diminish a table, conceptually, molecule by molecule: when is a table not a table? No stipulations will avail us here, however arbitrary. If the term 'table' is to be reconciled with bivalence, we must posit an exact demarcation, exact to the last molecule, even though we cannot specify it. We must hold that there are physical objects, coincident except for one molecule, such that one is a table and the other is not.

de Finetti (1974), publishing in his landmark book *Theory of Probability*, gets his readers' attention quickly by proclaiming, "Probability does not exist; it is a subjective description of a person's uncertainty. We should be normative about uncertainty and not descriptive" (p. x). He further emphasizes that the frequentist view of probability (objectivist view) "requires individual trials to be equally probable and stochastically independent" (p. x). In discussing the difference between possibility and probability, he states: "The logic of certainty furnishes us with the range of possibility (and the possible has no gradations); probability is an additional notion that one applies within the range of possibility, thus giving rise to graduations ('more or less' probable) that are meaningless in the logic of uncertainty" (p. 218). de Finetti gives us warnings: "The calculus of probability can say absolutely nothing about reality," and in referring to the dangers implicit in attempts to confuse certainty with high probability, he states:

> We have to stress this point because these attempts assume many forms and are always dangerous. In one sentence: to make a mistake of this kind leaves one inevitably faced with all sorts of

fallacious arguments and contradictions whenever an attempt is made to state, on the basis of probabilistic considerations, that something must occur, or that its occurrence confirms or disproves some probabilistic assumptions (p. 215).

In a discussion about the use of such vague terms as *very probable* or *practically certain*, or *almost impossible*, de Finetti states:

> The field of probability and statistics is then transformed into a Tower of Babel, in which only the most naive amateur claims to understand what he says and hears, and this because, in a language devoid of convention, the fundamental distinctions between what is certain and what is not, and between what is impossible and what is not, are abolished. Certainty and impossibility then become confused with high or low degrees of a subjective probability, which is itself denied precisely by this falsification of the language. On the contrary, the preservation of a clear, terse distinction between certainty and uncertainty, impossibility and possibility, is the unique and essential precondition for making meaningful statements (which could be either right or wrong), whereas the alternative transforms every sentence into a nonsense (p. 213).

The Utility of Fuzzy Systems

Several sources have shown and proven that fuzzy systems are universal approximators (Kosko, 1994; Ying et al., 1999). These proofs stem from the isomorphism between two algebras—an abstract algebra (one dealing with groups, fields, and rings) and a linear algebra (one dealing with vector spaces, state vectors, and transition matrices)—and the structure of a fuzzy system, which comprises an implication between actions and conclusions (antecedents and consequents). The reason for this isomorphism is that both entities (algebra and fuzzy systems) involve a mapping between elements of two or more domains. Just as an algebraic function maps an input variable to an output variable, a fuzzy system maps an input group to an output group; in the latter these groups can be linguistic propositions or other forms of fuzzy information. The foundation on which fuzzy systems theory rests is a fundamental theorem from real analysis in algebra known as the *Stone–Weierstrass theorem*, which was first developed in the late nineteenth century by Weierstrass (1885), then simplified by Stone (1937).

In the coming years it will be the consequence of this isomorphism that will make fuzzy systems more and more popular as solution schemes, and it will make fuzzy systems theory a routine offering in the classroom as opposed to its previous status as a "new, but curious technology." Fuzzy systems, or whatever label scientists eventually come to call it in the future, will be a standard course in any science or engineering curriculum. It contains all of what algebra has to offer, plus more, because it can handle all kinds of information not just numerical quantities.

Although fuzzy systems are shown to be universal approximators to algebraic functions, it is not this attribute that actually makes them valuable to us in understanding new or evolving problems. Rather, the primary benefit of fuzzy systems theory is to approximate system behavior in which analytic functions or numerical relations do not exist. Hence, fuzzy systems have high potential to understand the systems that are devoid of analytic formulations: complex systems. Complex systems can be new systems that have not been tested; they can be systems involved with the human condition such as biological or medical systems; or they can be social, economic, or political systems, in which the vast arrays of inputs and outputs could not all possibly be captured analytically or controlled in any conventional sense. Moreover, the relationship between the causes and effects of these systems is generally not understood but often can be observed.

Alternatively, fuzzy systems theory can have utility in assessing some of our more conventional, less complex systems. For example, for some problems exact solutions are not always necessary. An approximate, but fast, solution can be useful in making preliminary design decisions; or as an initial estimate in a more accurate numerical technique to save computational costs; or in the myriad situations in which the inputs to a problem are vague, ambiguous, or not known at all. For example, suppose we need a controller to bring an aircraft out of a vertical dive. Conventional controllers cannot handle this scenario because they are restricted to linear ranges of variables; a dive situation is highly non-linear. In this case, we could use a fuzzy controller, which is adept at handling nonlinear situations albeit in an imprecise fashion, to bring the plane out of the dive into a more linear range, then hand off the control of the aircraft to a conventional, linear, highly accurate controller. Examples of other situations in which exact solutions are not warranted abound in our daily lives. For example, in the following quote from a popular science fiction movie,

> *C-3PO*: Sir, the possibility of successfully navigating an asteroid field is approximately 3,720 to 1!
> *Han Solo*: Never tell me the odds!
>
> Characters in the movie *Star Wars: The Empire Strikes Back* (Episode V), 1980.

we have an illustration of where the input information (the odds of navigating through an asteroid field) is useless, so how does one make a decision in the presence of this information?

Hence, fuzzy systems are useful in two general contexts: (1) in situations involving highly complex systems whose behaviors are not well understood and (2) in situations where an approximate, but fast, solution is warranted.

As pointed out by Ben-Haim (2001), there is a distinction between models of systems and models of uncertainty. A fuzzy system can be thought of as an aggregation of both because it attempts to understand a system for which no model exists, and it does so with information that can be uncertain in a sense of being vague, or fuzzy, or imprecise, or altogether lacking. Systems whose behaviors are both understood and controllable are of the kind which exhibit a certain robustness to spurious changes. In this sense, robust systems are ones whose output (such as a decision system) does not change significantly under the influence of changes in the inputs because the system has been designed to operate within some window of uncertain conditions. It is maintained that fuzzy systems too are robust. They are robust because the uncertainties contained in both the inputs and outputs of the system are used in formulating the system structure itself, unlike conventional systems analysis that first poses a model, based on a collective set of assumptions needed to formulate a mathematical form, then uncertainties in each of the parameters of that mathematical abstraction are considered.

The positing of a mathematical form for our system can be our first mistake, and any subsequent uncertainty analysis of this mathematical abstraction could be misleading. We call this the *optimist's dilemma*: find out how a chicken clucks by first "assuming a spherical chicken." Once the sphericity of the chicken has been assumed, there are all kinds of elegant solutions that can be found; we can predict any number of sophisticated clucking sounds with our model. Unfortunately, when we monitor a real chicken it does not cluck the way we predict. The point being made here is that there are few physical and no mathematical abstractions that can be made to solve some of our complex problems, so we need new tools to deal with complexity; fuzzy systems and their associated developments can be one of these newer tools.

The power of fuzzy logic in terms of its impact on research and commercial markets is without debate. Thousands of researchers are working with fuzzy logic and producing patents and research papers. According to a report (Singh et al., 2013) on the impact of fuzzy logic as of March 4, 2013, there were 26 research journals on the theory or applications of fuzzy logic, there were 89,365 publications on theory or applications of fuzzy logic in the INSPEC database, there were 22,657 publications on theory or applications of fuzzy logic in the MathSciNet database, there were 16,898 patent applications and patents issued related to fuzzy logic in the United States, and there were 7,149 patent applications and patents issued related to fuzzy logic in Japan. The number of research contributions and commercial applications is growing daily and is growing at an increasing rate.

Limitations of Fuzzy Systems

However, this is not to suggest that we can now stop looking for additional tools to evaluate imprecision or to assess methods for achieving approximate but credible solutions to complex problems. Realistically, even fuzzy systems, as they are posed now, can be described as shallow models in the sense that they are primarily used in deductive reasoning. This is the kind of reasoning in which we infer the specific from the general. For example, in the game of tic-tac-toe, there are only a few moves for the entire game; we can deduce our next move from the previous move and our knowledge of the game. It is this kind of reasoning that we also called *shallow reasoning*, because our knowledge, as expressed linguistically, is of a shallow and meager kind. In contrast to this is the kind of reasoning that is inductive, where we infer the general from the particular; this method of inference is called *deep*, because our knowledge is of a deep and substantial kind—a game of chess would be closer to an inductive kind of model.

We should understand the distinction between using mathematical models to account for observed data and using mathematical models to describe the underlying process by which the observed data are generated or produced by nature (Arciszewski, Sauer, and Schum, 2003). Models of systems where the behavior can be observed, and whose predictions can only account for these observed data, are said to be shallow because they do not account for the underlying realities. Deep models, those of the inductive kind, are alleged to capture the physical process by which nature has produced the results we have observed. In his *Republic* (360 B.C./1991), Plato suggests the idea that things that are perceived are only imperfect copies of the true reality that can only be comprehended by pure thought. Plato was fond of mathematics, and he saw in its precise structure of logic idealized abstraction and separation from the material world. He thought of these things being so important that above the doorway to his Academy was placed the inscription "Let no one ignorant of mathematics enter here." In Plato's doctrine of forms, he argued that the phenomenal world was a mere shadowy image of the eternal, immutable real world, and that matter was docile and disorderly governed by a mind that was the source of coherence, harmony, and orderliness. He argued that if man was occupied with the things of the senses, then he could never gain true knowledge. In his work the *Phaedo*, he declares that as mere mortals we cannot expect to attain absolute truth about the universe, but instead must be content with developing a descriptive picture—a model (Barrow, 2000).

Centuries later, Galileo was advised by his inquisitors that he must not say that his mathematical models were describing the realities of nature, but rather that they simply were adequate models of the observations he made with his telescope (Drake, 1957); hence, that they were solely deductive. In this regard, models that only attempt to replicate some phenomenological behavior are considered shallow models or models of the deductive kind, and they lack the knowledge needed for true understanding of a physical process. The system that emerges under inductive reasoning will have connections with both evolution and complexity. How do humans reason in situations that are complicated or ill-defined? Modern psychology tells us that as humans we are only moderately good at deductive logic, and we make only moderate use of it. Bu, we are superb at seeing or recognizing or matching patterns—behaviors that confer obvious evolutionary benefits. In problems of complication then, we look for patterns, and we simplify the problem by using these to construct temporary internal models or hypotheses or schemata to work with (Bower and Hilgard, 1981). We carry out localized deductions based on our current hypotheses and we act on these deductions. Then, as feedback from the environment comes in, we may strengthen or weaken our beliefs in our current hypotheses, discarding some when they cease to perform, and replacing them as needed with new ones. In other words, where we cannot fully reason or lack full definition of the problem, we use simple models to fill the gaps in our understanding; such behavior is inductive.

Some sophisticated models may, in fact, be a complex weave of deductive and inductive steps. But even our so-called "deep models" may not be deep enough. An illustration of this comes from a recent popular decision problem, articulated as the El Farol problem by Arthur (1994). This problem involves a decision-making scenario in which inductive reasoning is assumed and modeled, and its implications are examined. El Farol is a bar in Santa Fe, New Mexico, where on one night of the week in particular there is popular Irish music offered. Suppose N bar patrons decide independently each week whether to go to El Farol on this certain night. For simplicity, we set $N = 100$. Space in the bar is limited, and the evening is enjoyable if things are not too crowded, specifically, if fewer than 60% of the possible 100 are present. There is no way to tell the number coming for sure in advance, therefore a bar patron goes—deems it worth going—if he expects fewer than 60 to show up or stays home if he expects more than 60 to go; there is no need that utilities differ much above and below 60. Choices are unaffected by previous visits; there is no collusion or prior communication among the bar patrons and the only information available is the number who came in past weeks. Of interest is the dynamics of the number of bar patrons attending from week to week.

There are two interesting features of this problem. First, if there was an obvious model that all bar patrons could use to forecast attendance and on which to base their decisions, then a deductive solution would be possible. But no such model exists in this case. Given the numbers attending in the recent past, a large number of expectational models might be reasonable and defensible. Thus, not knowing which model other patrons might choose, a reference patron cannot choose his in a well-defined way. There is no deductively rational solution, that is, no "correct" expectational model. From the patrons' viewpoint, the problem is ill-defined and they are propelled into a realm of induction. Second, any commonality of expectations gets disintegrated: if everyone believes few will go, then all will go. But this would invalidate that belief. Similarly, if all believe most will go, nobody will go, invalidating that belief. Expectations will be forced to differ, but not in a methodical, predictive way.

Scientists have long been uneasy with the assumption of perfect, deductive rationality in decision contexts that are complicated and potentially ill-defined. The level at which humans can apply perfect rationality is surprisingly modest. Yet, it has not been clear how to deal with imperfect or bounded rationality. From the inductive example given in the El Farol problem, it would be easy to suggest that as humans in these contexts we use inductive reasoning: we induce a variety of working hypotheses, act on the most credible, and replace hypotheses with new ones if they cease to work. Such reasoning can be modeled in a variety of ways. Usually, this leads to a rich psychological world in which peoples' ideas or mental models compete for survival against other peoples' ideas or mental models, a world that is both evolutionary and complex. And, although this seems the best course of action for modeling complex questions and problems, this text reveals a few ideas about models which go beyond those of the rule-based kind. These are briefly introduced in Chapter 12 (genetically evolved fuzzy cognitive maps and fuzzy agent-based models).

The Illusion: Ignoring Uncertainty and Accuracy

A slight variation in the axioms at the foundation of a theory can result in huge changes at the frontier.

Stanley P. Gudder, 1988
Author, *Quantum Probability*

The uninitiated often claim that fuzzy set theory is just another form of probability theory in disguise. This statement, of course, is simply not true. Gaines (1978) does an eloquent job of addressing this issue. Historically, probability and fuzzy sets have been presented as distinct theoretical foundations for reasoning and decision making in situations involving uncertainty. Yet, when one examines the underlying axioms of both probability and fuzzy set theories, the two theories differ by only one axiom in a total of 16 axioms needed for a complete representation! Gaines established a common basis for both forms of logic of uncertainty in which a basic uncertainty logic is defined in terms of valuation on a lattice of propositions. Addition of the axiom of the excluded middle to the basic logic gives a standard probability logic. Alternatively, addition of a requirement for strong truth-functionality gives a fuzzy logic. The quote by Stanley Gudder is quite instructive in this case: probability theory and fuzzy set theory each satisfy a different set of axioms; hence, neither theory should be held to the standards of the others' axiomatic constraints.

Basic statistical analysis is founded on probability theory or stationary random processes, whereas most experimental results contain both random (typically noise) and nonrandom processes. One class of random processes—stationary random processes—exhibits the following three characteristics: (1) The sample space on which the processes are defined cannot change from one experiment to another; that is, the outcome space cannot change. (2) The frequency of occurrence, or probability, of an event within that sample space is constant and cannot change from trial to trial or experiment to experiment. (3) The outcomes must be repeatable from experiment to experiment. The outcome of one trial does not influence the outcome of a previous or future trial. There are more general classes of random processes than the class mentioned here. However, fuzzy sets are not governed by these characteristics.

Stationary random processes are those that arise out of chance, in which the chances represent frequencies of occurrence that can be measured. Problems like picking colored balls out of an urn, coin and dice tossing, and many card games are good examples of stationary random processes. How many of the decisions that humans must make every day could be categorized as random? How about the uncertainty in the weather, is this random? How about your uncertainty in choosing clothes for the next day, or which car to buy, or your preference in colors, are these random uncertainties? How about the risk in whether a substance consumed by an individual now will cause cancer in that individual 15 years from now, is this a form of random uncertainty? Although it is possible to model all of these forms of uncertainty with various classes of random processes, the solutions may not be reliable. Treatment of these forms of uncertainty using fuzzy set theory should also be done with caution. One needs to study the character of the uncertainty and then choose an appropriate approach to develop a model of the process. Features of a problem that vary in time and space should be considered. For example, when the weather report suggests that there is a 60% chance of rain tomorrow, does this mean that there has been rain on tomorrow's date for 60 of the last 100 years? Does it mean that somewhere in your community 60% of the land area will receive rain? Does it mean that 60% of the time it will be raining and 40% of the time it will not be raining? Humans often deal with these forms of uncertainty linguistically, such as, "It will likely rain tomorrow." And, with this crude assessment of the possibility of rain, humans can still make appropriately accurate decisions about the weather.

Random errors will generally average out over time or space. Nonrandom errors, such as some unknown form of bias (often called a *systematic error*) in an experiment, will not generally average out and will likely grow larger with time. The systematic errors generally arise from causes about which we are ignorant, for which we lack information, or that we cannot control. Distinguishing between random and nonrandom errors is a difficult problem in many situations, and to quantify this distinction often results in the illusion that the analyst knows the extent and character of each type of error. In all likelihood, nonrandom errors can increase without bounds. Moreover, variability of the random kind cannot be reduced with additional information, although it can be quantified. By contrast, nonrandom uncertainty, which too can be quantified with various theories, can be reduced with the acquisition of additional information.

It is historically interesting that the word *statistics* is derived from the now-obsolete term *statist*, which means *an expert in statesmanship*. Statistics were the numerical facts that statists used to describe the operations of states. To many people, statistics, and other recent methods to represent uncertainty such as evidence theory and fuzzy set theory, are still the facts by which politicians, newspapers, insurance sellers, and other broker occupations approach us as potential customers for their services or products! The air of sophistication that these methods provide to an issue should not be the basis for making a decision; it should be made only after a good balance has been achieved between the information content in a problem and the proper representation tool to assess it.

Popular lore suggests that the various uncertainty theories allow engineers to fool themselves in a highly sophisticated way when looking at relatively incoherent heaps of data (computational or experimental), as if this form of deception is any more palatable than just plain ignorance. All too often, scientists and engineers are led to use these theories as a crutch to explain vagaries in their models or in their data. For example, in probability applications the assumption of independent random variables is often assumed to provide a simpler method to prescribe joint probability distribution functions. An analogous assumption, called *noninteractive sets* (see Chapter 2 in Ross, 2004), is used in fuzzy applications to develop joint membership functions from individual membership functions for sets from different universes of discourse. Should one ignore apparently aberrant information or consider all information in the model whether or not it conforms to the engineers' preconceptions? Additional experiments to increase understanding cost money, and yet, they might increase the uncertainty by revealing conflicting information. It could best be said that statistics alone, or fuzzy sets alone, or evidence theory alone, are individually insufficient to explain many of the imponderables that people face every day. Collectively they could be powerful. A poem by Cunningham (1971) titled "Meditation on Statistical Method" provides a good lesson in caution for any technologist pondering the thought that ignoring uncertainty (again, using statistics because of the era of the poem) in a problem will somehow make its solution seem more accurate.

> Plato despair!
> We prove by norms
> How numbers bear
> Empiric forms,
>
> How random wrongs
> Will average right
> If time be long
> And error slight;
>
> But in our hearts
> Hyperbole
> Curves and departs
> To infinity.
>
> Error is boundless.
> Nor hope nor doubt,
> Though both be groundless,
> Will average out.

Uncertainty and Information

Information is the resolution of uncertainty.

Claude Shannon, twentieth century mathematician

Only a small portion of knowledge (information) for a typical problem might be regarded as certain or deterministic. Unfortunately, the vast majority of the material taught in engineering classes is based on the presumption that knowledge involved is deterministic. Most processes are neatly and surreptitiously reduced to closed-form algorithms: equations and formulas. When students graduate, it seems that their biggest fear upon entering the real world is "forgetting the correct formula." These formulas typically describe a deterministic process, one where there is no uncertainty in the physics of the process (i.e., the right formula) and there is no uncertainty in the parameters of the process (i.e., the coefficients are known with impunity). It is only after we leave the university, it seems, that we realize we were duped in academia and that the information we have for a particular problem virtually always contains uncertainty. For how many of our problems can we say that the information content is known absolutely, that is, with no ignorance, no vagueness, no imprecision, or no element of chance? Uncertain information can take on many different forms. There is uncertainty that arises because of complexity, for example, the complexity in the reliability network of a nuclear reactor. There is uncertainty that arises from ignorance, from various classes of randomness, from the inability to perform adequate measurements, from lack of knowledge, or from the fuzziness inherent in our natural language.

The nature of uncertainty in a problem is an important point that engineers should ponder before their selection of an appropriate method to express the uncertainty. Fuzzy sets provide a mathematical way to represent vagueness and fuzziness in humanistic systems. For example, suppose you are teaching your child to bake cookies and you want to give instructions about when to take the cookies out of the oven. You could say to take them out when the temperature inside the cookie dough reaches 375° F, or you could advise your child to take them out when the tops of the cookies turn *light brown*. Which instruction would you give? Most likely, you would use the second of the two instructions. The first instruction is too precise to implement practically; in this case precision is not useful. The vague term *light brown* is useful in this context and can be acted on even by a child. We all use vague terms, imprecise information, and other fuzzy data just as easily as we deal with situations governed by chance, where probability techniques are warranted and useful. Hence, our sophisticated computational methods should be able to represent and manipulate a variety of uncertainties. Other representations of uncertainties resulting from ambiguity, nonspecificity, beliefs, and ignorance are introduced in Chapter 13. The one uncertainty that is not addressed in this text is the one termed *unknown*. The statement by a recent U.S. politician, is an interesting diversion that suggests why a method to quantify *unknownness* is perhaps a bit premature.

> The Unknown
> As we know,
> There are known knowns.
> There are things we know we know.
> We also know
> There are known unknowns.
> That is to say
> We know there are some things
> We do not know.
> But there are also unknown unknowns,
> The ones we don't know
> We don't know.
> —Feb. 12, 2002, Donald Rumsfeld, U.S. Secretary of Defense

Fuzzy Sets and Membership

The foregoing sections discuss the various elements of uncertainty. Making decisions about processes that contain nonrandom uncertainty, such as the uncertainty in natural language, has been shown to be less than perfect. The idea proposed by Lotfi Zadeh suggested that *set membership* is the key to decision making when faced with uncertainty. In fact, Zadeh made the following statement in his seminal paper of 1965:

> The notion of a fuzzy set provides a convenient point of departure for the construction of a conceptual framework which parallels in many respects the framework used in the case of ordinary sets, but is more general than the latter and, potentially, may prove to have a much wider scope of applicability, particularly in the fields of pattern classification and information processing. Essentially, such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables (p. 339).

As an example, we can easily assess whether someone is taller than 6 feet. In a binary sense, the person either is or is not, based on the accuracy, or imprecision, of our measuring device. For example, if "tall" is a set defined as heights equal to or greater than 6 feet, a computer would not recognize an individual of height 5′11.999″ as being a member of the set "tall." But how do we assess the uncertainty in the following question: Is the person *nearly* 6 feet tall? The uncertainty in this case is the result of the vagueness or ambiguity of the adjective *nearly*. A 5′11″ person could clearly be a member of the set of "nearly 6 feet tall" people. In the first situation, the uncertainty of whether a person, whose height is unknown, is 6 feet or not is binary; the person either is or is not, and we can produce a probability assessment of that prospect based on height data from many people. But the uncertainty of whether a person is nearly 6 feet is non-random. The degree to which the person approaches a height of 6 feet is fuzzy. In reality, "tallness" is a matter of degree and is relative. Among peoples of the Tutsi tribe in Rwanda and Burundi, a height for a male of 6 feet is considered short. So, 6 feet can be tall in one context and short in another. In the real (fuzzy) world, the set of tall people can overlap with the set of not-tall people, an impossibility when one follows the precepts of classical binary logic (this is discussed in Chapter 5).

This notion of set membership, then, is central to the representation of objects within a universe by sets defined on the universe. Classical sets contain objects that satisfy precise properties of membership; fuzzy sets contain objects that satisfy imprecise properties of membership, that is, membership of an object in a fuzzy set can be approximate. For example, the set of heights *from 5 to 7 feet* is precise (crisp); the set of heights in the region *around 6 feet* is imprecise, or fuzzy. To elaborate, suppose we have an exhaustive collection of individual elements (singletons) x, which make up a universe of information (discourse), X. Further, various combinations of these individual elements make up sets, say A, on the universe. For crisp sets, an element x in the universe X is either a member of some crisp set A or not. This binary issue of membership can be represented mathematically with the indicator function,

$$\chi_A(x) = \begin{cases} 1, & x \in \mathrm{A} \\ 0, & x \notin \mathrm{A} \end{cases}, \tag{1.1}$$

where the symbol $\chi_{\mathrm{A}}(x)$ gives the indication of an unambiguous membership of element x in set A, and the symbols $\in$ and $\notin$ denote contained in and not contained in, respectively. For our example of the universe of heights of people, suppose set A is the crisp set of all people with $5.0 \le x \le 7.0$ feet, shown in Figure 1.1a. A particular individual, x_1, has a height of 6.0 feet. The membership of this individual in crisp set A is equal to 1, or full membership, given symbolically as $\chi_{\mathrm{A}}(x_1) = 1$. Another individual, say x_2, has a height of 4.99 feet. The membership of this individual in set A is equal to 0, or no membership, hence $\chi_{\mathrm{A}}(x_2) = 0$, also seen in Figure 1.1a. In these cases the membership in a set is binary, either an element is a member of a set or it is not.

Zadeh extended the notion of binary membership to accommodate various "degrees of membership" on the real continuous interval [0, 1], where the endpoints of 0 and 1 conform to no membership and full membership, respectively, just as the indicator function does for crisp sets, but where the infinite number of values in between the endpoints can represent various degrees of membership for an element x in some set on the universe. The sets on the universe X that can

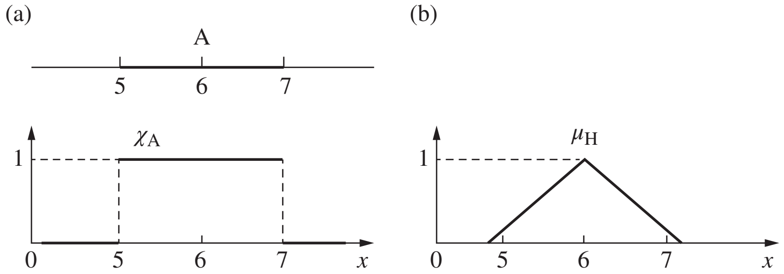

Figure 1.1 Height membership functions for (a) a crisp set A and (b) a fuzzy set H.

accommodate "degrees of membership" were termed by Zadeh as *fuzzy sets*. Continuing further on the example on heights, consider a set H consisting of heights *near 6 feet*. Because the property *near 6 feet* is fuzzy, there is no unique membership function for H. Rather, the analyst must decide what the membership function, denoted μ_H, should look like. Plausible properties of this function might be (1) normality ($\mu_H(6) = 1$), (2) monotonicity (the closer H is to 6, the closer μ_H is to 1), and (3) symmetry (numbers equidistant from 6 should have the same value of μ_H) (Bezdek, 1993). Such a membership function is illustrated in Figure 1.1b. A key difference between crisp and fuzzy sets is their membership function; a crisp set has a unique membership function, whereas a fuzzy set can have an infinite number of membership functions to represent it. For fuzzy sets, the uniqueness is sacrificed, but flexibility is gained because the membership function can be adjusted to maximize the utility for a particular application. It should be noted that a crisp set is a *special case* of a fuzzy set; it is a fuzzy set with no ambiguity on its boundaries.

James Bezdek (1993) provided one of the most lucid comparisons between crisp and fuzzy sets. It bears repeating here. Crisp sets of real objects are equivalent to, and isomorphically described by, a unique membership function, such as χ_A in Figure 1.1a. But there is no set-theoretic equivalent of "real objects" corresponding to χ_A. Fuzzy sets are always *functions*, which map a universe of objects, say X, onto the unit interval [0, 1]; that is, the fuzzy set H is the *function* μ_H that carries X into [0, 1]. Hence, *every* function that maps X onto [0, 1] is a fuzzy set. Although this statement is true in a formal mathematical sense, many functions that qualify on the basis of this definition cannot be suitable fuzzy sets. But, they *become* fuzzy sets when, and only when, they match some intuitively plausible semantic description of imprecise properties of the objects in X.

The membership function embodies the mathematical representation of membership in a set, and the notation used throughout this text for a fuzzy set is a set symbol with a tilde underscore, say $\underset{\sim}{A}$, where the functional mapping is given as

$$\mu_{\underset{\sim}{A}}(x) \in [0, 1], \tag{1.2}$$

and the symbol $\mu_{\underset{\sim}{A}}(x)$ is the degree of membership of element x in fuzzy set $\underset{\sim}{A}$. Therefore, $\mu_{\underset{\sim}{A}}(x)$ is a value on the unit interval that measures the degree to which element x belongs to fuzzy set $\underset{\sim}{A}$; equivalently, $\mu_{\underset{\sim}{A}}(x)$ = degree to which $x \in \underset{\sim}{A}$.

Chance versus Fuzziness

Suppose you are a basketball recruiter and are looking for a "very tall" player for the center position on a men's team. One of your information sources tells you that a hot prospect in Oregon has a 95% chance of being taller than 7 feet. Another of your sources tells you that a good player in Louisiana has a high membership in the set of "very tall" people. The problem with the information from the first source is that it is a probabilistic quantity. There is a 5% chance that the Oregon player is not taller than 7 feet and could, conceivably, be someone of extremely short stature. The second source of information would, in this case, contain a different kind of uncertainty for the recruiter; it is a fuzziness resulting from the linguistic qualifier *very tall* because if the player turned out to be shorter than 7 feet tall there is still a high likelihood that he would be quite tall.

Another example involves a personal choice. Suppose you are seated at a table on which rest two glasses of liquid. The liquid in the first glass is described to you as having a 95% chance of being healthful and good. The liquid in the second glass is described as having a 0.95 membership in the class of "healthful and good" liquids. Which glass would you select, keeping in mind that the first glass has a 5% chance of being filled with nonhealthful liquids, including poisons (Bezdek, 1993)?

What philosophical distinction can be made regarding these two forms of information? Suppose we are allowed to measure the basketball players' heights and test the liquids in the glasses. The prior probability of 0.95 in each case becomes a posterior probability of 1.0 or 0; that is, either the player is or is not taller than 7 feet and the liquid is either benign or not. However, the membership value of 0.95, which measures the extent to which the player's height is taller than 7 feet or the drinkability of the liquid is "healthful and good," remains 0.95 after measuring or testing. These two examples illustrate clearly the difference in the information content between chance and fuzziness.

This brings us to the clearest distinction between fuzziness and chance. Fuzziness describes the lack of distinction of an event, whereas chance describes the uncertainty in the occurrence of the event. The event will occur or not occur; but is the description of the event clear enough to measure its occurrence or nonoccurrence? Consider the following geometric questions, which serve to illustrate our ability to address fuzziness (lack of distinctiveness) with certain mathematical relations. The geometric shape in Figure 1.2a can resemble a disk, a cylinder, or a

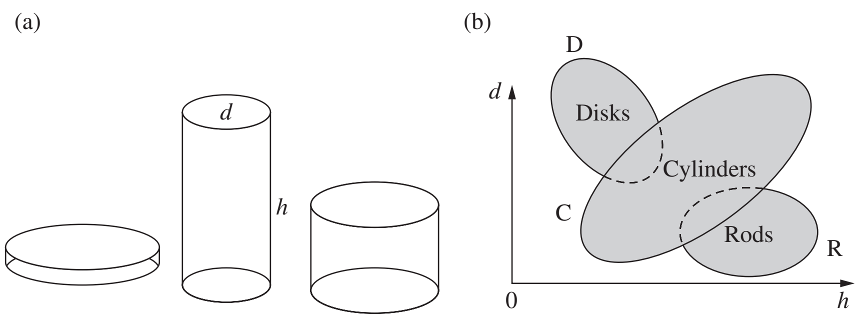

Figure 1.2 Relationship between (a) mathematical terms and (b) fuzzy linguistic terms.

rod depending on the aspect ratio of d/h. For $d/h \ll 1$, the shape of the object approaches a long rod; in fact, as $d/h \to 0$ the shape approaches a line. For $d/h \gg 1$, the object approaches the shape of a flat disk; as $d/h \to \infty$ the object approaches a circular area. For other values of this aspect ratio, for example, for $d/h \approx 1$, the shape is typical of what we would call a *right circular cylinder* (see Figure 1.2b).

The geometric shape in Figure 1.3a is an ellipse, with parameters a and b. Under what conditions of these two parameters will a general elliptic shape become a circle? Mathematically, we know that a circle results when $a/b = 1$, and hence this is a specific, crisp geometric shape. We know that when $a/b \ll 1$ or $a/b \gg 1$, we clearly have an elliptic shape, and as $a/b \to \infty$, a line segment results. Using this knowledge, we can develop a description of the membership function to describe the geometric set we call an approximate circle. Without a theoretical development, the following expression describing a Gaussian curve (for this membership function all points on the real line have nonzero membership; this can be an advantage or disadvantage depending on the nature of the problem) offers a good approximation for the membership function of the fuzzy set "approximate circle," denoted $\underset{\sim}{C}$:

$$\mu_{\underset{\sim}{C}}\left(\frac{a}{b}\right) = \exp\left[-3\left(\frac{a}{b} - 1\right)^2\right] \tag{1.3}$$

Figure 1.3b is a plot of the membership function given in Equation (1.3). As the elliptic ratio a/b approaches a value of unity, the membership value approaches unity; for $a/b = 1$, we have an unambiguous circle. As $a/b \to \infty$ or $a/b \to 0$, we get a line segment; hence, the membership of the shape in the fuzzy set $\underset{\sim}{C}$ approaches zero because a line segment is not similar in shape to a circle. In Figure 1.3b, we see that as we get farther from $a/b = 1$ our membership in the set "approximate circle" gets smaller and smaller. All values of a/b, which have a membership value of unity, are called the *prototypes*; in this case $a/b = 1$ is the only prototype for the set "approximate circle," because at this value it is exactly a circle.

Suppose we were to place in a bag a large number of generally elliptical two-dimensional shapes and ask the question: What is the probability of randomly selecting an "approximate circle" from the bag? We would not be able to answer this question without first assessing the two different kinds of uncertainty. First, we would have to address the issue of fuzziness

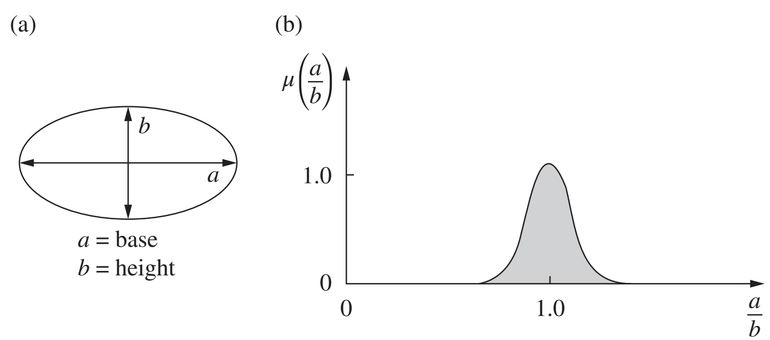

Figure 1.3 The (a) geometric shape and (b) membership function for an approximate circle.

in the meaning of the term *approximate circle* by selecting a value of membership, which we would be willing to call the shape an approximate circle; for example, any shape with a membership value above 0.9 in the fuzzy set "approximate circle" would be considered a circle. Second, we would have to know the proportion of the shapes in the bag that have membership values above 0.9. The first issue is one of assessing fuzziness and the second relates to the frequencies required to address questions of chance.

Intuition of Uncertainty: Fuzzy versus Probability

It is instructive to see how the propagation of uncertainty in a simple nonlinear model can reveal vast differences in the results between a probability model and a fuzzy model, and whether these would conform to our intuition. Suppose we have the simple model, $y = \sin(x)$, and we know that the input parameter, x, is uncertain. We want to model the uncertainty in the input x using a probability density function and also to model the uncertainty in x using a fuzzy membership function. It is important that these two functions look the same, geometrically. Figure 1.4 shows the modeling issues, and results. In Figure 1.4a, the uncertainty in the input is modeled as a uniform probability density function; each element in the universe of the input has equal frequency of occurrence. In Figure 1.4b, the uncertainty in the input is modeled as a fuzzy membership function; here each element in the universe of the input ($-\pi/2$ to $+\pi/2$) has an equal membership of unity.

To show how the uncertainty in the output, y, is determined we make use of two standard propagation approaches. In probability theory this propagation from uncertainty in the input to uncertainty in the output is made by using what is called *derived distributions* (Benjamin and Cornell, 1970). In fuzzy set theory, the propagation of uncertainty in the input to uncertainty in the output is developed using the *extension principle* (Zadeh, 1975). For our model, we have the propagation model, $y = \sin(x)$; we define the uncertainty in the input by a function $f(x)$, and we define the uncertainty in the output by a function $f(y)$. In probability theory we will use the uniform density function, $f(x) = 1/\pi$, to model the input uncertainty. Using the *derived distribution* method, we get the calculus relation,

$$f(x)dx = f(y)dy = \frac{1}{\pi}dx \tag{1.4}$$

Equation (1.4) then becomes, when considering both monotonically increasing and decreasing functions,

$$f(y) = f(x)\left|\frac{1}{dy/dx}\right| \tag{1.5}$$

Now, we define the inverse function of the output, y, as

$$y = \sin x;\text{ hence, } x = \sin^{-1} y \tag{1.6}$$

Combining Equation (1.6) with Equation (1.5) yields,

$$f(y) = f\left(\sin^{-1}y\right)\left|\frac{d\left[\sin^{-1}y\right]}{dy}\right| = f\left(\sin^{-1}y\right)\left|\frac{d[1/\sin y]}{dy}\right| = f(x)\left|\frac{1}{\sqrt{1-y^2}}\right| = \frac{1}{\pi}\left|\frac{1}{\sqrt{1-y^2}}\right| \tag{1.7}$$

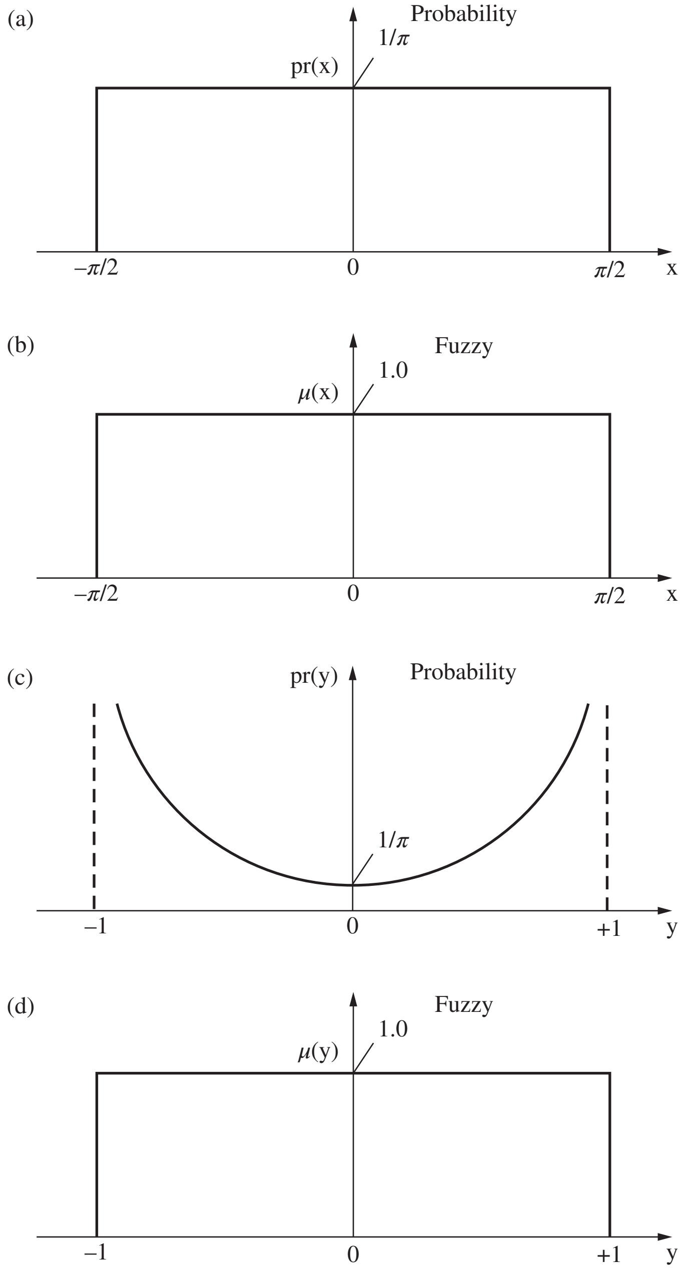

Figure 1.4 Comparison of probability and fuzzy approaches to uncertainty propagation and intuitive understanding of results: (a) input x is uniformly distributed, (b) input x is a fuzzy set with no ambiguity, (c) output y is distributed as a saddle function, and (d) output y is a fuzzy set with no ambiguity.

Now, we want to evaluate Equation (1.7) for the output at the points $y = 0$ and $y = 1$,

$$y(0) = f(0)\left|\frac{1}{\sqrt{1-0}}\right| = \frac{1}{\pi}\cdot 1 = \frac{1}{\pi} \tag{1.8a}$$

$$y(1) = y(-1) = \sin^{-1}(1)\left|\frac{1}{\sqrt{1-1}}\right| = \infty \tag{1.8b}$$

As can be seen in Figure 1.4c the derived output probability density function for y is a saddle function (see Equation 1.7). As seen in Figure 1.4d the output for the fuzzy case is a uniform membership function, which looks just like the input membership function; this fuzzy result comes from Zadeh's *extension principle* which is illustrated in detail in Chapter 12 for a harmonic function. The extension principle is used extensively in the area of fuzzy arithmetic, which is briefly summarized in Chapter 12 of this text.

With regard to the problem just addressed in Figure 1.4 and Equations (1.4 to 1.8) we now encounter the following question. Based on knowing that the input is an uncertain function that is equally frequent at any value in the universe of the input, which of the approaches seems more intuitive? In the case of the probabilistic model, the output with the lowest density value ($f(y) = 1/\pi$) occurs at the mean value ($x = 0$) of the input, and the highest density of the output (∞) occurs at the extremes ($y = \pm 1$) of the output (see Figure 1.4c). In the fuzzy model, the same uncertainty in the input results in the same uncertainty in the output (see Figure 1.4d). It might be clear to the reader which of these models is more intuitive, but at the very least it shows how some models can produce counterintuitive results!

Sets as Points in Hypercubes

There is an interesting geometric analog for illustrating the idea of set membership (Kosko, 1992). Heretofore, we have described a fuzzy set $\underset{\sim}{A}$ defined on a universe X. For a universe with only one element, the membership function is defined on the unit interval [0, 1]; for a two-element universe, the membership function is defined on the unit square; and for a three-element universe, the membership function is defined on the unit cube. All of these situations are shown in Figure 1.5. For a universe of n elements, we define the membership on the unit hypercube, $I^n = [0, 1]^n$.

The endpoints on the unit interval in Figure 1.5a, and the vertices of the unit square and the unit cube in Figure. 1.5b and c, respectively, represent the possible crisp subsets, or collections, of the elements of the universe in each figure. This collection of possible crisp (nonfuzzy) subsets of elements in a universe constitutes the power set of the universe. For example, in Figure 1.5c the universe comprises three elements, $X = \{x_1, x_2, x_3\}$. The point (0, 0, 1) represents the crisp subset in three-space, where x_1 and x_2 have no membership and element x_3 has full membership, that is, the subset $\{x_3\}$; the point (1, 1, 0) is the crisp subset where x_1 and x_2 have full membership and element x_3 has no membership, that is, the subset $\{x_1, x_2\}$; and so on for the other six vertices in Figure 1.5c. In general, there are 2^n subsets in the power set of a universe with n elements; geometrically, this universe is represented by a hypercube in n-space, where the 2^n vertices represent the collection of sets constituting the power set. Two points in the diagrams bear special note, as illustrated in Figure 1.5c. In this figure the point (1, 1, 1),

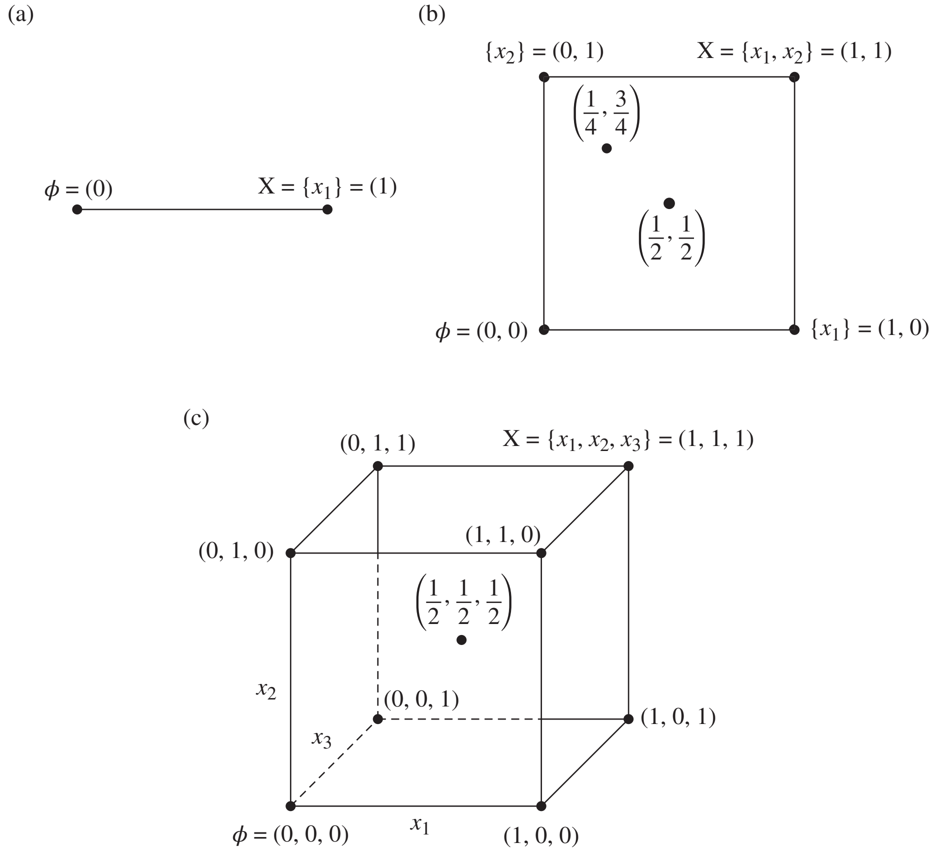

Figure 1.5 "Sets as points" (Kosko, 1992): (a) one-element universe, (b) two-element universe, (c) three-element universe.

where all elements in the universe have full membership, is called the *whole set*, X, and the point (0, 0, 0), where all elements in the universe have no membership, is called the *null set*, Ø.

The centroids of each of the diagrams in Figure 1.5 represent single points where the membership value for each element in the universe equals $\frac{1}{2}$. For example, the point $(\frac{1}{2}, \frac{1}{2})$ in Figure 1.5b is in the midpoint of the square. This midpoint in each of the three figures is a special point; it is the set of maximum "fuzziness." A membership value of $\frac{1}{2}$ indicates that the element belongs to the fuzzy set as much as it does not, that is, it holds equal membership in both the fuzzy set and its complement. In a geometric sense, this point is the location in the space that is farthest from any of the vertices and yet equidistant from all of them. In fact, all points interior to the vertices of the spaces represented in Figure 1.5 represent fuzzy sets, where the membership value of each variable is a number between 0 and 1. For example, in Figure 1.5b, the point $(\frac{1}{4}, \frac{3}{4})$ represents a fuzzy set where variable x_1 has a 0.25 degree of membership in the set and variable x_2 has a 0.75 degree of membership in the set. It is obvious by inspection of the diagrams in Figure 1.5 that, although the number of subsets in the power set is enumerated by the 2^n vertices, the number of fuzzy sets on the universe is infinite, as represented by the infinite number of points on the interior of each space.

Finally, the vertices of the cube in Figure 1.5c are the identical coordinates found in the value set, $V\{P(X)\}$, developed in Example 2.4 of the next chapter.

Summary

This chapter has discussed models with essentially two different kinds of information: fuzzy membership functions, which represent similarities of objects to nondistinct properties, and probabilities, which provide knowledge about relative frequencies. The value of either of these kinds of information in making decisions is a matter of preference; popular, but controversial, contrary views have been offered (Ross et al., 2002). Fuzzy models are *not* replacements for probability models. As seen in Figure 1.1, every crisp set is fuzzy, but the converse does not hold. An example (Fig. 1.4) was given that illustrates that the choice of an uncertainty model can lead to some strange counterintuitive results, and the reader is cautioned to exercise judgment in the selection of the most appropriate model that conforms to the actual uncertainty present in the problem. The idea that crisp sets are special forms of fuzzy sets was illustrated graphically in the section on sets as points, in which crisp sets are represented by the vertices of a unit hypercube. All other points within the unit hypercube, or along its edges, are graphically analogous to a fuzzy set. Fuzzy models are not that different from more familiar models. Sometimes they work better, and sometimes they do not. After all, the efficacy of a model in solving a problem should be the only criterion used to judge that model. Lately, a growing body of evidence suggests that fuzzy approaches to real problems are an effective alternative to previous, traditional methods.

References

Arciszewski, T., Sauer, T., and Schum, D. (2003). Conceptual designing: Chaos-based approach. *J. Intell. Fuzzy Syst.*, **13**: 45–60.

Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *Am. Econ. Rev.*, **84**: 406–411.

Barrow, J. (2000). *The Universe That Discovered Itself* (Oxford: Oxford University Press).

Bayes, T., and Price, M. (1763). An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions* (1683–1775), **53**: 370–418.

Ben-Haim, Y. (2001). *Information Gap Decision Theory: Decisions Under Severe Uncertainty, Series on Decision and Risk* (London: Academic Press).

Benjamin, J., and Cornell, A. (1970). *Probability, Statistics, and Decision for Civil Engineers* (New York: McGraw-Hill).

Bernoulli, J. (1713). *Mathematici Celeberrimi, Ars Conjectandi : Opus Posthumum. Accedit Tractatus De Seriebus Infinitis, Et Epistola Gallicè scripta De Ludo Pilae Reticularis* (Baae: Thurney).

Bezdek, J. (1993). Editorial: Fuzzy models—What are they, and why? *IEEE Trans. Fuzzy Syst.*, **1**: 1–5.

Black, M. (1937). Vagueness: An exercise in logical analysis. *Int. J. Gen. Syst.*, **17**: 107–128.

Bower, G., and Hilgard, E. (1981). *Theories of Learning* (Englewood Cliffs, NJ: Prentice Hall).

Cunningham, J. (1971). *The Collected Poems and Epigrams of J. V. Cunningham* (Chicago: Swallow Press).

de Finetti, B. (1974). *Theory of Probability* (New York: John Wiley & Sons, Ltd.).

de Moivre, A., and Utenhove, J. M. C. v. (1738). *The Doctrine of Chances: Or, a Method of Calculating the Probabilities of Events in Play*, 2nd ed., (London: H. Woodfall.)

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat.*, 325–339.

Drake, S. (1957). *Discoveries and Opinions of Galileo* (New York :Anchor Books).

Gaines, B. (1978). Fuzzy and probability uncertainty logics. *Inf. Control*, **38**: 154–169.

Klir, G., and Wierman, M. (1996). *Uncertainty-Based Information* (Heidelberg, Germany: Physica-Verlag).

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Upper Saddle River, NJ: Prentice Hall).
Kolata, G. (1991). Math problem, long baffling, slowly yields. *New York Times* (Mar 12), p. C1.
Kosko, B. (1992). *Neural Networks and Fuzzy Systems* (Englewood Cliffs, NJ: Prenctice Hall).
Kosko, B. (1994) Fuzzy systems as universal approximators. *IEEE Trans. Comput.*, **43**(11): 1329–1333.
Laplace, P. S. (1774/1986). Memoire sur la probabilite des causes par les evenemens [Memoir on the probability of the causes of events]. *Statist. Sci.*, **1**: 364–378.
Lehrer, J. (2007). *Proust Was a Neuroscientist* (Boston: Houghton Mifflin).
Lindley, D. (1987). Comment: A tale of two wells. *Stat. Sci.*, **2**: 38–40.
Lukasiewicz, J. with A. Tarski, (1930).Untersuchungen über den Aussagenkalkül, *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, cl. iii, **23**: 1–21.
Paulos, J. A. (1995). *A Mathematician Reads the Newspaper* (New York Basic Books).
Plato, & Jowett, B. (360/1991). *The Republic: The Complete and Unabridged Jowett Translation* (New York: Vintage Books).
Quine, W. (1981). *Theories and Things* (Cambridge, MA: Harvard University Press).
Rocha, A., Massad, E., and Pereira, Jr., A. (2005). *The Brain: Fuzzy Arithmetic to Quantum Computing* (Berlin: Springer).
Ross, T. (2004) *Fuzzy Logic with Engineering Applications*, 2nd (Chichester, UK: John Wiley & Sons, Ltd.).
Ross, T., Booker, J., and Parkinson, W. J. (2002). *Fuzzy Logic and Probability Applications: Bridging the Gap* (Philadelphia: Society for Industrial and Applied Mathematics).
Russell, B. (1923). Vagueness. *Australasian J Psych Phil.*, **1**: 88.
Shafer, G. (1976). *A Mathematical Theory of Evidence* (Princeton: Princeton University Press).
Singh, H., Gupta, M. M., Meitzler, T., Hou, Z. G., Garg, K. K., Solo, A. M. G., and Zadeh, L. A. (2013). Real life of applications of fuzzy logic. *Adv. Fuzzy Logic.* http://dx.dio.org/10.1155/2013/581879
Stigler, S. (1986). Laplace's 1774 memoir on inverse probability. *Stat. Sci.*, **1**: 359–378.
Stone, M. H. (1937). Applications of the theory of Boolean rings to general topology. *Trans. Am. Math. Soc.*, **41**: 375–481, 453–481.
Weierstrass, K. (1885). Mathematische Werke, Band 3, Abhandlungen III, 1–37, esp. p. 5, *Sitzungsber. koniglichen preuss. Akad. Wiss.*, July 9 and July 30.
Ying, H., Ding, Y., Li, S., and Shao, S. (1999) Fuzzy systems as universal approximators. *IEEE Trans. Syst. Man. Cybern. A Syst. Hum.*, **29**(5): 508–514.
Zadeh, L. (1965) *Fuzzy sets. Inf. Control*, **8**: 338–353.
Zadeh, L. (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man. Cybern.,* **SMC-3**: 28–44.
Zadeh, L. (1975) The concept of a linguistic variable and its application to approximate reasoning, Part I. *Inf. Sci.*, **8**: 199–249.
Zadeh, L. (1995) Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, **37**: 271–276.

Problems

1.1 Develop a reasonable membership function for the following fuzzy sets based on moving vehicles on a freeway for the speed range of 0–75 mph.
  a. Fast
  b. Moderate
  c. Slow

1.2 A region experiences three seasons during a year. Each of those seasons last for four months starting with winter, then spring, and summer. Develop a membership function for the winter and the summer season on a scale of calendar months.

1.3 For the cylindrical shapes shown in Figure 1.2, develop a membership function for each of the following shapes using the ratio *d/h*, and discuss the reason for any overlapping among the three membership functions:

a. Rod
b. Cylinder
c. Disk

1.4 The question of whether a glass of water is half-full or half-empty is an age-old philosophical issue. Such descriptions of the volume of liquid in a glass depend on the state of mind of the person asked the question. Develop membership functions for the fuzzy sets "half-full," "full." and "half-empty" using a scale of percent of total volume. Assume the maximum volume of water in the glass is V_0. Discuss whether the terms *half-full and half-empty* should have identical membership functions. Does your answer solve this ageless riddle?

1.5 The pH level for a drinking water standard should be between 6 and 8.5. Draw a reasonable membership function for drinking water with an optimum pH standard value on a scale of 0–14.

1.6 To generate electricity, a turbine should rotate "at least at a speed of 40 rpm." Draw a membership function to show the effect of speed on generating electricity using
a. Crisp membership function
b. Fuzzy membership function

1.7 According to Hooke's law: within the elastic limit, stress is directly proportional to strain. A mild steel shows this behavior for a stress up to 335 MPa. Draw both crisp and fuzzy membership functions showing that "mild steel" is within the elastic limit.

1.8 Develop algorithms for the following membership function shapes:
a. Triangular
b. Trapezoid
c. Gaussian

1.9 Water can be classified into three states: solid (ice), liquid (water), and gas (water vapor). These three states are functions of temperature. Draw membership functions for these three states of water in terms of their temperature; use either Celsius or Fahrenheit for your temperature scale.

1.10 A circular column loaded axially is assumed to be eccentric when the load is acting 5% off the axis, depending on the diameter of the column, d. As shown in Figure P1.10 we

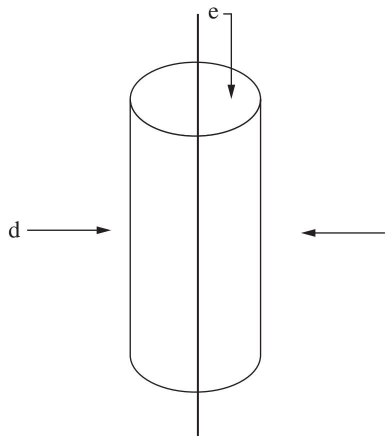

Figure P1.10

have the following conditions: $e/d = 0.05$ eccentric; $e/d < 0.05$ not very eccentric; $e/d > 0.05$ very eccentric. Develop a membership function for "eccentricity" on the scale of e/d ratios.

1.11 If the level of water in a dam is below 110 m height it is said to be a "safe" height. But if the level rises to more than 120 m, which is considered as "dangerous" height, then immediate opening of a gate is required. Draw a membership function for a "safe" water level.

1.12 Probability distributions can be shown to exist on certain planes that intersect the regions shown in Figure 1.5. Draw the points, lines, and planes on which probability distributions exist for the one-, two-, and three-element cases shown in Figure 1.5.

2

Classical Sets and Fuzzy Sets

Philosophical objections may be raised by the logical implications of building a mathematical structure on the premise of fuzziness, since it seems (at least superficially) necessary to require that an object be or not be an element of a given set. From an aesthetic viewpoint, this may be the most satisfactory state of affairs, but to the extent that mathematical structures are used to model physical actualities, it is often an unrealistic requirement.… Fuzzy sets have an intuitively plausible philosophical basis. Once this is accepted, analytical and practical considerations concerning fuzzy sets are in most respects quite orthodox.

James Bezdek, 1981
Professor, Computer Science

Quantum mechanics brought an unexpected fuzziness into physics because of quantum uncertainty, the Heisenberg uncertainty principle.

Edward Witten, twentieth-century American Mathematician

As alluded to in Chapter 1, the *universe of discourse* is the universe of all available information on a given problem. Once this universe is defined we are able to define certain events on this information space. I will describe sets as mathematical abstractions of these events and of the universe itself. Figure 2.1a shows an abstraction of a universe of discourse, say X, and a crisp (classical) set A somewhere in this universe. A classical set is defined by *crisp* boundaries, that is, there is no uncertainty in the prescription or location of the boundaries of the set, as shown in Figure 2.1a, where the boundary of crisp set A is an unambiguous line. A fuzzy set, on the other hand, is prescribed by vague or ambiguous properties; hence, its boundaries are ambiguously specified, as shown by the fuzzy boundary for set $\underset{\sim}{A}$ in Figure 2.1b.

The notion of set membership was introduced, from a one-dimensional viewpoint in Chapter 1. Figure 2.1 again helps to explain this idea, but from a two-dimensional perspective.

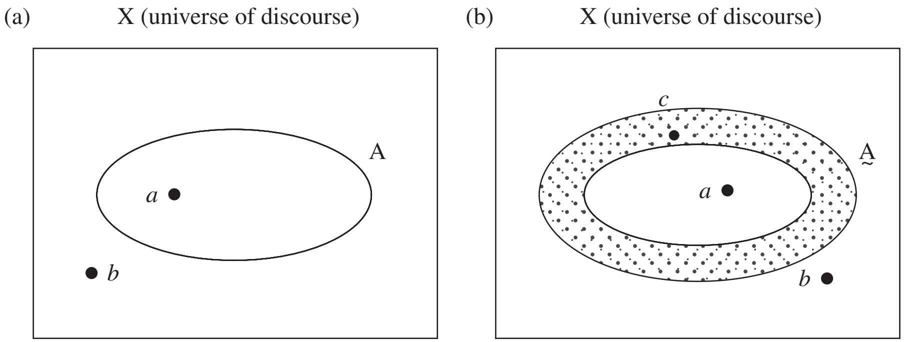

Figure 2.1 Diagrams for (a) crisp set boundary and (b) fuzzy set boundary.

Point *a* in Figure 2.1a is clearly a member of crisp set A; point *b* is unambiguously *not* a member of set A. Figure 2.1b shows the vague, ambiguous boundary of a fuzzy set $\underset{\sim}{A}$ on the same universe X: the shaded boundary represents the boundary region of $\underset{\sim}{A}$. In the central (unshaded) region of the fuzzy set, point *a* is clearly a full member of the set. Outside the boundary region of the fuzzy set, point *b* is clearly not a member of the fuzzy set. However, the membership of point *c*, which is on the boundary region, is ambiguous. If complete membership in a set (such as point *a* in Figure 2.1b) is represented by the number 1, and no membership in a set (such as point *b* in Figure 2.1b) is represented by 0, so then point *c* in Figure 2.1b must have some intermediate value of membership (partial membership in fuzzy set $\underset{\sim}{A}$) on the interval [0, 1]. Presumably, the membership of point *c* in $\underset{\sim}{A}$ approaches a value of 1 as it moves closer to the central (unshaded) region in Figure 2.1b of $\underset{\sim}{A}$ and the membership of point *c* in $\underset{\sim}{A}$ approaches a value of 0 as it moves closer to leaving the boundary region of $\underset{\sim}{A}$.

In this chapter, the precepts and operations of fuzzy sets are compared with those of classical sets. Several good books are available for reviewing this basic material (see for example, Dubois and Prade, 1980; Klir and Folger, 1988; Zimmermann, 1991; Klir and Yuan, 1995). Fuzzy sets embrace virtually all (with one exception, as will be seen) of the definitions, precepts, and axioms that define classical sets. As indicated in Chapter 1, crisp sets are a special case of fuzzy sets; they are sets without ambiguity in their membership (i.e., they are sets with unambiguous boundaries). It will be shown that fuzzy set theory is a mathematically rigorous and comprehensive set theory useful in characterizing concepts (sets) with natural ambiguity. It is instructive to introduce fuzzy sets by first reviewing the elements of classical (crisp) set theory.

Classical Sets

Define a universe of discourse, X, as a collection of objects all having the same characteristics. The individual elements in the universe X will be denoted as *x*. The features of the elements in X can be discrete, countable integers, or continuous valued quantities on the real line. Examples of elements of various universes might be as follows:

the clock speeds of computer CPUs;
the operating currents of an electronic motor;
the operating temperature of a heat pump (in degrees Celsius);

the Richter magnitudes of an earthquake;
the integers 1 to 10.

Most real-world engineering processes contain elements that are real and nonnegative. The first four items just named are examples of such elements. However, for purposes of modeling, most engineering problems are simplified to consider only integer values of the elements in a universe of discourse. So, for example, computer clock speeds might be measured in integer values of megahertz and heat pump temperatures might be measured in integer values of degree Celsius. Further, most engineering processes are simplified to consider only finite-sized universes. Although Richter magnitudes may not have a theoretical limit, we have not historically measured earthquake magnitudes much above 9; this value might be the upper bound in a structural engineering design problem. As another example, suppose you are interested in the stress under one leg of the chair in which you are sitting. You might argue that it is possible to get an infinite stress on one leg of the chair by sitting in the chair in such a manner that only one leg is supporting you and by letting the area of the tip of that leg approach zero. Although this is theoretically possible, in reality the chair leg will either buckle elastically as the tip area becomes small or yield plastically and fail because materials that have infinite strength have not yet been developed. Hence, choosing a universe that is discrete and finite or one that is continuous and infinite is a modeling choice; the choice does not alter the characterization of sets defined on the universe. If elements of a universe are continuous, then sets defined on the universe will be composed of continuous elements. For example, if the universe of discourse is defined as all Richter magnitudes up to a value of 9, then we can define a set of "destructive magnitudes," which might be composed of (1) all magnitudes greater than or equal to a value of 6 in the crisp case or (2) all magnitudes "approximately 6 and higher" in the fuzzy case.

A useful attribute of sets and the universes on which they are defined is a metric known as the *cardinality*, or the *cardinal number*. The total number of elements in a universe X is called its *cardinal number*, denoted n_x, where x again is a label for individual elements in the universe. Discrete universes that are composed of a countably finite collection of elements will have a finite cardinal number; continuous universes comprises an infinite collection of elements will have an infinite cardinality. Collections of elements within a universe are called *sets*, and collections of elements within sets are called *subsets*. Sets and subsets are terms that are often used synonymously because any set is also a subset of the universal set X. The collection of all possible sets in the universe is called the *whole set*.

For crisp sets A and B consisting of collections of some elements in X, the following notation is defined:

$x \in \mathrm{X}$	$\rightarrow$	x belongs to X
$x \in \mathrm{A}$	$\rightarrow$	x belongs to A
$x \notin \mathrm{A}$	$\rightarrow$	x does not belong to A

For sets A and B on X, we also have

$\mathrm{A} \subset \mathrm{B}$	$\rightarrow$	A is fully contained in B (if $x \in \mathrm{A}$, then $x \in \mathrm{B}$)
$\mathrm{A} \subseteq \mathrm{B}$	$\rightarrow$	A is contained in or is equivalent to B
$(\mathrm{A} \leftrightarrow \mathrm{B})$	$\rightarrow$	$\mathrm{A} \subseteq$ Band $\mathrm{B} \subseteq \mathrm{A}$ (A is equivalent to B)

We define the *null set*, ∅, as the set containing no elements, and the whole set, X, as the set of all elements in the universe. The null set is analogous to an impossible event, and the whole set is analogous to a certain event. All possible sets of X constitute a special set called the *power set*, P(X). For a specific universe X, the power set P(X) is enumerated in the following example.

Example 2.1
We have a universe composed of three elements, $X = \{a, b, c\}$, so the cardinal number is $n_x = 3$. The power set is

$$\mathrm{P(X)} = \{\emptyset,\ \{a\},\ \{b\},\ \{c\},\ \{a,b\},\ \{a,c\},\ \{b,c\},\ \{a,b,c\}\}.$$

The cardinality of the power set, denoted $n_{\mathrm{P(X)}}$, is found as

$$n_{\mathrm{P(X)}} = 2^{n_\mathrm{X}} = 2^3 = 8.$$

Note that if the cardinality of the universe is infinite, then the cardinality of the power set is also infinity, that is, $n_\mathrm{X} = \infty \Rightarrow n_{\mathrm{P(X)}} = \infty$.

Operations on Classical Sets

Let A and B be two sets on the universe X. The union between the two sets, denoted A ∪ B, represents all those elements in the universe that reside in (or belong to) the set A, the set B, or both sets A and B. (This operation is also called the *logical or*; another form of the union is the *exclusive or* operation. The *exclusive or* is described in Chapter 5.) The intersection of the two sets, denoted A ∩ B, represents all those elements in the universe X that simultaneously reside in (or belong to) both sets A and B. The complement of a set A, denoted Ā, is defined as the collection of all elements in the universe that do not reside in the set A. The difference of a set A with respect to B, denoted A | B, is defined as the collection of all elements in the universe that reside in A and that do not reside in B simultaneously. These operations are shown below in set-theoretic terms.

Union $$\mathrm{A} \cup \mathrm{B} = \{x \mid x \in \mathrm{A} \text{ or } x \in \mathrm{B}\}. \tag{2.1}$$

Intersection $$\mathrm{A} \cap \mathrm{B} = \{x \mid x \in \mathrm{A} \text{ and } x \in \mathrm{B}\}. \tag{2.2}$$

Complement $$\overline{\mathrm{A}} = \{x \mid x \notin \mathrm{A},\ x \in \mathrm{X}\}. \tag{2.3}$$

Difference $$\mathrm{A} \mid \mathrm{B} = \{x \mid x \in \mathrm{A} \text{ and } x \notin \mathrm{B}\}. \tag{2.4}$$

These four operations are shown in terms of Venn diagrams in Figures 2.2–2.5.

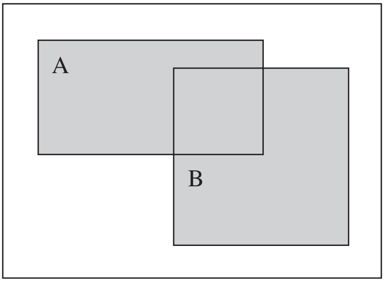

Figure 2.2 Union of sets A and B (logical or).

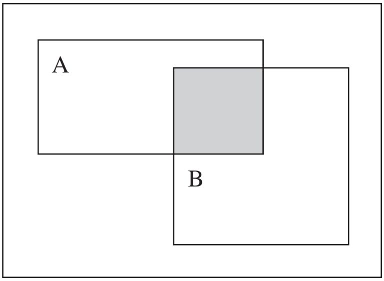

Figure 2.3 Intersection of sets A and B.

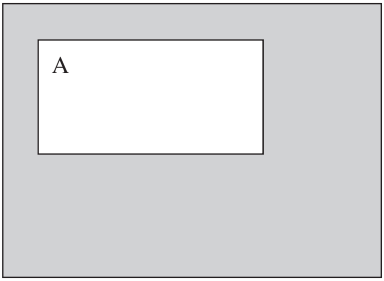

Figure 2.4 Complement of set A.

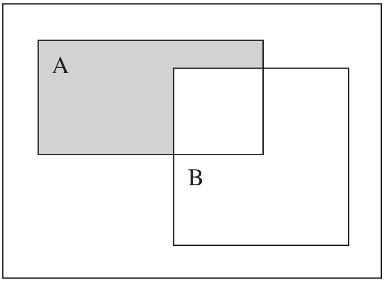

Figure 2.5 Difference operation A|B.

Properties of Classical (Crisp) Sets

Certain properties of sets are important because of their influence on the mathematical manipulation of sets. The most appropriate properties for defining classical sets and showing their similarity to fuzzy sets are as follows:

Commutativity

$$\begin{aligned} A \cup B &= B \cup A \\ A \cap B &= B \cap A. \end{aligned} \tag{2.5}$$

Associativity

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C \\ A \cap (B \cap C) &= (A \cap B) \cap C. \end{aligned} \tag{2.6}$$

Distributivity

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C). \end{aligned} \tag{2.7}$$

Idempotency

$$\begin{aligned} A \cup A &= A \\ A \cap A &= A. \end{aligned} \tag{2.8}$$

Identity

$$\begin{aligned} A \cup \emptyset &= A \\ A \cap X &= A \\ A \cap \emptyset &= \emptyset. \\ A \cup X &= X. \end{aligned} \tag{2.9}$$

Transitivity

$$\text{If } A \subseteq B \text{ and } B \subseteq C, \text{ then } A \subseteq C. \tag{2.10}$$

Involution

$$\overline{\overline{A}} = A. \tag{2.11}$$

Two special properties of set operations are known as the *excluded middle axioms* and *De Morgan's principles*. These properties are en umerated here for two sets A and B. The *excluded middle axioms* are important because these are the only set operations described here that are *not* valid for both classical sets and fuzzy sets. There are two excluded middle axioms, which are given in Equation (2.12). The first, called the *axiom of the excluded middle*, deals with the union of a set A and its complement; the second, called the *axiom of contradiction*, represents the intersection of a set A and its complement.

Axiom of the excluded middle

$$A \cup \overline{A} = X. \tag{2.12a}$$

Axiom of the contradiction

$$A \cap \overline{A} = \emptyset. \tag{2.12b}$$

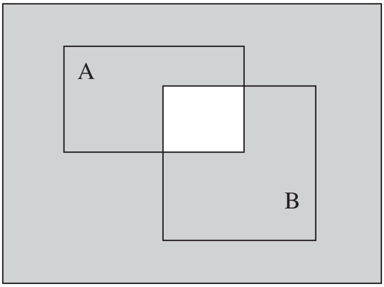

Figure 2.6 De Morgan's principle $\left(\overline{A \cap B}\right)$.

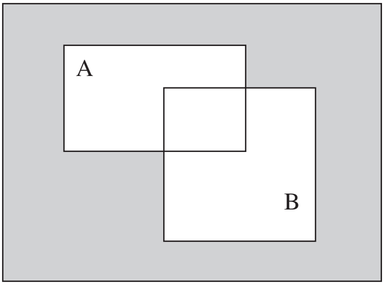

Figure 2.7 De Morgan's principle $\left(\overline{A \cup B}\right)$.

De Morgan's principles are important because of their usefulness in proving tautologies and contradictions in logic, as well as in a host of other set operations and proofs. De Morgan's principles are displayed in the shaded areas of the Venn diagrams in Figures 2.6 and 2.7 and described mathematically in Equation (2.13).

$$\overline{A \cap B} = \overline{A} \cup \overline{B}. \tag{2.13a}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}. \tag{2.13b}$$

In general, De Morgan's principles can be stated for n sets, as provided here for events, E_i:

$$\overline{E_1 \cup E_2 \cup \cdots \cup E_n} = \overline{E_1} \cap \overline{E_2} \cap \cdots \cap \overline{E_n}. \tag{2.14a}$$

$$\overline{E_1 \cap E_2 \cap \cdots \cap E_n} = \overline{E_1} \cup \overline{E_2} \cup \cdots \cup \overline{E_n}. \tag{2.14b}$$

From the general equations, which are given in Equation (2.14), for De Morgan's principles, there is a duality relation: the complement of a union or an intersection is equal to the intersection or union, respectively, of the respective complements. This result is powerful in dealing with set structures because we often have information about the complement of a set (or event) or the complement of combinations of sets (or events), rather than information about the sets themselves.

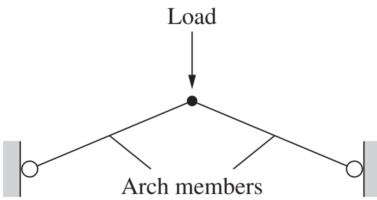

Figure 2.8 A two-member arch.

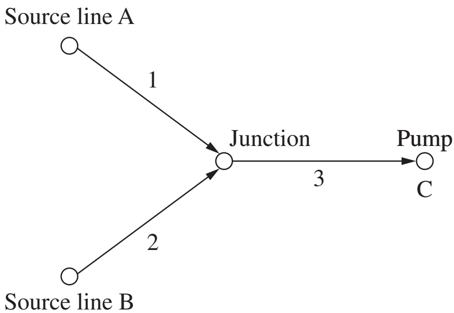

Figure 2.9 Hydraulic hose system.

Example 2.2
A shallow arch consists of two slender members as shown in Figure 2.8. If either of the members fails, then the arch will collapse. If E_1 = survival of member 1 and E_2 = survival of member 2, then survival of the arch = $E_1 \cap E_2$, and, conversely, collapse of the arch = $\overline{E_1 \cap E_2}$. Logically, collapse of the arch will occur if either of the members fails, that is, when $\overline{E_1} \cup \overline{E_2}$. Therefore,

$$\overline{E_1 \cap E_2} = \overline{E_1} \cup \overline{E_2},$$

which is an illustration of De Morgan's principle.

As Equation (2.14) suggests, De Morgan's principles are useful for compound events, as illustrated in the following example.

Example 2.3
For purposes of safety, the fluid supply for a hydraulic pump C in an airplane comes from two redundant source lines, A and B. The fluid is transported by high-pressure hoses consisting of branches 1, 2, and 3, as shown in Figure 2.9. Operating specifications for the pump indicate that either source line alone is capable of supplying the necessary fluid pressure to the pump. Denote E_1 = failure of branch 1, E_2 = failure of branch 2, and E_3 = failure of branch 3. Then insufficient pressure to operate the pump would be caused by $(E_1 \cap E_2) \cup E_3$, and sufficient pressure would

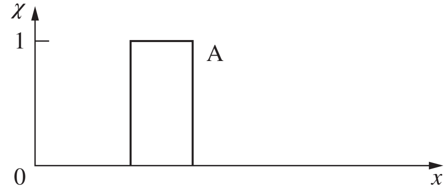

Figure 2.10 Membership function is a mapping for crisp set A.

be the complement of this event. Using De Morgan's principles, we can calculate the condition of sufficient pressure to be

$$\overline{(E_1 \cap E_2) \cup E_3} = \left(\overline{E_1} \cup \overline{E_2}\right) \cap \overline{E_3},$$

in which $\left(\overline{E_1} \cup \overline{E_2}\right)$ means the availability of pressure at the junction, and $\overline{E_3}$ means the absence of failure in branch 3.

Mapping of Classical Sets to Functions

Mapping is an important concept in relating set-theoretic forms to function-theoretic representations of information. In its most general form it can be used to map elements or subsets in one universe of discourse to elements or sets in another universe. Suppose X and Y are two different universes of discourse (information). If an element x is contained in X and corresponds to an element y contained in Y, it is generally termed a mapping from X to Y, or f: X → Y. As a mapping, the characteristic (indicator) function χ_{A} is defined as

$$\chi_{\mathrm{A}}(x) = \begin{cases} 1, & x \in \mathrm{A} \\ 0, & x \notin \mathrm{A} \end{cases}, \tag{2.15}$$

where χ_{A} expresses "membership" in set A for the element x in the universe. This membership idea is a mapping from an element x in universe X to one of the two elements in universe Y, that is, to the elements 0 or 1, as shown in Figure 2.10.

For any set A defined on the universe X, there exists a function-theoretic set, called a *value set*, denoted V(A), under the mapping of the characteristic function, χ. By convention, the null set ∅ is assigned the membership value 0 and the whole set X is assigned the membership value 1.

Example 2.4

Continuing with the example (Example 2.1) of a universe with three elements, $X = \{a, b, c\}$, we desire to map the elements of the power set of X, that is, P(X), to a universe, Y, consisting of only two elements (the characteristic function),

$$\mathrm{Y} = \{0, 1\}.$$

As before, the elements of the power set are enumerated.

$$\mathrm{P(X)} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{b,c\}, \{a,c\}, \{a,b,c\}\}.$$

Thus, the elements in the value set V(A) as determined from the mapping are

$$\mathrm{V\{P(X)\}} = \{\{0, 0, 0\}, \{1, 0, 0\}, \{0, 1, 0\}, \{0, 0, 1\}, \{1, 1, 0\}, \{0, 1, 1\}, \{1, 0, 1\}, \{1, 1, 1\}\}.$$

For example, the third subset in the power set P(X) is the element b. For this subset there is no a, so a value of 0 goes in the first position of the data triplet; there is a b, so a value of 1 goes in the second position of the data triplet; and there is no c, so a value of 0 goes in the third position of the data triplet. Hence, the third subset of the value set is the data triplet, $\{0, 1, 0\}$, as already seen. The value set has a graphical analog that is described in Chapter 1 in the section Sets as Points in Hypercubes.

Now, define two sets, A and B, on the universe X. The union of these two sets in terms of function-theoretic terms is given as follows (the symbol $\vee$ is the maximum operator and $\wedge$ is the minimum operator):

Union $$\mathrm{A} \cup \mathrm{B} \rightarrow \chi_{\mathrm{A}\cup\mathrm{B}}(x) = \chi_{\mathrm{A}}(x) \vee \chi_{\mathrm{B}}(x) = \max(\chi_{\mathrm{A}}(x), \chi_{\mathrm{B}}(x)). \tag{2.16}$$

The intersection of these two sets in function-theoretic terms is given as follows:

Intersection $$\mathrm{A} \cap \mathrm{B} \rightarrow \chi_{\mathrm{A}\cap\mathrm{B}}(x) = \chi_{\mathrm{A}}(x) \wedge \chi_{\mathrm{B}}(x) = \min(\chi_{\mathrm{A}}(x), \chi_{\mathrm{B}}(x)). \tag{2.17}$$

The complement of a single set on universe X, say A, is given as follows:

Complement $$\overline{\mathrm{A}} \rightarrow \chi_{\overline{\mathrm{A}}}(x) = 1 - \chi_{\mathrm{A}}(x). \tag{2.18}$$

For two sets on the same universe, say A and B, if one set (A) is contained in another set (B), then

Containment $$\mathrm{A} \subseteq \mathrm{B} \rightarrow \chi_{\mathrm{A}}(x) \leq \chi_{\mathrm{B}}(x). \tag{2.19}$$

Function-theoretic operators for union and intersection (other than maximum and minimum, respectively) are discussed in the literature (Gupta and Qi, 1991).

Fuzzy Sets

In classical, or crisp, sets the transition for an element in the universe between membership and non-membership in a given set is abrupt and well defined (said to be *crisp*). For an element in a universe that contains fuzzy sets, this transition can be gradual. This transition among various degrees of membership can be thought of as conforming to the fact that the boundaries of the fuzzy sets are vague and ambiguous. Hence, membership of an element from the universe in this set is measured by a function that attempts to describe vagueness and ambiguity.

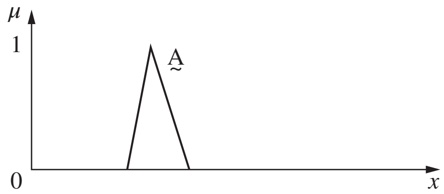

Figure 2.11 Membership function for fuzzy set $\underset{\sim}{A}$.

A fuzzy set, then, is a set containing elements that have varying degrees of membership in the set. This idea is in contrast with classical, or crisp, sets because members of a crisp set would not be members unless their membership is full, or complete, in that set (i.e., their membership is assigned a value of 1). Elements in a fuzzy set, because their membership need not be complete, can also be members of other fuzzy sets on the same universe.

Elements of a fuzzy set are mapped to a universe of *membership values* using a function-theoretic form. As mentioned in Chapter 1, Equation (1.2), fuzzy sets are denoted in this text by a set symbol with a tilde understrike. So, for example, $\underset{\sim}{A}$ would be the *fuzzy set A*. This function maps elements of a fuzzy set $\underset{\sim}{A}$ to a real numbered value on the interval 0–1. If an element in the universe, say x, is a member of fuzzy set $\underset{\sim}{A}$, then this mapping is given by Equation (1.2), or $\mu_{\underset{\sim}{A}}(x) \in [0, 1]$. This mapping is shown in Figure 2.11 for a typical fuzzy set.

A notation convention for fuzzy sets when the universe of discourse, X, is discrete and finite, is as follows for a fuzzy set $\underset{\sim}{A}$:

$$\underset{\sim}{A} = \left\{ \frac{\mu_{\underset{\sim}{A}}(x_1)}{x_1} + \frac{\mu_{\underset{\sim}{A}}(x_2)}{x_2} + \cdots \right\} = \left\{ \sum_i \frac{\mu_{\underset{\sim}{A}}(x_i)}{x_i} \right\}. \tag{2.20}$$

When the universe, X, is continuous and infinite, the fuzzy set $\underset{\sim}{A}$ is denoted as

$$\underset{\sim}{A} = \left\{ \int \frac{\mu_{\underset{\sim}{A}}(x)}{x} \right\}. \tag{2.21}$$

In both notations, the horizontal bar is not a quotient but rather a delimiter. The numerator in each term is the membership value in set $\underset{\sim}{A}$ associated with the element of the universe indicated in the denominator. In the first notation, the summation symbol is not for algebraic summation, but rather denotes the collection or aggregation of each element; hence, the "+" signs in the first notation are not the algebraic "add" but are an aggregation or collection operator. In the second notation, the integral sign is not an algebraic integral but a continuous function-theoretic aggregation operator for continuous variables. Both notations are due to Zadeh (1965).

Fuzzy Set Operations

Define three fuzzy sets $\underset{\sim}{A}$, $\underset{\sim}{B}$, and $\underset{\sim}{C}$ on the universe X. For a given element x of the universe, the following function-theoretic operations for the set-theoretic operations of union, intersection, and complement are defined for $\underset{\sim}{A}$, $\underset{\sim}{B}$,and $\underset{\sim}{C}$ on X:

Union $$\mu_{\underset{\sim}{A}\cup\underset{\sim}{B}}(x)=\mu_{\underset{\sim}{A}}(x)\vee\mu_{\underset{\sim}{B}}(x). \tag{2.22}$$

Intersection $$\mu_{\underset{\sim}{A}\cap\underset{\sim}{B}}(x)=\mu_{\underset{\sim}{A}}(x)\wedge\mu_{\underset{\sim}{B}}(x). \tag{2.23}$$

Complement $$\mu_{\overline{\underset{\sim}{A}}}(x)=1-\mu_{\underset{\sim}{A}}(x). \tag{2.24}$$

Venn diagrams for these operations, extended to consider fuzzy sets, are shown in Figures 2.12–2.14. The operations given in Equations (2.22) to (2.24) are known as the

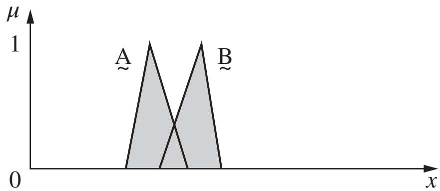

Figure 2.12 Union of fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$.

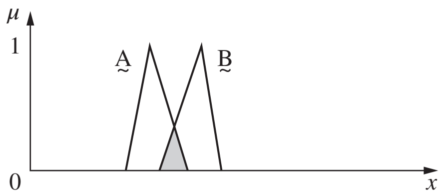

Figure 2.13 Intersection of fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$.

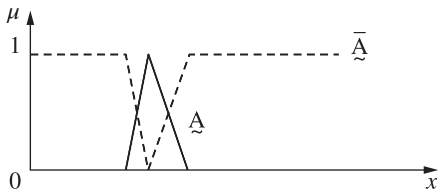

Figure 2.14 Complement of fuzzy set $\underset{\sim}{A}$.

standard fuzzy operations. There are many other fuzzy operations, and a discussion of these is given later in this chapter.

Any fuzzy set $\underset{\sim}{\mathrm{A}}$ defined on a universe X is a subset of that universe. Also by definition, just as with classical sets, the membership value of any element x in the null set $\varnothing$ is 0, and the membership value of any element x in the whole set X is 1. Note that the null set and the whole set are not fuzzy sets in this context (no tilde understrike). The appropriate notation for these ideas is as follows:

$$\underset{\sim}{\mathrm{A}} \subseteq \mathrm{X} \Rightarrow \mu_{\underset{\sim}{\mathrm{A}}}(x) \le \mu_{\mathrm{X}}(x). \tag{2.25a}$$

$$\text{For all } x \in \mathrm{X},\ \mu_{\varnothing}(x) = 0. \tag{2.25b}$$

$$\text{For all } x \in \mathrm{X},\ \mu_{\mathrm{X}}(x) = 0. \tag{2.25c}$$

The collection of all fuzzy sets and fuzzy subsets on X is denoted as the fuzzy power set $\mathrm{P}(\underset{\sim}{\mathrm{X}})$. It should be obvious, based on the fact that all fuzzy sets can overlap, that the cardinality, $n_{\mathrm{P(X)}}$, of the fuzzy power set is infinite; that is, $n_{\mathrm{P(X)}} = \infty$.

De Morgan's principles for classical sets also hold for fuzzy sets, as denoted by the following expressions:

$$\overline{\underset{\sim}{\mathrm{A}} \cap \underset{\sim}{\mathrm{B}}} = \overline{\underset{\sim}{\mathrm{A}}} \cup \overline{\underset{\sim}{\mathrm{B}}}. \tag{2.26a}$$

$$\overline{\underset{\sim}{\mathrm{A}} \cup \underset{\sim}{\mathrm{B}}} = \overline{\underset{\sim}{\mathrm{A}}} \cap \overline{\underset{\sim}{\mathrm{B}}}. \tag{2.26b}$$

As enumerated before, all other operations on classical sets also hold for fuzzy sets, except for the excluded middle axioms. These two axioms do not hold for fuzzy sets because they do not form part of the basic axiomatic structure of fuzzy sets (see Gaines, 1978). Since fuzzy sets can overlap, a set and its complement can also overlap. The *excluded middle axioms*, extended for fuzzy sets, are expressed as

$$\underset{\sim}{\mathrm{A}} \cup \overline{\underset{\sim}{\mathrm{A}}} \neq \mathrm{X}. \tag{2.27a}$$

$$\underset{\sim}{\mathrm{A}} \cap \overline{\underset{\sim}{\mathrm{A}}} \neq \varnothing. \tag{2.27b}$$

Extended Venn diagrams comparing the *excluded middle axioms* for classical (crisp) sets and fuzzy sets are shown in Figures 2.15 and 2.16, respectively.

Properties of Fuzzy Sets

Fuzzy sets follow the same properties as crisp sets. Because of this and because the membership values of a crisp set are a subset of the interval [0, 1], classical sets can be thought of as a special case of fuzzy sets. Hence, the properties given in Equations (2.5) through (2.11) are identical to those for fuzzy sets.

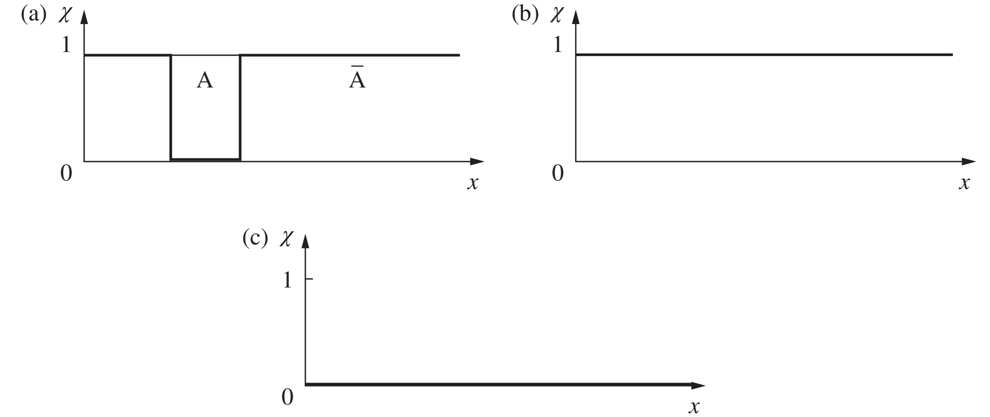

Figure 2.15 Excluded middle axioms for crisp sets. (a) Crisp set A and its complement; (b) crisp $A \cup \overline{A} = X$ (axiom of excluded middle); and (c) crisp $A \cap \overline{A} = \emptyset$ (axiom of contradiction).

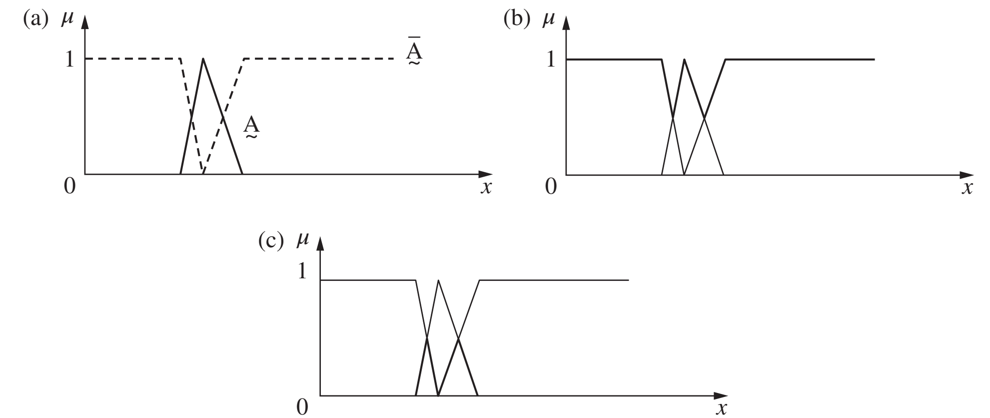

Figure 2.16 Excluded middle axioms for fuzzy sets are not valid. (a) Fuzzy set $\underset{\sim}{A}$ and its complement; (b) fuzzy $\underset{\sim}{A} \cap \overline{\underset{\sim}{A}} \neq X$. (axiom of excluded middle); and (c) fuzzy $A \cap \overline{A} \neq \emptyset$ (axiom of contradiction).

Example 2.5

Consider a simple hollow shaft of approximately 1-m radius and wall thickness $1/(2\pi)$ m. The shaft is built by stacking a ductile section, D, of the appropriate cross section over a brittle section, B, as shown in Figure 2.17. A downward force P and a torque T are simultaneously applied to the shaft. Because of the dimensions chosen, the nominal shear stress on any element in the shaft is T (pascals) and the nominal vertical component of stress in the shaft is P (pascals). We also assume that the failure properties of both B and D are not known with any certainty.

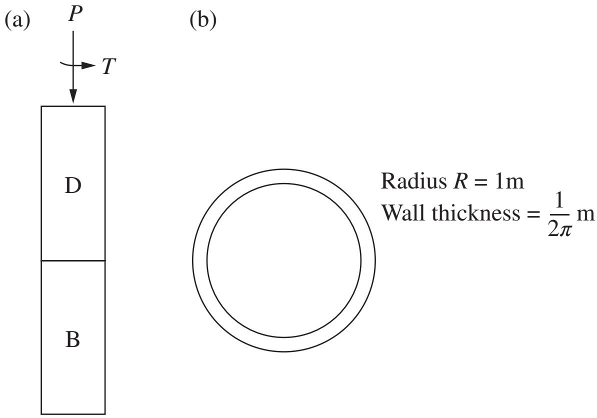

Figure 2.17 (a) Axial and (b) cross-sectional views of example hollow shaft.

We define the fuzzy set $\underset{\sim}{A}$ to be the region in (P, T) space for which material D is "safe," using as a metric the failure function $\mu_A = f([P^2 + 4T^2]^{1/2})$. Similarly, we define the set $\underset{\sim}{B}$ to be the region in (P, T) space for which material B is "safe" using as a metric the failure function $\mu_B = g(P - \beta|T|)$, where β is an assumed material parameter. The functions f and g will, of course, be membership functions on the interval [0, 1]. Their exact specification is not important at this point. What is useful, however, before specifying f and g, is to discuss the basic set operations in the context of this problem. This discussion is summarized as follows:

1. $\underset{\sim}{A} \cup \underset{\sim}{B}$ is the set of loadings for which one expects that either material B or material D will be "safe."
2. $\underset{\sim}{A} \cap \underset{\sim}{B}$ is the set of loadings for which one expects that both material B and material D are "safe."
3. $\overline{\underset{\sim}{A}}$ and $\overline{\underset{\sim}{B}}$ are the sets of loadings for which material D and material B are unsafe, respectively.
4. $\underset{\sim}{A}|\underset{\sim}{B}$ is the set of loadings for which the ductile material is safe but the brittle material is in jeopardy.
5. $\underset{\sim}{B}|\underset{\sim}{A}$ is the set of loadings for which the brittle material is safe but the ductile material is in jeopardy.
6. De Morgan's principle $\overline{\underset{\sim}{A} \cap \underset{\sim}{B}} = \overline{\underset{\sim}{A}} \cup \overline{\underset{\sim}{B}}$ asserts that the loadings that are not safe with respect to both materials are the union of those that are unsafe with respect to the brittle material with those that are unsafe with respect to the ductile material.
7. De Morgan's principle $\overline{\underset{\sim}{A} \cup \underset{\sim}{B}} = \overline{\underset{\sim}{A}} \cap \overline{\underset{\sim}{B}}$ asserts that the loads that are safe for neither material D nor material B are the intersection of those that are unsafe for material D with those that are unsafe for material B.

To illustrate these ideas numerically, let us say we have two discrete fuzzy sets, namely,

$$\underset{\sim}{A} = \left\{\frac{0}{1} + \frac{1}{2} + \frac{0.5}{3} + \frac{0.3}{4} + \frac{0.2}{5}\right\} \text{ and } \underset{\sim}{B} = \left\{\frac{0}{1} + \frac{0.5}{2} + \frac{0.7}{3} + \frac{0.2}{4} + \frac{0.4}{5}\right\}.$$

We can now calculate several of the operations just discussed (membership for element 1 in both $\underset{\sim}{A}$ and $\underset{\sim}{B}$ is implicitly 0):

Complement

$$\overline{\underset{\sim}{A}} = \left\{ \frac{1}{1} + \frac{0}{2} + \frac{0.5}{3} + \frac{0.7}{4} + \frac{0.8}{5} \right\}.$$

$$\overline{\underset{\sim}{B}} = \left\{ \frac{1}{1} + \frac{0.5}{2} + \frac{0.3}{3} + \frac{0.8}{4} + \frac{0.6}{5} \right\}.$$

Union

$$\underset{\sim}{A} \cup \underset{\sim}{B} = \left\{ \frac{0}{1} + \frac{1}{2} + \frac{0.7}{3} + \frac{0.3}{4} + \frac{0.4}{5} \right\}.$$

Intersection

$$\underset{\sim}{A} \cap \underset{\sim}{B} = \left\{ \frac{0}{1} + \frac{0.5}{2} + \frac{0.5}{3} + \frac{0.2}{4} + \frac{0.2}{5} \right\}.$$

Difference

$$\underset{\sim}{A} | \underset{\sim}{B} = \underset{\sim}{A} \cap \overline{\underset{\sim}{B}} = \left\{ \frac{0}{1} + \frac{0.5}{2} + \frac{0.3}{3} + \frac{0.3}{4} + \frac{0.2}{5} \right\}.$$

$$\underset{\sim}{B} | \underset{\sim}{A} = \underset{\sim}{B} \cap \overline{\underset{\sim}{A}} = \left\{ \frac{0}{1} + \frac{0}{2} + \frac{0.5}{3} + \frac{0.2}{4} + \frac{0.4}{5} \right\}.$$

DeMorgan' sprinciples

$$\overline{\underset{\sim}{A} \cup \underset{\sim}{B}} = \overline{\underset{\sim}{A}} \cap \overline{\underset{\sim}{B}} = \left\{ \frac{1}{1} + \frac{0}{2} + \frac{0.3}{3} + \frac{0.7}{4} + \frac{0.6}{5} \right\}.$$

$$\overline{\underset{\sim}{A} \cap \underset{\sim}{B}} = \overline{\underset{\sim}{A}} \cup \overline{\underset{\sim}{B}} = \left\{ \frac{1}{1} + \frac{0.5}{2} + \frac{0.5}{3} + \frac{0.8}{4} + \frac{0.8}{5} \right\}.$$

Example 2.6

In chemical engineering we have a natural gas stream that mixed with an amine solution in an absorber. Suppose the selection of an appropriate analyzer to monitor the "sales gas" sour-gas concentration is important. This selection process can be complicated by the fact that one type of analyzer, say A, does not provide an average suitable pressure range but it does give a borderline value of instrument dead time; in contrast another analyzer, say B, may give a good value of process dead time but a poor pressure range. Suppose for this problem we consider three analyzers: A, B, and C.

Let

$$\underset{\sim}{P} = \left\{ \frac{0.7}{A} + \frac{0.3}{B} + \frac{0.9}{C} \right\}$$

represent the fuzzy set showing the pressure range suitability of analyzers A, B, and C (a membership of 0 is not suitable, a value of 1 is excellent).

Also let

$$\underset{\sim}{\mathrm{OT}} = \left\{ \frac{0.5}{\mathrm{A}} + \frac{0.9}{\mathrm{B}} + \frac{0.4}{\mathrm{C}} \right\}$$

represent the fuzzy set showing the instrument dead time suitability of analyzers A, B, and C (again, 0 is not suitable and 1 is excellent).

$\overline{\underset{\sim}{\mathrm{P}}}$ and $\overline{\underset{\sim}{\mathrm{OT}}}$ will show the analyzers that are not suitable for pressure range and instrument dead time, respectively:

$$\overline{\underset{\sim}{\mathrm{P}}} = \left\{ \frac{0.3}{\mathrm{A}} + \frac{0.7}{\mathrm{B}} + \frac{0.1}{\mathrm{C}} \right\} \text{ and } \overline{\underset{\sim}{\mathrm{OT}}} = \left\{ \frac{0.5}{\mathrm{A}} + \frac{0.1}{\mathrm{B}} + \frac{0.6}{\mathrm{C}} \right\},$$

$$\text{therefore } \overline{P} \cap \overline{\underset{\sim}{\mathrm{OT}}} = \left\{ \frac{0.3}{\mathrm{A}} + \frac{0.1}{\mathrm{B}} + \frac{0.1}{\mathrm{C}} \right\}.$$

$\underset{\sim}{\mathrm{P}} \cup \underset{\sim}{\mathrm{OT}}$ will show which analyzer is most suitable in either category:

$$\underset{\sim}{\mathrm{P}} \cup \underset{\sim}{\mathrm{OT}} = \left\{ \frac{0.7}{\mathrm{A}} + \frac{0.9}{\mathrm{B}} + \frac{0.9}{\mathrm{C}} \right\}.$$

$\underset{\sim}{\mathrm{P}} \cap \underset{\sim}{\mathrm{OT}}$ will show which analyzer is suitable in both categories:

$$\underset{\sim}{\mathrm{P}} \cap \underset{\sim}{\mathrm{OT}} = \left\{ \frac{0.5}{\mathrm{A}} + \frac{0.3}{\mathrm{B}} + \frac{0.4}{\mathrm{C}} \right\}.$$

Example 2.7

One of the crucial manufacturing operations associated with building the external fuel tank for the Space Shuttle involves the spray-on foam insulation (SOFI) process, which combines two critical component chemicals in a spray gun under high pressure and a precise temperature and flow rate. Control of these parameters to *near* set-point values is crucial for satisfying a number of important specification requirements. Specification requirements consist of aerodynamic, mechanical, chemical, and thermodynamic properties.

Fuzzy characterization techniques could be employed to enhance initial screening experiments; for example, to determine the critical values of both flow and temperature. The true levels can be approximated only in the real world. If we target a low flow rate for 48 1b min^{-1}, it may be 38–58 1b min^{-1}. Also, if we target a high temperature for 135° F, it may be 133–137° F.

How the imprecision of the experimental setup influences the variabilities of key process end results could be modeled using fuzzy set methods, for example, high flow with high temperature, low flow with low temperature, and so on. Examples are shown in Figure 2.18, for low flow rate and high temperature.

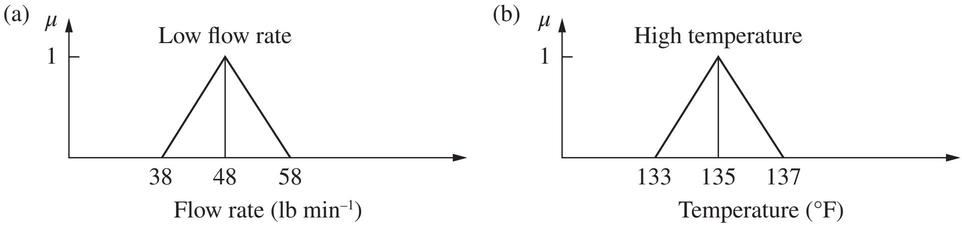

Figure 2.18 Foam insulation membership function for (a) low flow rate and (b) high temperature.

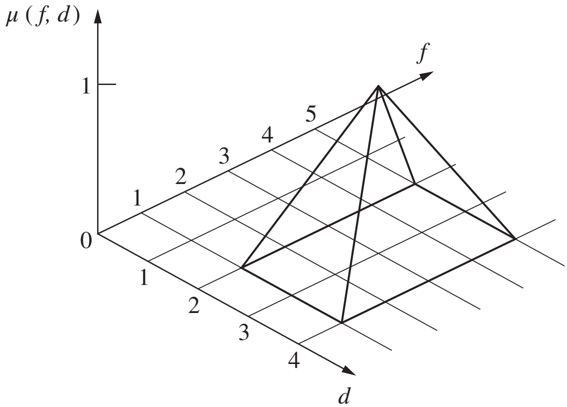

Figure 2.19 Three-dimensional image of the intersection of two fuzzy sets, that is, $\underset{\sim}{F} \cap \underset{\sim}{D}$.

Suppose we have a fuzzy set for flow, normalized on a universe of integers [1, 2, 3, 4, 5], and a fuzzy set for temperature, normalized on a universe of integers [1, 2, 3, 4], as follows:

$$\underset{\sim}{F} = \left\{ \frac{0}{1} + \frac{0.5}{2} + \frac{1}{3} + \frac{0.5}{4} + \frac{0}{5} \right\} \quad \text{and} \quad \underset{\sim}{D} = \left\{ \frac{0}{1} + \frac{0}{2} + \frac{1}{3} + \frac{0}{4} \right\}.$$

Further, suppose that we are interested in how flow and temperature are related in a pairwise sense; we could take the intersection of these two sets. A three-dimensional image should be constructed when we take the union or intersection of sets from two different universes. For example, the intersection of $\underset{\sim}{F}$ and $\underset{\sim}{D}$ is given in Figure 2.19. The idea of combining membership functions from two different universes in an orthogonal form, as indicated in Figure 2.19, is associated with what is termed *noninteractive fuzzy sets*, and this is described in Ross (2004).

Alternative Fuzzy Set Operations

The operations on fuzzy sets listed as Equations (2.22) through (2.24) are called the *standard fuzzy operations*. These operations are the same as those for classical sets, when the range of membership values is restricted to the unit interval. However, these standard fuzzy operations are not the only operations that can be applied to fuzzy sets. For each of the three standard operations, there exists a broad class of functions whose members can be considered fuzzy generalizations of the standard operations. Functions that qualify as fuzzy intersections and fuzzy unions are usually referred to in the literature as *t-norms* and *t-conorms* (or *s-norms*), respectively (e.g., Klir and Yuan, 1995; Klement, Mesiar, and Pap, 2000). These t-norms and t-conorms are so named because they were originally introduced as *triangular norms* and *triangular conorms*, respectively, by Menger (1942) in his study of statistical metric spaces.

The standard fuzzy operations have special significance when compared to all of the other t-norms and t-conorms. The standard fuzzy intersection, min operator, when applied to a fuzzy set produces the largest membership value of all the t-norms, and the standard fuzzy union, max operator, when applied to a fuzzy set produces the smallest membership value of all the t-conorms. These features of the standard fuzzy intersection and union are significant because they both prevent the compounding of errors in the operands (Klir and Yuan, 1995). Most of the alternative norms lack this significance.

Aggregation operations on fuzzy sets are operations by which several fuzzy sets are combined in a desirable way to produce a single fuzzy set. For example, suppose a computer's performance in three test trials is described as excellent, very good, and nominal, and each of these linguistic labels is represented by a fuzzy set on the universe [0, 100]. Then, a useful aggregation operation would produce a meaningful expression, in terms of a single fuzzy set, of the overall performance of the computer. The standard fuzzy intersections and unions qualify as aggregation operations on fuzzy sets and, although they are defined for only two arguments, the fact that they have a property of associativity provides a mechanism for extending their definitions to three or more arguments. Other common aggregation operations, such as *averaging operations* and *ordered weighted averaging operations*, can be found in the literature (see Klir and Yuan, 1995). The averaging operations have their own range that happens to fill the gap between the largest intersection (the min operator) and the smallest union (the max operator). These averaging operations on fuzzy sets have no counterparts in classical set theory and, because of this, extensions of fuzzy sets into fuzzy logic allow for the latter to be much more expressive in natural categories revealed by empirical data or required by intuition (Belohlavek et al., 2002).

Summary

In this chapter we have developed the basic definitions for, properties of, and operations on crisp sets and fuzzy sets. It has been shown that the only basic axioms not common to both crisp and fuzzy sets are the two excluded middle axioms; however, these axioms are not part of the axiomatic structure of fuzzy set theory (Chapter 1). All other operations detailed here are common to both crisp and fuzzy sets; however, other operations such as aggregation and averaging operators that are allowed in fuzzy sets have no counterparts in classical set theory. For many situations in reasoning, the excluded middle axioms present constraints on reasoning (see Chapters 5 and 13). Aside from the difference of set membership being an infinite-valued idea as opposed to a binary-valued quantity, fuzzy sets are handled and treated in the same mathematical form as are crisp sets. The principle of *noninteractivity* between sets was mentioned and is analogous to the assumption of independence in probability modeling. Noninteractive fuzzy sets will become a necessary idea in fuzzy systems simulation when inputs from a variety of universes are aggregated in a collective sense to propagate an output; this propagation process is discussed in more detail in Chapters 5 and 8. Finally, it was pointed out that there are many other operations, called *norms*, that can be used to extend fuzzy intersections, unions, and complements, but such extensions are beyond the scope of this text.

References

Belohlavek, R., Klir, G., Lewis, H., and Way, E. (2002). On the capability of fuzzy set theory to represent concepts. *Int. J. Gen. Syst.*, **31**, 569–585.

Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms* (New York: Plenum Press).

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems, Theory and Applications* (New York: Academic Press).

Gaines, B. (1978). Fuzzy and probability uncertainty logics. *Inf. Control*, **38**, 154–169.

Gupta, M., and Qi, J. (1991). Theory of T-norms and fuzzy inference methods. *Fuzzy Sets Syst.*, **40**, 431–450.

Klement, E., Mesiar, R., and Pap, E. (2000). *Triangular Norms* (Boston: Kluwer Academic).

Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Upper Saddle River, NJ: Prentice Hall).

Menger, K. (1942). Statistical metrics. *Proc. Natl. Acad. Sci. U.S.A.*, **28**, 535–537.

Ross, T. (2004). *Fuzzy Logic with Engineering Applications* (Chichester, UK: John Wiley & Sons, Ltd.).

Zadeh, L. (1965). *Fuzzy sets. Inf. Control*, **8**, 338–353.

Zimmermann, H. (1991) *Fuzzy Set Theory and Its Applications*, 2nd ed. (Dordrecht, Netherlands: Kluwer Academic).

Problems

2.1 The direction of the wind was in a North–South direction from morning until noon. Then it changed direction to East–West. The time noon is considered a transition period for changing the wind direction from North–South to East–West as seen in Figure P2.1. Find the intersection, union, and difference for the two directions of wind flow.

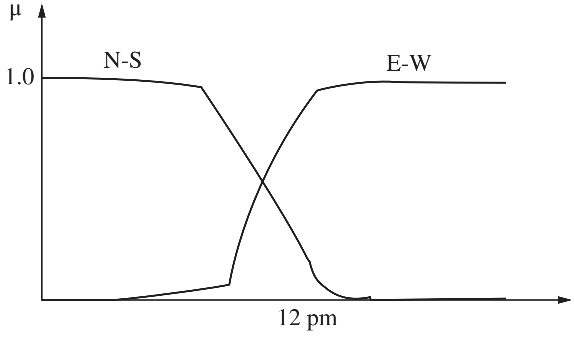

Figure P2.1

2.2 Two types of steel are tested four times each for their tensile strength. Let us consider fuzzy sets '$\underset{\sim}{A}$' and '$\underset{\sim}{B}$' to be the two types of steel on the universe of strengths (no dimensional units) where four tests X = {1, 2, 3, 4} were conducted on each of the two steel types. The following given sets represent the tensile strengths of each steel. Compare the tensile strength of these two steels by finding the following.

$$\underset{\sim}{A} = \left\{ \frac{0.4}{1} + \frac{0.35}{2} + \frac{0.5}{3} + \frac{0.6}{4} \right\} \quad \underset{\sim}{B} = \left\{ \frac{0.7}{1} + \frac{0.75}{2} + \frac{0.65}{3} + \frac{0.8}{4} \right\}$$

a. $\underset{\sim}{A} \cap \underset{\sim}{B}$
b. $\underset{\sim}{A} \cup \underset{\sim}{B}$
c. $\overline{\underset{\sim}{A}}$, $\overline{\underset{\sim}{B}}$
d. $\underset{\sim}{A} - \underset{\sim}{B}$

2.3 In determining corporate profitability, many construction companies must make decisions based on the particular client's spending habits, such as the amount the client spends and their capacity for spending. Many of these attributes are fuzzy. A client which spends a "large amount" is considered to be "profitable" to the construction company. A "large" amount of spending is a fuzzy variable, as is a "profitable" return. These two fuzzy sets should have some overlap, but they should not be defined on an identical range.

$\underset{\sim}{A}$ = {"large" spenders}.
$\underset{\sim}{B}$ = {"profitable" clients}.

For the two fuzzy sets shown in Figure P2.3, find the following properties graphically:

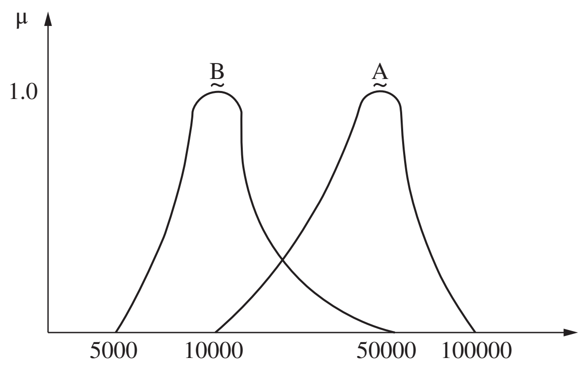

Figure P2.3

a. $\underset{\sim}{A} \cup \underset{\sim}{B}$: all clients deemed profitable or who are large spenders.
b. $\underset{\sim}{A} \cap \underset{\sim}{B}$: all clients deemed profitable and large spenders.
c. $\overline{\underset{\sim}{A}}$ *and* $\overline{\underset{\sim}{B}}$: Those clients (i) deemed not profitable and (ii) deemed not large spenders (separately).
d. $\overline{\underset{\sim}{A}} | \overline{\underset{\sim}{B}}$: Entities deemed profitable clients but not large spenders.
e. $\overline{\underset{\sim}{A} U \underset{\sim}{B}} = \overline{\underset{\sim}{A}} \cap \overline{\underset{\sim}{B}}$ (De Morgan's principle).

2.4 A Motorbike ($\underset{\sim}{M}$), Car ($\underset{\sim}{C}$), and Truck ($\underset{\sim}{T}$) are tested for their speed. Speed is directly influenced by the Horsepower (HP) of the engine. Let 'X' denotes speeds HP, then X = {10, 30, 100, 200, 500}. Their membership function values are given in the table:

Horsepower (HP)	Motorbike ($\underset{\sim}{M}$)	Car ($\underset{\sim}{C}$)	Truck ($\underset{\sim}{T}$)
10	0.2	0	0
30	0.5	0.3	0
100	0.9	0.7	0.2
200	1	0.9	0.4
500	1	1	0.7

Find the following:

a. $\underset{\sim}{M} \cup \underset{\sim}{C}$
b. $\underset{\sim}{C} \cap \underset{\sim}{T}$
c. $\overline{\underset{\sim}{M}}$
d. $\overline{\underset{\sim}{T}}$
e. $\underset{\sim}{M} | \underset{\sim}{T}$
f. $\overline{\underset{\sim}{C} \cap \underset{\sim}{T}}$
g. $\underset{\sim}{M} \cup (\underset{\sim}{C} \cap \underset{\sim}{T})$

2.5 A company is thinking to organize a weeklong event during the month of February in two different cities. The event will take place outside, and the company wants to avoid the rain. The weather department released a prediction of "chances of not having rainfall" for three cities during the four weeks of the month of February. The fuzzy sets given show the assessment of rain in each city for each of the four weeks of February

$$\underset{\sim}{\text{City1}} = \left\{ \frac{0.8}{week1} + \frac{0.5}{week2} + \frac{0.4}{week3} + \frac{0.8}{week4} \right\}$$

$$\underset{\sim}{\text{City2}} = \left\{ \frac{0.7}{week1} + \frac{0.6}{week2} + \frac{0.8}{week3} + \frac{0.9}{week4} \right\}$$

$$\underset{\sim}{\text{City3}} = \left\{ \frac{0.3}{week1} + \frac{0.9}{week2} + \frac{0.75}{week3} + \frac{0.5}{week4} \right\}.$$

The company wants to find the best two cities and the best week to hold the event. What are the cities and the week for the best time to try to hold the event?

2.6 A survey was conducted to study the demand for electricity and water in five parts of a city. It was found that peak demand for both water and electricity was not same in all parts in the city. The membership functions for these two sets is shown in Figure P2.6. For these two fuzzy sets, compute the union, intersection, and difference. The five parts of the city are denoted as X = {1, 2, 3, 4, 5}.

2.7 Consider a local area network (LAN) of interconnected workstations that communicate using Ethernet protocols at a maximum rate of 10 Mbps (megabits per second). Traffic rates on the network can be expressed as the peak value of the total bandwidth (BW) used, and two fuzzy sets, "Quiet" and "Congested," can be used to describe the perceived

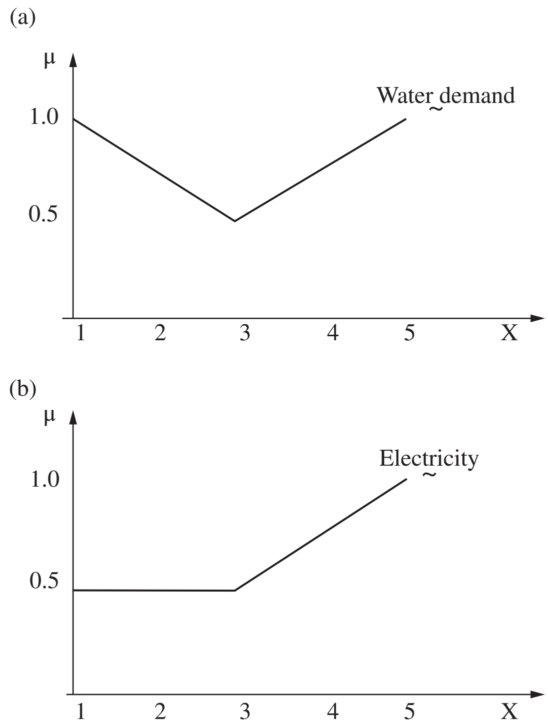

Figure P2.6

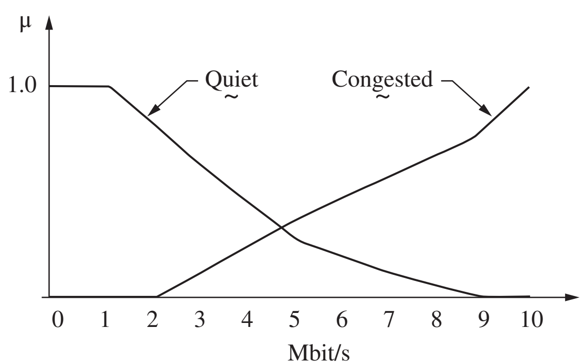

Figure P2.7

data loading on the LAN. If the discrete universal set X = {0, 1, 2, 5, 7, 9, 10} represents BW usage, in Mbps, then the membership functions of the fuzzy sets Quiet ($\underset{\sim}{Q}$) and Congested ($\underset{\sim}{C}$) are shown in Figure P2.7.

For these fuzzy sets graphically find union, intersection, complement of each and difference.

2.8 An engineer is asked to develop a glass-break detector/discriminator for use with residential alarm systems. The detector should be able to distinguish between the breaking of a pane of glass (a window) and a drinking glass. From analysis it has been determined that the sound of a shattering window pane contains most of its energy at frequencies centered about 4 kHz, whereas the sound of a shattering drinking glass contains most of its energy at frequencies centered about 8 kHz. The frequency spectra of the two shattering sounds overlap. The membership functions for the window pane and the glass are given as $\mu_A(x)$ and $\mu_B(x)$, respectively. Illustrate the basic operations of union, intersection, complement and difference for the following membership functions:

$$X = \{0, 1 \ldots 10\}; \sigma = 2; \mu_{\underset{\sim}{A}} = 4; \mu_{\underset{\sim}{B}} = 8.$$

$$\mu_{\underset{\sim}{A}}(\mathrm{x}) = \exp\left[\frac{-(x-\mu_{\underset{\sim}{A}})^2}{2\sigma^2}\right]$$

$$\mu_{\underset{\sim}{B}}(\mathrm{x}) = \exp\left[\frac{-(x-\mu_{\underset{\sim}{B}})^2}{2\sigma^2}\right]$$

3

Classical Relations and Fuzzy Relations

[A]ssonance means getting the rhyme wrong.
Michael Caine as Professor Bryant in the movie *Educating Rita*, 1983

Education is the path from cocky ignorance to miserable uncertainty.
Mark Twain, nineteenth-century writer

This chapter introduces the notion of a relation as the basic idea behind numerous operations on sets such as Cartesian products, composition of relations, and equivalence properties. Like a set, a relation is of fundamental importance in all fields based on engineering, science, and mathematics. It is also associated with graph theory, a subject of wide impact in design and data manipulation. Relations can also be used to represent *similarity*, a notion that is important to many different technologies and, as expressed in the humorous metaphorical quote about *assonance*, a concept that is a key ingredient in our natural language and its many uses, for example, its use in poems. The *American Heritage Dictionary* defines *assonance* as "approximate agreement or partial similarity"; assonance is an example of a prototypical fuzzy concept.

Understanding relations is central to the understanding of a great many areas addressed in this text. As the second quote reveals, there is likely a relationship between not knowing anything and knowing too much that is tied to education (little of it and much of it). Relations are intimately involved in logic, approximate reasoning, rule-based systems, nonlinear simulation, synthetic evaluation, classification, pattern recognition, and control. Relations will be referred to repeatedly in this text in many different applications areas. Relations represent mappings for sets just as mathematical functions do; relations are also useful in representing connectives in logic (Chapter 5).

This chapter begins by describing Cartesian products as a means of producing ordered relationships among sets. Following this is an introduction to classical (crisp) relations, that is, structures that represent the presence or absence of correlation, interaction, or propinquity between the elements of two or more crisp sets; in this case, a set could also be the universe. There are only two degrees of relationship between elements of the sets in a crisp relation: the relationships "completely related" and "not related" in a binary sense. Basic operations, properties, and the cardinality of relations are explained and illustrated. Two composition methods to relate elements of three or more universes are illustrated.

Fuzzy relations are then developed by allowing the relationship between elements of two or more sets to take on an infinite number of degrees of relationship between the extremes of "completely related" and "not related." In this sense, fuzzy relations are to crisp relations as fuzzy sets are to crisp sets; crisp sets and relations are constrained realizations of fuzzy sets and relations. Operations, properties, and cardinality of fuzzy relations are introduced and illustrated, as are Cartesian products and compositions of fuzzy relations. Some engineering examples are given to illustrate various issues associated with relations. The reader can consult the literature for more details on relations (e.g., Gill, 1976; Dubois and Prade, 1980; Kandel, 1985; Klir and Folger, 1988; Zadeh, 1971).

This chapter contains a section on tolerance and equivalence relations—both classical and fuzzy—which is introduced for use in later chapters of the text. Both tolerance and equivalence relations are illustrated with some examples. Finally, the chapter concludes with a section on value assignments, which discusses various methods to develop the elements of relations and a list of additional composition operators. These assignment methods are discussed, and a few examples are given in the area of similarity methods.

Cartesian Product

An ordered sequence of r elements, written in the form $(a_1, a_2, a_3, \ldots, a_r)$, is called an *ordered r-tuple*; an unordered r-tuple is simply a collection of r elements without restrictions on order. In a ubiquitous special case where $r = 2$, the r-tuple is referred to as an ordered *pair*. For crisp sets $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_r$, the set of all r-tuples $(a_1, a_2, a_3, \ldots, a_r)$, where $a_1 \in \mathrm{A}_1$, $a_2 \in \mathrm{A}_2$, and $a_r \in \mathrm{A}_r$, is called the *Cartesian product* of $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_r$, and is denoted by $\mathrm{A}_1 \times \mathrm{A}_2 \times \cdots \times \mathrm{A}_r$. The Cartesian product of two or more sets is *not* the same thing as the arithmetic product of two or more sets. The latter is dealt with in Chapter 12, when the extension principle is introduced.

When all the A_r are identical and equal to A, the Cartesian product $\mathrm{A}_1 \times \mathrm{A}_2 \times \cdots \times \mathrm{A}_r$ can be denoted as A^r.

Example 3.1

The elements in two sets A and B are given as $\mathrm{A} = \{0, 1\}$ and $\mathrm{B} = \{a, b, c\}$. Various Cartesian products of these two sets can be written as shown:

$$\mathrm{A} \times \mathrm{B} = \{(0,a),(0,b),(0,c),(1,a),(1,b),(1,c)\}.$$

$$\mathrm{B} \times \mathrm{A} = \{(a,0),(a,1),(b,0),(b,1),(c,0),(c,1)\}.$$

$$\mathrm{A} \times \mathrm{A} = \mathrm{A}^2 = \{(0,0),(0,1),(1,0),(1,1)\}.$$

$$\mathrm{B} \times \mathrm{B} = \mathrm{B}^2 = \{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}.$$

Crisp Relations

A subset of the Cartesian product $A_1 \times A_2 \times \cdots \times A_r$ is called an *r-ary relation* over $A_1, A_2, \ldots, A_r$. Again, the most common case is for $r = 2$; in this situation, the relation is a subset of the Cartesian product $A_1 \times A_2$ (i.e., a set of pairs, the first coordinate of which is from A_1 and the second from A_2). This subset of the full Cartesian product is called a *binary relation from* A_1 *into* A_2. If three, four, or five sets are involved in a subset of the full Cartesian product, the relations are called *ternary*, *quaternary*, and *quinary*, respectively. In this text, whenever the term *relation* is used without qualification, it is taken to mean a *binary relation*.

The Cartesian product of two universes X and Y is determined as

$$X \times Y = \{(x, y) | x \in X, y \in Y\}, \tag{3.1}$$

which forms an ordered pair of every $x \in X$ with every $y \in Y$, forming *unconstrained* matches between X and Y. That is, every element in universe X is related completely to every element in universe Y. The *strength* of this relationship between ordered pairs of elements in each universe is measured by the characteristic function, denoted χ, where a value of unity is associated with *complete relationship* and a value of zero is associated with *no relationship*, that is,

$$\chi_{X \times Y}(x, y) = \begin{cases} 1, & (x, y) \in X \times Y \\ 0, & (x, y) \notin X \times Y \end{cases}. \tag{3.2}$$

One can think of this strength of relation as a mapping from ordered pairs of the universe, or ordered pairs of sets defined on the universes, to the characteristic function. When the universes, or sets, are finite, the relation can be conveniently represented by a matrix, called a *relation matrix*. An r-ary relation can be represented by an r-dimensional relation matrix. Hence, binary relations can be represented by two-dimensional matrices (which are used throughout this text).

An example of the strength of relation for the unconstrained case is given in the Sagittal diagram shown in Figure 3.1 (a Sagittal diagram is simply a schematic depicting points as elements of universes and lines as relationships between points, or it can be a pictorial of the elements as nodes that are connected by directional lines, as seen in Figure 3.8). Lines in the Sagittal diagram and values of unity in the *relation matrix*

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \end{array}$$

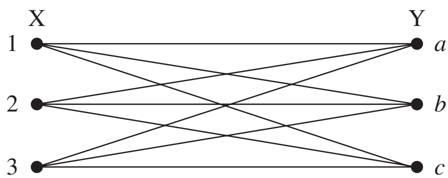

Figure 3.1 Sagittal diagram of an unconstrained relation.

correspond to the ordered pairs of mappings in the relation. Here, the elements in the two universes are defined as X = {1, 2, 3} and Y = {a, b, c}.

A more general crisp relation, R, exists when matches between elements in two universes are *constrained*. Again, the characteristic function is used to assign values of relationship in the mapping of the Cartesian space X × Y to the binary values of (0, 1):

$$\chi_{\mathrm{R}}(x,y) = \begin{cases} 1, & (x,y) \in \mathrm{R} \\ 0, & (x,y) \notin \mathrm{R} \end{cases}. \tag{3.3}$$

Example 3.2

In many biological models, members of certain species can reproduce only with certain members of another species. Hence, only some elements in two or more universes have a relationship (nonzero) in the Cartesian product. An example is shown in Figure 3.2 for 2 two-member species, that is, for X = {1, 2} and for Y = {a, b}. In this case, the locations of zeros in the relation matrix

$$\mathrm{R} = \{(1,a),(2,b)\} \quad \mathrm{R} \subset \mathrm{X} \times \mathrm{Y}; \quad \mathrm{R} = \begin{array}{c c} & \begin{array}{cc} a & b \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}.$$

and the absence of lines in the Sagittal diagram correspond to pairs of elements between the two universes where there is "no relation"; that is, the strength of the relationship is zero.

Special cases of the constrained and the unconstrained Cartesian product for sets where $r = 2$ (i.e., for A^2) are called the *identity relation* and the *universal relation*, respectively. For example, for A = {0, 1, 2} the universal relation, denoted $\mathrm{U_A}$, and the identity relation, denoted $\mathrm{I_A}$, are found to be

$$\mathrm{U_A} = \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}.$$
$$\mathrm{I_A} = \{(0,0),(1,1),(2,2)\}.$$

Example 3.3

Relations can also be defined for continuous universes. Consider, for example, the continuous relation defined by the following expression:

$$\mathrm{R} = \{(x,y) \mid y \geq 2x, x \in \mathrm{X}, y \in \mathrm{Y}\},$$

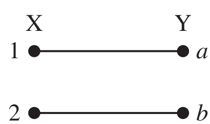

Figure 3.2 Relation matrix and Sagittal diagram for a constrained relation.

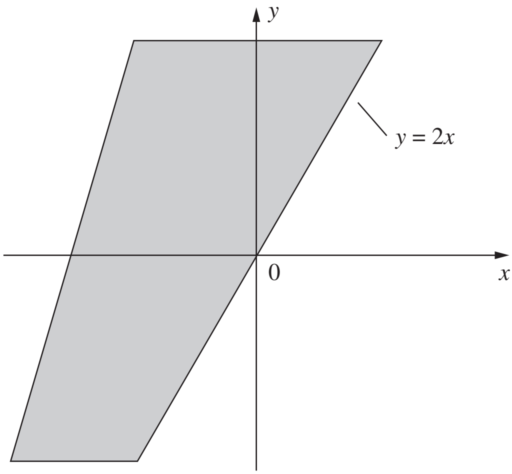

Figure 3.3 Relation corresponding to the expression $y \geq 2x$.

which is also given in function-theoretic form using the characteristic function as

$$\chi_{R}(x, y) = \begin{cases} 1, & y \geq 2x \\ 0, & y < 2x \end{cases}.$$

Graphically, this relation is equivalent to the shaded region shown in Figure 3.3.

Cardinality of Crisp Relations

Suppose n elements of the universe X are related (paired) to m elements of the universe Y. If the cardinality of X is n_X and the cardinality of Y is n_Y, then the cardinality of the relation, R, between these two universes is $n_{X \times Y} = n_X * n_Y$. The cardinality of the power set describing this relation, P(X × Y), is then $n_{p(X \times Y)} = 2^{(nXnY)}$.

Operations on Crisp Relations

Define R and S as two separate relations on the Cartesian universe X × Y, and define the null relation and the complete relation as the relation matrices **O** and **E**, respectively. An example of a 4 × 4 form of the **O** and **E** matrices is given here:

$$\mathbf{O} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The following function-theoretic operations for the two crisp relations (R, S) can now be defined.

Union $$\mathrm{R} \cup \mathrm{S} \rightarrow \chi_{\mathrm{R}\cup\mathrm{S}}(x,y) : \chi_{\mathrm{R}\cup\mathrm{S}}(x,y) = \max[\chi_{\mathrm{R}}(x,y), \chi_{\mathrm{S}}(x,y)]. \tag{3.4}$$

Intersection $$\mathrm{R} \cap \mathrm{S} \rightarrow \chi_{\mathrm{R}\cap\mathrm{S}}(x,y) : \chi_{\mathrm{R}\cap\mathrm{S}}(x,y) = \min[\chi_{\mathrm{R}}(x,y), \chi_{\mathrm{S}}(x,y)] \tag{3.5}$$

Complement $$\overline{\mathrm{R}} \rightarrow \chi_{\overline{\mathrm{R}}}(x,y) : \chi_{\overline{\mathrm{R}}}(x,y) = 1 - \chi_{\mathrm{R}}(x,y). \tag{3.6}$$

Containment $$\mathrm{R} \subset \mathrm{S} \rightarrow \chi_{\mathrm{R}\subset\mathrm{S}}(x,y) : \chi_{\mathrm{R}}(x,y) \le \chi_{\mathrm{S}}(x,y). \tag{3.7}$$

Identity $$\emptyset \rightarrow \mathbf{O} \quad \text{and} \quad \mathrm{X} \rightarrow \mathbf{E}. \tag{3.8}$$

Properties of Crisp Relations

The properties of commutativity, associativity, distributivity, involution, and idempotency all hold for crisp relations just as they do for classical set operations. Moreover, *De Morgan*'s *principles* and the *excluded middle axioms* also hold for crisp (classical) relations just as they do for crisp (classical) sets. The null relation, **O**, and the complete relation, **E**, are analogous to the null set, Ø, and the whole set, X, respectively, in the set-theoretic case (Chapter 2).

Composition

Let R be a relation that relates, or maps, elements from universe X to universe Y, and let S be a relation that relates, or maps, elements from universe Y to universe Z.

A useful question we seek to answer is whether we can find a relation, T, that relates the same elements in universe X that R contains to the same elements in universe Z that S contains. It turns out that we can find such a relation using an operation known as *composition*. From the Sagittal diagram in Figure 3.4, we see that the only "path" between relation

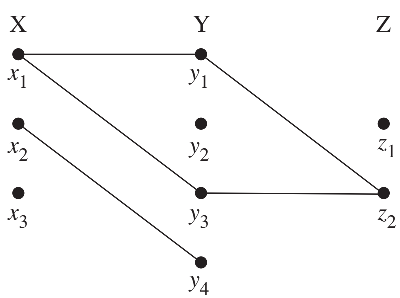

Figure 3.4 Sagittal diagram relating elements of three universes.

R and relation S is the two routes that start at x_1 and end at z_2 (i.e., $x_1 - y_1 - z_2$ and $x_1 - y_3 - z_2$). Hence, we wish to find a relation T that relates the ordered pair (x_1, z_2), that is, $(x_1, z_2) \in \mathrm{T}$. In this example,

$$\mathrm{R} = \{(x_1, y_1), (x_1, y_3), (x_2, y_4)\}.$$
$$\mathrm{S} = \{(y_1, z_2), (y_3, z_2)\}.$$

There are two common forms of the composition operation: one is called the *max–min composition*, and the other, the *max–product composition*. (Five other forms of the composition operator are available for certain logic issues; these are described at the end of this chapter.) The max–min composition is defined by the set-theoretic and membership function-theoretic expressions

$$\begin{aligned} &\mathrm{T} = \mathrm{R} \circ \mathrm{S}, \\ &\chi_{\mathrm{T}}(x,z) = \bigvee_{y \in \mathrm{Y}} (\chi_{\mathrm{R}}(x,y) \wedge \chi_{\mathrm{S}}(y,z)). \end{aligned} \tag{3.9}$$

and the max–product (sometimes called *max–dot*) composition is defined by the set-theoretic and membership function-theoretic expressions

$$\begin{aligned} &\mathrm{T} = \mathrm{R} \circ \mathrm{S}, \\ &\chi_{\mathrm{T}}(x,z) = \bigvee_{y \in \mathrm{Y}} (\chi_{\mathrm{R}}(x,y) \bullet \chi_{\mathrm{S}}(y,z)). \end{aligned} \tag{3.10}$$

Here, the symbol, •, is arithmetic product.

There is an interesting physical analogy for the max–min composition operator. Figure 3.5 illustrates a system comprising several chains placed together in a parallel fashion. In the system, each chain comprises a number of chain links. If we were to take one of the chains out of the system, place it in a tensile test machine, and exert a large tensile force on the chain, we would find that the chain would break at its weakest link.

Hence, the strength of one chain is equal to the strength of its weakest link; in other words, the *minimum* ($\wedge$) strength of all the links in the chain governs the strength of the overall chain. Now, if we were to place the entire chain system in a tensile device and exert a tensile force on the chain system, we would find that the chain system would continue to carry increasing loads until the last chain in the system breaks. That is, weaker chains would break with an increasing load until the strongest chain is left alone, and eventually it would break; in other words, the *maximum* ($\vee$) strength of all the chains in the chain system would govern the overall strength of the chain system. Each chain in the system is analogous to the min operation in the max–min

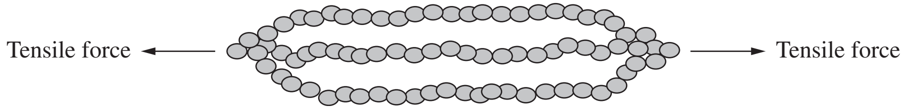

Figure 3.5 Chain strength analogy for max–min composition.

composition, and the overall chain system strength is analogous to the max operation in the max–min composition.

Example 3.4
The matrix expression for the crisp relations shown in Figure 3.4 can be found using the max–min composition operation. Relation matrices for R and S would be expressed as

$$\mathrm{R} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad \text{and} \quad \mathrm{S} = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \begin{array}{c} \begin{array}{cc} z_1 & z_2 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \end{array}.$$

The resulting relation T would then be determined by max–min composition (Equation (3.9)) or max–product composition (Equation (3.10)). (In the crisp case, these forms of the composition operators produce identical results; other forms of this operator, such as those listed at the end of this chapter, will not produce identical results.) For example,

$$\mu_{\mathrm{T}}(x_1, z_1) = \max[\min(1,0), \min(0,0), \min(1,0), \min(0,0)] = 0,$$
$$\mu_{\mathrm{T}}(x_1, z_2) = \max[\min(1,1), \min(0,0), \min(1,1), \min(0,0)] = 1,$$

and for the rest,

$$\mathrm{T} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cc} z_1 & z_2 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \end{array}.$$

Fuzzy Relations

Fuzzy relations also map elements of one universe, say X, to those of another universe, say Y, through the Cartesian product of the two universes. However, the "strength" of the relation between ordered pairs of the two universes is not measured with the characteristic function, but rather with a membership function expressing various "degrees" of strength of the relation on the unit interval [0,1]. Hence, a fuzzy relation $\underset{\sim}{\mathrm{R}}$ is a mapping from the Cartesian space $\mathrm{X} \times \mathrm{Y}$ to the interval [0,1], where the strength of the mapping is expressed by the membership function of the relation for ordered pairs from the two universes, or $\mu_{\underset{\sim}{\mathrm{R}}}(x,y)$.

Cardinality of Fuzzy Relations

Because the cardinality of fuzzy sets on any universe is infinity, the cardinality of a fuzzy relation between two or more universes is also infinity.

Operations on Fuzzy Relations

Let $\underset{\sim}{R}$ and $\underset{\sim}{S}$ be fuzzy relations on the Cartesian space X × Y. Then the following operations apply for the membership values for various set operations (these are similar to the same operations on crisp sets, Equations (3.4) to (3.8):

Union $$\mu_{\underset{\sim}{R}\cup\underset{\sim}{S}}(x,y)=\max(\mu_{\underset{\sim}{R}}(x,y),\mu_{\underset{\sim}{S}}(x,y)). \tag{3.11}$$

Intersection $$\mu_{\underset{\sim}{R}\cap\underset{\sim}{S}}(x,y)=\min(\mu_{\underset{\sim}{R}}(x,y),\mu_{\underset{\sim}{S}}(x,y)). \tag{3.12}$$

Complement $$\mu_{\overline{\underset{\sim}{R}}}(x,y)=1-\mu_{\underset{\sim}{R}}(x,y). \tag{3.13}$$

Containment $$\underset{\sim}{R}\subset\underset{\sim}{S}\Rightarrow\mu_{\underset{\sim}{R}}(x,y)\le\mu_{\underset{\sim}{S}}(x,y). \tag{3.14}$$

Properties of Fuzzy Relations

Just as for crisp relations, the properties of commutativity, associativity, distributivity, involution, and idempotency all hold for fuzzy relations. Moreover, De Morgan's principles hold for fuzzy relations just as they do for crisp (classical) relations, and the null relation, **O**, and the complete relation, **E**, are analogous to the null set and the whole set in set-theoretic form, respectively. Fuzzy relations are not constrained, as is the case for fuzzy sets in general, by the excluded middle axioms. Because a fuzzy relation $\underset{\sim}{R}$ is also a fuzzy set, there is overlap between a relation and its complement; hence,

$$\underset{\sim}{R}\cup\overline{\underset{\sim}{R}}\neq\mathbf{E}.$$

$$\underset{\sim}{R}\cap\overline{\underset{\sim}{R}}\neq\mathbf{O}.$$

As seen in the foregoing expressions, the *excluded middle axioms* for fuzzy relations do not result, in general, in the null relation, **O**, or the complete relation, **E**.

Fuzzy Cartesian Product and Composition

Because fuzzy relations in general are fuzzy sets, we can define the Cartesian product to be a relation between two or more fuzzy sets. Let $\underset{\sim}{A}$ be a fuzzy set on universe X and $\underset{\sim}{B}$ be a fuzzy set on universe Y, then the Cartesian product between fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ will result in a fuzzy relation R, which is contained within the full Cartesian product space, or

$$\underset{\sim}{A}\times\underset{\sim}{B}=\underset{\sim}{R}\subset X\times Y, \tag{3.15}$$

where the fuzzy relation $\underset{\sim}{R}$ has membership function

$$\mu_{\underset{\sim}{R}}(x,y)=\mu_{\underset{\sim}{A}\times\underset{\sim}{B}}(x,y)=\min(\mu_{\underset{\sim}{A}}(x),\mu_{\underset{\sim}{B}}(y)). \tag{3.16}$$

The Cartesian product defined by $A \times B = R$ (Equation (3.15)) is implemented in the same fashion as is the cross product of two vectors. Again, the Cartesian product is *not* the same operation as the arithmetic product. In the case of two-dimensional relations ($r = 2$), the former employs the idea of pairing of elements among sets, whereas the latter uses actual arithmetic products between elements of sets. Each of the fuzzy sets could be thought of as a vector of membership values; each value is associated with a particular element in each set. For example, for a fuzzy set (vector) $\underset{\sim}{A}$ that has four elements, hence column vector of size 4×1, and for a fuzzy set (vector) $\underset{\sim}{B}$ that has five elements, hence a row vector size of 1×5, the resulting fuzzy relation $\underset{\sim}{R}$ will be represented by a matrix of size 4×5, that is, $\underset{\sim}{R}$ will have four rows and five columns. This idea is illustrated in the following example.

Example 3.5

Suppose we have two fuzzy sets, $\underset{\sim}{A}$ defined on a universe of three discrete temperatures, $X = \{x_1, x_2, x_3\}$, and $\underset{\sim}{B}$ defined on a universe of two discrete pressures, $Y = \{y_1, y_2\}$, and we want to find the fuzzy Cartesian product between them. Fuzzy set $\underset{\sim}{A}$ could represent the "ambient" temperature; fuzzy set $\underset{\sim}{B}$ the "near-optimum" pressure for a certain heat exchanger; and the Cartesian product might represent the conditions (temperature–pressure pairs) of the exchanger that are associated with "efficient" operations. For example, let

$$\underset{\sim}{A} = \frac{0.2}{x_1} + \frac{0.5}{x_2} + \frac{1}{x_3} \quad \text{and} \quad \underset{\sim}{B} = \frac{0.3}{y_1} + \frac{0.9}{y_2}.$$

Note that $\underset{\sim}{A}$ can be represented as a column vector of size 3×1 and $\underset{\sim}{B}$ can be represented by a row vector of 1×2. Then the fuzzy Cartesian product, using Equation (3.16), results in a fuzzy relation $\underset{\sim}{R}$ (of size 3×2) representing "efficient" conditions, or

$$\underset{\sim}{A} \times \underset{\sim}{B} = \underset{\sim}{R} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{bmatrix} 0.2 & 0.2 \\ 0.3 & 0.5 \\ 0.3 & 0.9 \end{bmatrix} \end{array}.$$

Fuzzy composition can be defined just as it is for crisp (binary) relations. Suppose $\underset{\sim}{R}$ is a fuzzy relation on the Cartesian space $X \times Y$, $\underset{\sim}{S}$ is a fuzzy relation on $Y \times Z$, and $\underset{\sim}{T}$ is a fuzzy relation on $X \times Z$, then fuzzy max–min composition is defined in terms of the set-theoretic notation and membership function-theoretic notation in the following manner:

$$\underset{\sim}{T} = \underset{\sim}{R} \circ \underset{\sim}{S},$$
$$\mu_{\underset{\sim}{T}}(x,z) = \bigvee_{y \in Y} \left(\mu_{\underset{\sim}{R}}(x,y) \wedge \mu_{\underset{\sim}{S}}(y,z)\right), \tag{3.17a}$$

and fuzzy max–product composition is defined in terms of the membership function-theoretic notation as

$$\mu_{\underset{\sim}{T}}(x,z) = \bigvee_{y \in Y} \left(\mu_{\underset{\sim}{R}}(x,y) \bullet \mu_{\underset{\sim}{S}}(y,z)\right). \tag{3.17b}$$

It should be pointed out that neither crisp nor fuzzy compositions are commutative in general; that is,

$$\underset{\sim}{R} \circ \underset{\sim}{S} \neq \underset{\sim}{S} \circ \underset{\sim}{R}. \tag{3.18}$$

Equation (3.18) is general for any matrix operation, fuzzy or otherwise, that must satisfy consistency between the cardinal counts of elements in respective universes. Even for the case of square matrices, the composition converse, represented by Equation (3.18), is not guaranteed.

Example 3.6

Let us extend the information contained in the Sagittal diagram shown in Figure 3.4 to include fuzzy relationships for X × Y (denoted by the fuzzy relation R) and Y × Z (denoted by the fuzzy relation S). In this case, we change the elements of the universes to

$$X = \{x_1, x_2\}, \ \ Y = \{y_1, y_2\}, \ \text{and} \ \ Z = \{z_1, z_2, z_3\}.$$

Consider the following fuzzy relations:

$$\underset{\sim}{R} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{bmatrix} 0.7 & 0.5 \\ 0.8 & 0.4 \end{bmatrix} \end{array} \quad \text{and} \quad \underset{\sim}{S} = \begin{array}{c} \\ y_1 \\ y_2 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{bmatrix} 0.9 & 0.6 & 0.2 \\ 0.1 & 0.7 & 0.5 \end{bmatrix} \end{array}.$$

Then, the resulting relation, T, which relates elements of universe X to elements of universe Z, that is, defined on Cartesian space X × Z, can be found by max–min composition (Equation (3.17a)) to be, for example,

$$\mu_{\underset{\sim}{T}}(x_1, z_1) = \max[\min(0.7, 0.9), \min(0.5, 0.1)] = 0.7,$$

and the rest

$$\underset{\sim}{T} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} z1 & z2 & z3 \end{array} \\ \begin{bmatrix} 0.7 & 0.6 & 0.5 \\ 0.8 & 0.6 & 0.4 \end{bmatrix} \end{array}.$$

and by max–product composition (Equation (3.17b)) to be, for example,

$$\mu_{\underset{\sim}{T}}(x_2, z_2) = \max[(0.8 \cdot 0.6), (0.4 \cdot 0.7)] = 0.48,$$

and the rest

$$\underset{\sim}{T} = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{bmatrix} 0.63 & 0.42 & 0.25 \\ 0.72 & 0.48 & 0.20 \end{bmatrix} \end{array}.$$

We now illustrate the use of relations with fuzzy sets for three examples from the fields of medicine, electrical engineering, and civil engineering.

Example 3.7

A certain type of virus attacks cells of the human body. The infected cells can be visualized using a special microscope. The microscope generates digital images that medical doctors can analyze and identify the infected cells. The virus causes the infected cells to have a black spot, within a darker gray region (Figure 3.6).

A digital image process can be applied to the image. This processing generates two variables: the first variable, P, is related to black spot quantity (black pixels) and the second variable, S, is related to the shape of the black spot, that is, if they are circular or elliptic. In these images, it is often difficult to actually count the number of black pixels, or to identify a perfect circular cluster of pixels; hence, both these variables must be estimated in a linguistic way.

Suppose that we have two fuzzy sets: $\underset{\sim}{P}$ that represents the number of black pixels (e.g., none with black pixels, C_1, a few with black pixels, C_2, and a lot of black pixels, C_3) and $\underset{\sim}{S}$ that represents the shape of the black pixel clusters (e.g., S_1 is an ellipse and S_2 is a circle). So, we have

$$\underset{\sim}{P} = \left\{ \frac{0.1}{C_1} + \frac{0.5}{C_2} + \frac{1.0}{C_3} \right\} \quad \text{and} \quad \underset{\sim}{S} = \left\{ \frac{0.3}{S_1} + \frac{0.8}{S_2} \right\},$$

and we want to find the relationship between quantity of black pixels in the virus and the shape of the black pixel clusters. Using a Cartesian product between $\underset{\sim}{P}$ and $\underset{\sim}{S}$ gives

$$\underset{\sim}{R} = \underset{\sim}{P} \times \underset{\sim}{S} = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \end{array} \begin{array}{c} \begin{array}{cc} S_1 & S_2 \end{array} \\ \begin{bmatrix} 0.1 & 0.1 \\ 0.3 & 0.5 \\ 0.3 & 0.8 \end{bmatrix} \end{array}.$$

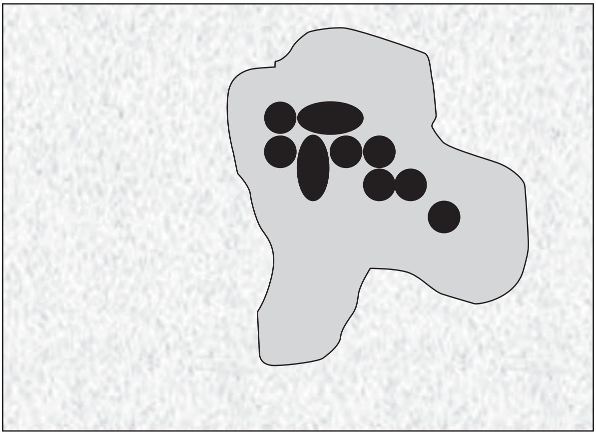

Figure 3.6 An infected cell shows black spots with different shapes in a micrograph.

Now, suppose another microscope image is taken and the number of black pixels is slightly different; let the new black pixel quantity be represented by a fuzzy set, $\underset{\sim}{P}'$:

$$\underset{\sim}{P}' = \left\{ \frac{0.4}{C_1} + \frac{0.7}{C_2} + \frac{0.6}{C_3} \right\}.$$

Using max–min composition with the relation $\underset{\sim}{R}$ will yield a new value for the fuzzy set of pixel cluster shapes that are associated with the new black pixel quantity:

$$\underset{\sim}{S}' = \underset{\sim}{P}' \circ \underset{\sim}{R} = [0.4 \;\; 0.7 \;\; 0.6] \circ \begin{bmatrix} 0.1 & 0.1 \\ 0.3 & 0.5 \\ 0.3 & 0.8 \end{bmatrix} = [0.3 \;\; 0.6].$$

Example 3.8

Suppose we are interested in understanding the speed control of the direct current (DC) shunt motor under no-load condition, as shown diagrammatically in Figure 3.7. Initially, the series resistance R_{se} in Figure 3.7 should be kept in the cut-in position for the following reasons:

1. The back electromagnetic force, given by $E_b = kN\phi$—where k is a constant of proportionality, N is the motor speed, and ϕ is the flux (which is proportional to input voltage, V)—is equal to zero because the motor speed is equal to zero initially.
2. We have $V = E_b + I_a(R_a + R_{se})$, therefore $I_a = (V - E_b)/(R_a + R_{se})$, where I_a is the armature current and R_a is the armature resistance. Because E_b is equal to zero initially, the armature current will be $I_a = V/(R_a + R_{se})$, which is going to be quite large initially and may destroy the armature.

On the basis of both cases 1 and 2, keeping the series resistance R_{se} in the cut-in position will restrict the speed to a low value. Hence, if the rated no-load speed of the motor is 1500 rpm, then the resistance in series with the armature, or the shunt resistance R_{sh}, has to be varied.

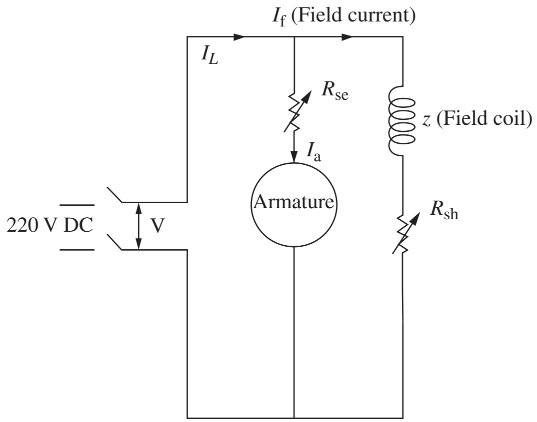

Figure 3.7 A DC shunt motor system.

Two methods provide this type of control: armature control and field control. For example, in armature control, suppose that ϕ (flux) is maintained at some constant value, then motor speed N is proportional to E_b.

If R_{se} is decreased step by step from its high value, I_a (armature current) increases. Hence, this method increases I_a. On the other hand, as I_a is increased the motor speed N increases. These two possible approaches to control could have been done manually or automatically. Either way, however, results in at least two problems, presuming we do not want to change the design of the armature:

What should be the minimum and maximum level of R_{se}?
What should be the minimum and maximum value of I_a?

Now let us suppose that load on the motor is taken into consideration. Then the problem of control becomes twofold. First, owing to fluctuations in the load, the armature current may change, resulting in change in the motor speed. Second, as a result of changes in speed, the armature resistance control must be accomplished to maintain the motor's rated speed. Such control issues become important in applications involving electric trains and a large number of consumer appliances making use of small batteries to run their motors.

We wish to use concepts of fuzzy sets to address this problem. Let $\underset{\sim}{R}_{se}$ be a fuzzy set representing a number of possible values for series resistance, say s_n values, given as

$$\underset{\sim}{R}_{se} = \{R_{s_1}, R_{s_2}, R_{s_3}, \ldots, R_{s_n}\},$$

and let $\underset{\sim}{I}_a$ be a fuzzy set having a number of possible values of the armature current, say m values, given as

$$\underset{\sim}{I}_a = \{I_1, I_2, I_3, \ldots, I_m\}.$$

The fuzzy sets $\underset{\sim}{R}_{se}$ and $\underset{\sim}{I}_a$ can be related through a fuzzy relation, say $\underset{\sim}{R}$, which would allow for the establishment of various degrees of relationship between pairs of resistance and current. In this way, the resistance–current pairings could conform to the modeler's intuition about the trade-offs involved in control of the armature.

Let $\underset{\sim}{N}$ be another fuzzy set having numerous values for the motor speed, say v values, given as

$$\underset{\sim}{N} = \{N_1, N_2, N_3, \ldots, N_v\}.$$

Now, we can determine another fuzzy relation, say $\underset{\sim}{S}$, to relate current to motor speed, that is, $\underset{\sim}{I}_a$ to $\underset{\sim}{N}$.

Using the operation of composition, we could then compute a relation, say $\underset{\sim}{T}$, to be used to relate series resistance to motor speed, that is, $\underset{\sim}{R}_{se}$ to $\underset{\sim}{N}$. The operations needed to develop these relations are as follows—two fuzzy Cartesian products and one composition:

$$\underset{\sim}{R} = \underset{\sim}{R}_{se} \times \underset{\sim}{I}_a,$$

$$\underset{\sim}{S} = \underset{\sim}{I}_a \times \mathrm{N},$$

$$\underset{\sim}{T} = \underset{\sim}{R} \circ \underset{\sim}{S}.$$

Suppose the membership functions for both series resistance $\underset{\sim}{R}_{se}$ and armature current $\underset{\sim}{I}_a$ are given in terms of percentages *of their respective rated values*, that is,

$$\mu_{R_{se}}(\%se) = \frac{0.3}{30} + \frac{0.7}{60} + \frac{1.0}{100} + \frac{0.2}{120}$$

and

$$\mu_{I_a}(\%a) = \frac{0.2}{20} + \frac{0.4}{40} + \frac{0.6}{60} + \frac{0.8}{80} + \frac{1.0}{100} + \frac{0.1}{120},$$

and the membership value for $\underset{\sim}{N}$ is given in units of motor speed in rpm,

$$\mu_N(\text{rpm}) = \frac{0.33}{500} + \frac{0.67}{1000} + \frac{1.0}{1500} + \frac{0.15}{1800}.$$

The following relations then result from use of the Cartesian product to determine $\underset{\sim}{R}$ and $\underset{\sim}{S}$:

$$\underset{\sim}{R} = \begin{array}{c} \\ 30 \\ 60 \\ 100 \\ 120 \end{array} \begin{array}{c} \begin{array}{cccccc} 20 & 40 & 60 & 80 & 100 & 120 \end{array} \\ \begin{bmatrix} 0.2 & 0.3 & 0.3 & 0.3 & 0.3 & 0.1 \\ 0.2 & 0.4 & 0.6 & 0.7 & 0.7 & 0.1 \\ 0.2 & 0.4 & 0.6 & 0.8 & 1 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

and

$$\underset{\sim}{S} = \begin{array}{c} \\ 20 \\ 40 \\ 60 \\ 80 \\ 100 \\ 120 \end{array} \begin{array}{c} \begin{array}{cccc} 500 & 1000 & 1500 & 1800 \end{array} \\ \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.15 \\ 0.33 & 0.4 & 0.4 & 0.15 \\ 0.33 & 0.6 & 0.6 & 0.15 \\ 0.33 & 0.67 & 0.8 & 0.15 \\ 0.33 & 0.67 & 1 & 0.15 \\ 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix} \end{array}.$$

For example, $\mu_{\underset{\sim}{R}}(60, 40) = \min(0.7, 0.4) = 0.4$, $\mu_{\underset{\sim}{R}}(100, 80) = \min(1.0, 0.8) = 0.8$, and $\mu_{\underset{\sim}{S}}(80, 1000) = \min(0.8, 0.67) = 0.67$.

The following relation results from a max–min composition for $\underset{\sim}{T}$:

$$\underset{\sim}{T} = \underset{\sim}{R} \circ \underset{\sim}{S} = \begin{array}{c} \\ 30 \\ 60 \\ 100 \\ 120 \end{array} \begin{array}{c} \begin{array}{cccc} 500 & 1000 & 1500 & 1800 \end{array} \\ \begin{bmatrix} 0.3 & 0.3 & 0.3 & 0.15 \\ 0.33 & 0.67 & 0.7 & 0.15 \\ 0.33 & 0.67 & 1 & 0.15 \\ 0.2 & 0.2 & 0.2 & 0.15 \end{bmatrix} \end{array}.$$

For instance,

$$\mu_{\underset{\sim}{T}}(60,1500) = \max[\min(0.2, 0.2), \min(0.4, 0.4), \min(0.6, 0.6),$$
$$\min(0.7, 0.8), \min(0.7, 1.0), \min(0.1, 0.1)].$$
$$= \max[0.2, 0.4, 0.6, 0.7, 0.7, 0.1] = 0.7.$$

Example 3.9

In the city of Calgary, Alberta, there are a significant number of neighborhood ponds that store overland flow from rainstorms and release the water downstream at a controlled rate to reduce or eliminate flooding in downstream areas. To illustrate a relation using the Cartesian product, let us compare the level in the neighborhood pond system based on a 1-in-100-year storm volume capacity with the closest three rain gauge stations that measure total rainfall.

Let $\underset{\sim}{A}$ = pond system relative depths based on 1-in-100-year capacity (assume the capacities of four ponds are p_1, p^2, p_3, and p_4, and all combine to form one outfall to the trunk sewer). Let $\underset{\sim}{B}$ = total rainfall for event based on 1-in-100-year values from three different rain gauge stations, g_1, g_2, and g_3. Suppose we have the following specific fuzzy sets:

$$\underset{\sim}{A} = \frac{0.2}{p_1} + \frac{0.6}{p_2} + \frac{0.5}{p_3} + \frac{0.9}{p_4}.$$

$$\underset{\sim}{B} = \frac{0.4}{g_1} + \frac{0.7}{g_2} + \frac{0.8}{g_3}.$$

The Cartesian product of these two fuzzy sets could then be formed:

$$\underset{\sim}{A} \times \underset{\sim}{B} = \underset{\sim}{C} = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \begin{array}{c} \begin{array}{ccc} g_1 & g_2 & g_3 \end{array} \\ \begin{bmatrix} 0.2 & 0.2 & 0.2 \\ 0.4 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.5 \\ 0.4 & 0.7 & 0.8 \end{bmatrix} \end{array}.$$

The meaning of this Cartesian product would be to relate the rain gauge's prediction of large storms to the actual pond performance during rain events. Higher values indicate designs and station information that could model and control flooding in a reasonable way. Lower relative values may indicate a design problem or a nonrepresentative gauge location.

To illustrate composition for the same situation, let us try to determine if the rainstorms are widespread or localized. Let us compare the results from a pond system well removed from the previous location during the same storm.

Suppose we have a relationship between the capacity of five more ponds within a new pond system ($p_5, \ldots, p_9$) and the rainfall data from the original rainfall gauges (g_1, g_2, and g_3). This relation is given as

$$\underset{\sim}{D} = \begin{array}{c} \\ g_1 \\ g_2 \\ g_3 \end{array} \begin{array}{c} \begin{array}{ccccc} p_5 & p_6 & p_7 & p_8 & p_9 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.5 & 0.2 & 0.1 \\ 0.4 & 0.7 & 0.5 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.8 & 0.9 & 0.8 \end{bmatrix} \end{array}.$$

Let $\underset{\sim}{E}$ be a fuzzy max–min composition for the two ponding systems:

$$\underset{\sim}{E} = \underset{\sim}{C} \circ \underset{\sim}{D} = \begin{array}{c} \\ p_1 \\ p_2 \\ p_3 \\ p_4 \end{array} \begin{array}{c} \begin{array}{ccccc} p_5 & p_6 & p_7 & p_8 & p_9 \end{array} \\ \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.6 & 0.6 & 0.6 & 0.6 \\ 0.4 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.4 & 0.7 & 0.8 & 0.8 & 0.8 \end{bmatrix} \end{array}.$$

For example,

$$\mu_{\underset{\sim}{E}}(p_2, p_7) = \max[\min(0.4, 0.5), \min(0.6, 0.5), \min(0.6, 0.8)] = 0.6.$$

This new relation, $\underset{\sim}{E}$, actually represents the character of the rainstorm for the two geographically separated pond systems: the first system from the four ponds $p_1,\ldots, p_4$ and the second system from the ponds $p_5,\ldots, p_9$. If the numbers in this relation are large, it means that the rainstorm is widespread, whereas if the numbers are closer to zero, then the rainstorm is more localized and the original rain gauges are not a good predictor for both systems.

Tolerance and Equivalence Relations

Relations can exhibit various useful properties, a few of which are discussed here. As mentioned in the introduction, relations can be used in graph theory (Gill, 1976; Zadeh, 1971). Consider the simple graphs in Figure 3.8. This figure describes a universe of three elements, which are labeled as the vertices of this graph, 1, 2, and 3, or in set notation, $X = \{1, 2, 3\}$. The useful properties we wish to discuss are reflexivity, symmetry, and transitivity (there are other properties of relations that are the antonyms of these three, i.e., irreflexivity, asymmetry, and nontransitivity; these, and an additional property of asymmetry, are not discussed in this text). When a relation is reflexive every vertex in the graph originates a single loop, as shown in Figure 3.8a. If a relation is symmetric, then in the graph for every edge pointing (the arrows on the edge lines in Figure 3.8b) from vertex i to vertex j ($i, j = 1, 2, 3$), there is an edge pointing in the opposite direction, that is, from vertex j to vertex i. When a relation is transitive, then for every pair of edges in the graph, one pointing from vertex i to vertex j and the other from vertex j to vertex k ($i, j, k = 1, 2, 3$), there is an edge pointing from vertex i directly to vertex k, as seen in Figure 3.8c (e.g., an arrow from vertex 1 to vertex 2, an arrow from vertex 2 to vertex 3, and an arrow from vertex 1 to vertex 3).

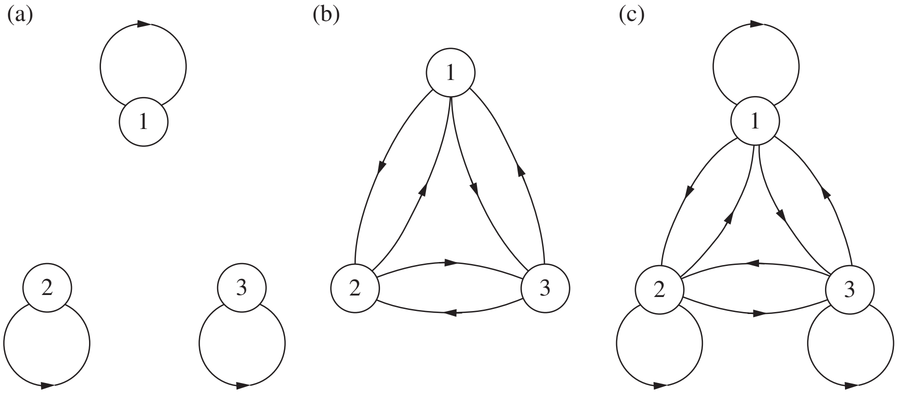

Figure 3.8 Three-vertex graphs for properties of (a) reflexivity, (b) symmetry, and (c) transitivity. Adapted from Gill, 1976.

Crisp Equivalence Relation

A relation R on a universe X can also be thought of as a relation from X to X. The relation R is an equivalence relation if it has the following three properties: (1) reflexivity, (2) symmetry, and (3) transitivity. For example, for a matrix relation, the following properties will hold:

Reflexivity
$$(x_i, x_i) \in \mathrm{R} \text{ or } \chi_{\mathrm{R}}(x_i, x_i) = 1. \tag{3.19a}$$

Symmetry
$$\begin{aligned}&(x_i, x_j) \in \mathrm{R} \rightarrow (x_j, x_i) \in \mathrm{R}\\ &\quad\text{or } \chi_{\mathrm{R}}(x_i, x_j) = \chi_{\mathrm{R}}(x_j, x_i).\end{aligned} \tag{3.19b}$$

Transitivity
$$\begin{aligned}&(x_i, x_j) \in \mathrm{R} \text{ and } (x_j, x_k) \in \mathrm{R} \rightarrow (x_i, x_k) \in \mathrm{R}\\ &\text{or } \chi_{\mathrm{R}}(x_i, x_j) \text{ and } \chi_{\mathrm{R}}(x_j, x_k) = 1 \rightarrow \chi_{\mathrm{R}}(x_i, x_k) = 1.\end{aligned} \tag{3.19c}$$

The most familiar equivalence relation is that of equality among elements of a set. Other examples of equivalence relations include the relation of parallelism among lines in plane geometry, the relation of similarity among triangles, the relation "works in the same building as" among workers of a given city, and others.

Crisp Tolerance Relation

A tolerance relation R (also called a *proximity* relation) on a universe X is a relation that exhibits only the properties of reflexivity and symmetry. A tolerance relation, R, can be reformed into an equivalence relation by at most $(n-1)$ compositions with itself, where n is the cardinal number of the set defining R, in this case X, that is,

$$\mathrm{R}_1^{n-1} = \mathrm{R}_1 \circ \mathrm{R}_1 \circ \cdots \circ \mathrm{R}_1 = \mathrm{R}. \tag{3.20}$$

Example 3.10

Suppose in an airline transportation system we have a universe composed of five elements: the cities Omaha, Chicago, Rome, London, and Detroit. The airline is studying locations of potential hubs in various countries and must consider air mileage between cities and takeoff and landing policies in the various countries. These cities can be enumerated as the elements of a set, that is,

$$\mathrm{X} = \{x_1, x_2, x_3, x_4, x_5\} = \{\text{Omaha, Chicago, Rome, London, Detroit}\}.$$

Further, suppose we have a tolerance relation, R_1, that expresses relationships among these cities:

$$\mathrm{R}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

This relation is reflexive and symmetric. The graph for this tolerance relation would involve five vertices (five elements in the relation), as shown in Figure 3.9. The property of reflexivity (diagonal elements equal unity) simply indicates that a city is totally related to itself. The property of symmetry might represent proximity: Omaha and Chicago (x_1 and x_2) are close (in a binary sense) geographically, and Chicago and Detroit (x_2 and x_5) are close geographically. This relation, R_1, does not have properties of transitivity, for example,

$$(x_1, x_2) \in \mathrm{R}_1 \;\; (x_2, x_5) \in \mathrm{R}_1 \;\text{ but }\; (x_1, x_5) \notin \mathrm{R}_1.$$

R_1 can become an equivalence relation through one ($1 \le n$, where $n = 5$) composition. Using Equation (3.20), we get

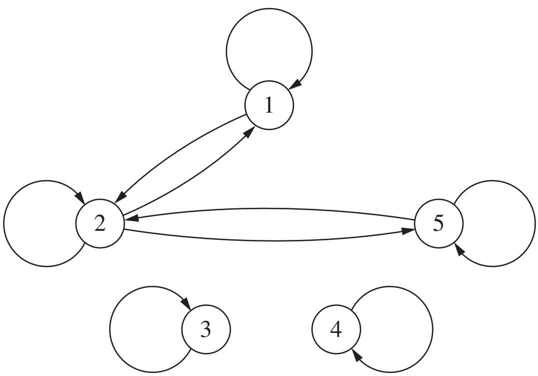

Figure 3.9 Five-vertex graph of tolerance relation (reflexive and symmetric) in Example 3.10.

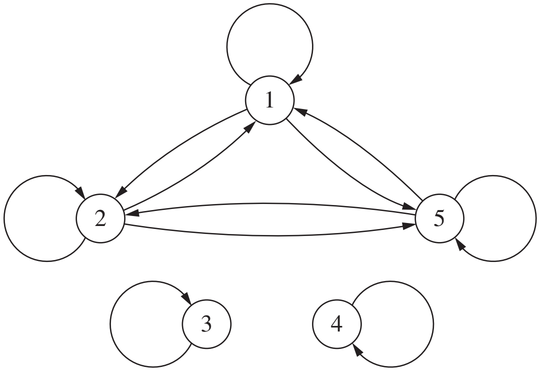

Figure 3.10 Five-vertex graph of equivalence relation (reflexive, symmetric, and transitive) in Example 3.10.

$$\mathrm{R}_1 \circ \mathrm{R}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} = \mathrm{R}.$$

Now, we see in this matrix that transitivity holds, that is, $(x_1, x_5) \in \mathrm{R}_1$, and R is an equivalence relation. Although the point is not important here, we will see in Chapter 10 that equivalence relations also have certain special properties useful in classification. For instance, in this example, the equivalence relation expressed in the foregoing R matrix could represent cities in separate countries. Inspection of the matrix shows that the first, second, and fifth columns are identical, that is, Omaha, Chicago, and Detroit are in the same class; and columns the third and fourth are unique, indicating that Rome and London are cities each in their own class; these three different classes could represent distinct countries. The graph for this equivalence relation would involve five vertices (five elements in the relation) as shown in Figure 3.10.

Fuzzy Tolerance and Equivalence Relations

A fuzzy relation, $\underset{\sim}{\mathrm{R}}$, on a single universe X is also a relation from X to X. It is a fuzzy equivalence relation if all three of the following properties for matrix relations define it:

Reflexivity $$\mu_{\underset{\sim}{\mathrm{R}}}(x_i, x_i) = 1. \tag{3.21a}$$

Symmetry $$\mu_{\underset{\sim}{\mathrm{R}}}\left(x_i, x_j\right) = \mu_{\underset{\sim}{\mathrm{R}}}\left(x_j, x_i\right). \tag{3.21b}$$

Transitivity $$\mu_{\underset{\sim}{\mathrm{R}}}\left(x_i, x_j\right) = \lambda_1 \text{ and } \mu_{\underset{\sim}{\mathrm{R}}}\left(x_j, x_k\right) = \lambda_2 \rightarrow \mu_{\underset{\sim}{\mathrm{R}}}\left(x_j, x_k\right) = \lambda, \quad \text{where } \lambda \geq \min[\lambda_1, \lambda_2]. \tag{3.21c}$$

Looking at the physical analog (see Figure 3.5) of a composition operation, we see it comprises a parallel system of chains, where each chain represents a particular path through the chain system. The physical analogy behind transitivity is that the shorter the chain, the stronger the relation (the stronger is the chain system). In particular, the strength of the link between two elements must be greater than or equal to the strength of any indirect chain involving other elements, that is, Equation (3.21c) (Dubois and Prade, 1980).

It can be shown that any fuzzy tolerance relation, $\underset{\sim}{R}_1$, that has properties of reflexivity and symmetry can be reformed into a fuzzy equivalence relation by at most $(n - 1)$ compositions, just as a crisp tolerance relation can be reformed into a crisp equivalence relation. That is,

$$\underset{\sim}{R}_1^{n-1} = \underset{\sim}{R}_1 \circ \underset{\sim}{R}_1 \circ \cdots \circ \underset{\sim}{R}_1 = \underset{\sim}{R}. \tag{3.22}$$

Example 3.11

Suppose, in a biotechnology experiment, five potentially new strains of bacteria have been detected in the area around an anaerobic corrosion pit on a new aluminum–lithium alloy used in the fuel tanks of a new experimental aircraft. To propose methods to eliminate the biocorrosion caused by these bacteria, the five strains must first be categorized. One way to categorize them is to compare them to one another. In a pairwise comparison, the following "similarity" relation, $\underset{\sim}{R}_1$, is developed. For example, the first strain (column 1) has a strength of similarity to the second strain of 0.8, to the third strain a strength of 0 (i.e., no relation), to the fourth strain a strength of 0.1, and so on. Because the relation is for pairwise similarity it will be reflexive and symmetric. Hence,

$$\underset{\sim}{R}_1 = \begin{bmatrix} 1 & 0.8 & 0 & 0.1 & 0.2 \\ 0.8 & 1 & 0.4 & 0 & 0.9 \\ 0 & 0.4 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0.5 \\ 0.2 & 0.9 & 0 & 0.5 & 1 \end{bmatrix}$$

is reflexive and symmetric. However, it is not transitive, for example,

$$\mu_{\underset{\sim}{R}}(x_1, x_2) = 0.8, \ \ \mu_{\underset{\sim}{R}}(x_2, x_5) = 0.9 \geq 0.8,$$

but

$$\mu_{\underset{\sim}{R}}(x_1, x_5) = 0.2 \leq \min(0.8, 0.9).$$

One composition results in the following relation:

$$\underset{\sim}{R}_1^2 = \underset{\sim}{R}_1 \circ \underset{\sim}{R}_1 = \begin{bmatrix} 1 & 0.8 & 0.4 & 0.2 & 0.8 \\ 0.8 & 1 & 0.4 & 0.5 & 0.9 \\ 0.4 & 0.4 & 1 & 0 & 0.4 \\ 0.2 & 0.5 & 0 & 1 & 0.5 \\ 0.8 & 0.9 & 0.4 & 0.5 & 1 \end{bmatrix},$$

where transitivity still does not result; for example,

$$\mu^2_{\underset{\sim}{R}}(x_1, x_2) = 0.8 \geq 0.5 \quad \text{and} \quad \mu^2_{\underset{\sim}{R}}(x_2, x_4) = 0.5,$$

but

$$\mu^2_{\underset{\sim}{R}}(x_1, x_4) = 0.2 \leq \min(0.8, 0.5).$$

Finally, after one or two more compositions, transitivity results:

$$\underset{\sim}{R}_1^3 = \underset{\sim}{R}_1^4 = \underset{\sim}{R} = \begin{bmatrix} 1 & 0.8 & 0.4 & 0.5 & 0.8 \\ 0.8 & 1 & 0.4 & 0.5 & 0.9 \\ 0.4 & 0.4 & 1 & 0.4 & 0.4 \\ 0.5 & 0.5 & 0.4 & 1 & 0.5 \\ 0.8 & 0.9 & 0.4 & 0.5 & 1 \end{bmatrix}.$$

$$\underset{\sim}{R}_1^3(x_1, x_2) = 0.8 \geq 0.5.$$
$$\underset{\sim}{R}_1^3(x_2, x_4) = 0.5 \geq 0.5.$$
$$\underset{\sim}{R}_1^3(x_1, x_4) = 0.5 \geq 0.5.$$

Graphs can be drawn for fuzzy equivalence relations, but the arrows in the graphs between vertices will have various "strengths," that is, values on the interval [0, 1]. Once the fuzzy relation $\underset{\sim}{R}$ in Example 3.11 is an equivalence relation, it can be used in categorizing the various bacteria according to preestablished levels of confidence. These levels of confidence are illustrated with a method called *alpha-cuts* in Chapter 4, and the categorization idea is illustrated using classification in Chapter 10.

Value Assignments

An appropriate question regarding relations is as follows: Where do the membership values that are contained in a relation come from? The answer to this question is that there are at least seven different ways to develop the numerical values that characterize a relation:

1. Cartesian product
2. Closed-form expression
3. Lookup table
4. Linguistic rules of knowledge
5. Classification
6. Automated methods from input/output data
7. Similarity methods in data manipulation.

The first way is the one that has been illustrated so far in this chapter, to calculate relations from the Cartesian product of two or more fuzzy sets. A second way is through simple observation of a physical process. For a given set of inputs, we observe a process yielding a set of outputs. If there is no variation between specific input–output pairs, we may be led to model the process with a crisp relation. Moreover, if no variability exists, one might be able to express the relation as a closed-form algorithm of the form $\mathbf{Y} = f(\mathbf{X})$, where $\mathbf{X}$ is a vector of inputs and $\mathbf{Y}$ is a vector of outputs. If some variability exists, membership values on the interval [0, 1] may lead us to develop a fuzzy relation from a third approach, the use of a lookup table. Fuzzy relations can also be assembled from linguistic knowledge, expressed as if–then rules. Such knowledge may come from experts, from polls, or from consensus building. This fourth method is illustrated in more detail in Chapters 5 and 8. Relations also arise from notions of classification where issues associated with similarity are central to determining relationships among patterns or clusters of data. The ability to develop relations in classification, the fifth method, is developed in more detail in Chapter 10. The sixth method involves the development of membership functions from procedures used on input and output data, which could be observed and measured from some complex process; this method is the subject of Chapter 7.

One of the most prevalent forms of determining the values in relations, and which is simpler than the sixth method, is through manipulations of data, the seventh method. The more robust a data set, the more accurate the relational entities are in establishing relationships among elements of two or more data sets. This seventh way for determining value assignments for relations is actually a family of procedures termed *similarity methods* (see Zadeh, 1971; Dubois and Prade, 1980). All these methods attempt to determine some sort of similar pattern or structure in data through various metrics. There are many of these methods available, but the two most prevalent are discussed here.

Cosine Amplitude

A useful method is the cosine amplitude method. As with all the following methods, this similarity metric makes use of a collection of data samples, n data samples in particular. If these data samples are collected they form a data array, $\mathbf{X}$,

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}.$$

Each of the elements, $\mathbf{x}_i$, in the data array $\mathbf{X}$ is itself a vector of length m, that is,

$$\mathbf{x}_i = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_m}\}.$$

Hence, each of the data samples can be thought of as a point in m-dimensional space, where each point needs m coordinates for a complete description. Each element of a relation, r_{ij}, results from a pairwise comparison of two data samples, say $\mathbf{x}_i$ and $\mathbf{x}_j$, where the strength of the relationship between data sample $\mathbf{x}_i$ and data sample $\mathbf{x}_j$ is given by the membership value expressing that strength, that is, $r_{ij} = \mu_R(x_i, y_j)$. The relation matrix will be of size $n \times n$ and, as will be the case for all similarity relations, the matrix will be reflexive and symmetric, hence a tolerance relation. The cosine amplitude method calculates r_{ij} in the following manner, and guarantees, as do all the similarity methods, that $0 \le r_{ij} \le 1$:

$$r_{ij} = \frac{\left|\sum_{k=1}^{m} x_{ik}x_{jk}\right|}{\sqrt{\left(\sum_{k=1}^{m} x_{ik}^2\right)\left(\sum_{k=1}^{m} x_{jk}^2\right)}}, \text{ where } i, j = 1, 2, \ldots, n. \tag{3.23}$$

Close inspection of Equation (3.23) reveals that this method is related to the dot product for the cosine function. When two vectors are colinear (most similar), their dot product is unity; when the two vectors are at right angles to one another (most dissimilar), their dot product is zero.

Example 3.12 (Ross, 1995).
Five separate regions along the San Andreas fault in California have suffered damage from a recent earthquake. For purposes of assessing payouts from insurance companies to building owners, the five regions must be classified as to their damage levels. Expression of the damage in terms of relations will prove helpful.

Surveys are conducted of the buildings in each region. All the buildings in each region are described as being in one of three damage states: no damage, medium damage, and serious damage. Each region has each of these three damage states expressed as a percentage (ratio) of the total number of buildings. Hence, for this problem $n = 5$ and $m = 3$. The following table summarizes the findings of the survey team:

Regions	x_1	x_2	x_3	x_4	x_5
x_{i1}, Ratio with no damage	0.3	0.2	0.1	0.7	0.4
x_{i2}, Ratio with medium damage	0.6	0.4	0.6	0.2	0.6
x_{i3}, Ratio with serious damage	0.1	0.4	0.3	0.1	0.0

We wish to use the cosine amplitude method to express these data as a fuzzy relation. Equation (3.23) for an element in the fuzzy relation, r_{ij}, thus takes on the specific form

$$r_{ij} = \frac{\left|\sum_{k=1}^{3} x_{ik}x_{jk}\right|}{\sqrt{\left(\sum_{k=1}^{3} x_{ik}^2\right)\left(\sum_{k=1}^{3} x_{jk}^2\right)}}.$$

For example, for $i = 1$ and $j = 2$, we get

$$r_{12} = \frac{0.3\times 0.2 + 0.6\times 0.4 + 0.1\times 0.4}{\left[(0.3)^2 + 0.6^2 + 0.1^2)\left(0.2^2 + 0.4^2 + 0.4^2\right)\right]^{1/2}} = \frac{0.34}{[0.46\times 0.36]^{1/2}} = 0.836.$$

Computing the other elements of the relation results in the following tolerance relation:

$$\underset{\sim}{R}_1 = \begin{bmatrix} 1 & & & & \\ 0.836 & 1 & & \text{sym} & \\ 0.914 & 0.934 & 1 & & \\ 0.682 & 0.6 & 0.441 & 1 & \\ 0.982 & 0.74 & 0.818 & 0.774 & 1 \end{bmatrix},$$

and two compositions of $\underset{\sim}{R}_1$ produce the equivalence relation, R:

$$\underset{\sim}{R} = \underset{\sim}{R}_1^3 = \begin{bmatrix} 1 & & & & \\ 0.914 & 1 & & \text{sym} & \\ 0.914 & 0.934 & 1 & & \\ 0.774 & 0.774 & 0.774 & 1 & \\ 0.982 & 0.914 & 0.914 & 0.774 & 1 \end{bmatrix}.$$

The tolerance relation, $\underset{\sim}{R}_1$, expressed the pairwise similarity of damage for each of the regions; the equivalence relation, R, also expresses this same information but additionally can be used to classify the regions into categories with *like properties* (Chapter 10).

Max–Min Method

Another popular method, which is computationally simpler than the cosine amplitude method, is known as the *max–min method*. Although the name sounds similar to the max–min composition method, this similarity method is different from composition. It is found through simple min and max operations on pairs of the data points, x_{ij}, and is given as

$$r_{ij} = \frac{\sum_{k=1}^{m} \min(x_{ik}, x_{jk})}{\sum_{k=1}^{m} \max(x_{ik}, x_{jk})}, \quad \text{where } i,j = 1,2,\ldots,n. \tag{3.24}$$

Example 3.13

If we reconsider Example 3.12, the min–max method will produce the following result for $i = 1$, $j = 2$:

$$r_{12} = \frac{\sum_{k=1}^{3} [\min(0.3,0.2), \min(0.6,0.4), \min(0.1,0.4)]}{\sum_{k=1}^{3} [\max(0.3,0.2), \max(0.6,0.4), \max(0.1,0.4)]} = \frac{0.2+0.4+0.1}{0.3+0.6+0.4} = 0.538.$$

Computing the other elements of the relation results in the following tolerance relation:

$$\underset{\sim}{R}_1 = \begin{bmatrix} 1 & & & & \\ 0.538 & 1 & & \text{sym} & \\ 0.667 & 0.667 & 1 & & \\ 0.429 & 0.333 & 0.250 & 1 & \\ 0.818 & 0.429 & 0.538 & 0.429 & 1 \end{bmatrix}.$$

Other Similarity Methods

The list of other similarity methods is quite lengthy. Ross (1995) presents nine additional similarity methods, and others can be found in the literature.

Other Forms of the Composition Operation

Max–min and max–product (also referred to as *max–dot*) methods of composition of fuzzy relations are the two most commonly used techniques. Many other techniques are mentioned in the literature. Each method of composition of fuzzy relations reflects a special inference machine and has its own significance and applications. The max–min method is the one used by Zadeh in his original paper on approximate reasoning using natural language if–then rules. Many have claimed, since Zadeh's introduction, that this method of composition effectively expresses the approximate and interpolative reasoning used by humans when they employ linguistic propositions for deductive reasoning (Ross, 1995).

The following additional methods are among those proposed in the literature for the composition operation $\underset{\sim}{B} = \underset{\sim}{A} \circ \underset{\sim}{R}$, where $\underset{\sim}{A}$ is the input, or antecedent defined on the universe X, $\underset{\sim}{B}$ is the output, or consequent defined on universe Y, and $\underset{\sim}{R}$ is a fuzzy relation characterizing the relationship between specific inputs (x) and specific outputs (y):

$$\text{min}-\text{max} \qquad \mu_{\underset{\sim}{B}}(y) = \min_{x \in X}\{\max[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{R}}(x,y)]\}, \tag{3.25}$$

$$\text{max}-\text{max} \qquad \mu_{\underset{\sim}{B}}(y) = \max_{x \in X}\{\max[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{R}}(x,y)]\}, \tag{3.26}$$

$$\text{min}-\text{min} \qquad \mu_{\underset{\sim}{B}}(y) = \min_{x \in X}\{\min[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{R}}(x,y)]\}, \tag{3.27}$$

$$\text{max}-\text{average} \quad \mu_{\underset{\sim}{B}}(y) = \frac{1}{2}\max_{x \in X}[\mu_{\underset{\sim}{A}}(x) + \mu_{\underset{\sim}{R}}(x,y)], \tag{3.28}$$

$$\text{sum}-\text{product} \quad \mu_{\underset{\sim}{B}}(y) = f\left\{\sum_{x \in X}\left[\mu_{\underset{\sim}{A}}(x) \cdot \mu_{\underset{\sim}{R}}(x,y)\right]\right\}, \tag{3.29}$$

where $f(\cdot)$ is a logistic function (such as a sigmoid or a step function) that limits the value of the function within the interval [0, 1]. This composition method is commonly used in applications of artificial neural networks for mapping between parallel layers in a multilayer network.

It is left as an exercise for the reader (Problem 3.12) to determine the relationship among these additional forms of the composition operator for various combinations of the membership values for $\mu_{\underset{\sim}{A}}(x)$ and $\mu_{\underset{\sim}{R}}(x,y)$.

Summary

This chapter has shown some of the properties and operations of crisp and fuzzy relations. There are many more, but these will provide a sufficient foundation for the rest of the material in the text. The idea of a relation is most powerful; this modeling power is shown in subsequent chapters dealing with such issues as logic, nonlinear simulation, classification, and control. The idea of composition was introduced, and it can be seen in Chapter 12 that the composition of a relation is similar to a method used to extend fuzziness into functions, called the *extension principle*. Tolerance and equivalent relations hold some special properties, as is illustrated in Chapter 10, when they are used in similarity applications and classification applications, respectively. Finally, several similarity metrics were shown to be useful in developing the relational *strengths*, or distances, within fuzzy relations from data sets.

References

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications* (New York: Academic Press).
Gill, A. (1976). *Applied Algebra for the Computer Sciences* (Englewood Cliffs, NJ: Prentice Hall).
Kandel, A. (1985). *Fuzzy Mathematical Techniques with Applications* (Menlo Park, CA: Addison-Wesley).
Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).
Ross, T. (1995). *Fuzzy Logic and Engineering Applications* (New York: McGraw-Hill).
Zadeh, L. (1971). Similarity relations and fuzzy orderings. *Inf. Sci.*, **3**, 177–200.

Problems

General Relations

3.1 Suppose we have two fuzzy sets, $\underset{\sim}{P}$ defined on a universe of four discrete Pressures $[x_1, x_2, x_3, x_4]$ and $\underset{\sim}{T}$ defined on a universe of three discrete Temperatures $[y_1, y_2, y_3]$. These two fuzzy sets $\underset{\sim}{P}$ and $\underset{\sim}{T}$ are required to be optimum to maintain stability for a natural gas inside a tank.

The Cartesian product represents the optimum conditions (Pressure-Temperature). Let $\underset{\sim}{P} = \left\{\frac{0.4}{x_1} + \frac{0.3}{x_2} + \frac{0.5}{x_3} + \frac{0.6}{x_4}\right\}$ and $\underset{\sim}{T} = \left\{\frac{0.6}{y_1} + \frac{0.5}{y_2} + \frac{0.4}{y_3}\right\}$. Calculate the Cartesian product $\underset{\sim}{Q} = \underset{\sim}{P} \times \underset{\sim}{T}$.

3.2 In a water treatment process, we use a biological process to remove biodegradable organic matter. The organic matter is measured as the BOD, where the optimal BOD of effluent should be less than 20 mg/L. Let $\underset{\sim}{B}$ represent a fuzzy set "good effluent" on the universe of optical BOD values (20, 40, 60) as defined by the membership function

$$\mu_{\underset{\sim}{B}} = \frac{0.5}{60} + \frac{0.8}{40} + \frac{1}{20}.$$

The retention time is critical to a bioreactor; we try to find the retention time, measured in days. Let $\underset{\sim}{T}$ represent a fuzzy set called *optimal retention time* on the universe of days (6, 8, 10) as given by the membership function.

$$\mu_{\underset{\sim}{T}} = \frac{0.9}{10} + \frac{0.6}{8} + \frac{0.4}{6}$$

The utilization rate of organic food indicates the level of the treatment in the biological process, and this rate is measured on a universe of fractions from 0 to 1, where 1 is optimal. Fuzzy set $\underset{\sim}{U}$ will represent "high utilization rates," as defined by the membership function.

$$\mu_{\underset{\sim}{U}} = \frac{1}{0.9} + \frac{0.8}{0.8} + \frac{0.6}{0.7} + \frac{0.4}{0.6}$$

We can define the following relations: $\underset{\sim}{R} = \underset{\sim}{B} \times \underset{\sim}{T}$, which reflects how retention time affects BOD removal; other relations can be formed as follows:

$\underset{\sim}{S} = \underset{\sim}{T} \times \underset{\sim}{U}$, which relates how retention time affects organic food consumption; and
$\underset{\sim}{W} = \underset{\sim}{R} \circ \underset{\sim}{S}$, which represents the BOD removal and food utilization.

a. Find $\underset{\sim}{R}$ and $\underset{\sim}{S}$ using the Cartesian Products.
b. Find $\underset{\sim}{W}$ using max-min composition.
c. Find $\underset{\sim}{W}$ using max-product composition.

3.3 To make concrete, the main four components are Cement, Sand, Water and Gravel. The mixture is considered to be the best if the proportions of Cement, Sand and Gravel are 1:1.5:1. An amount of 40% by volume of water is added to make concrete paste. Now, to fill the slab with concrete the constructors need the exact proportion of concrete that produces no shortage and no waste. The mix proportion for different components in shown in fuzzy sets that follow:

$$\underset{\sim}{\text{Cement}} = \left\{\frac{0.4}{10} + \frac{0.3}{20} + \frac{0.9}{30} + \frac{0.6}{40} + \frac{0.4}{50}\right\}$$

on a universe of cubic-feet of cement.

$$\underset{\sim}{\text{Sand}} = \left\{\frac{0.3}{15} + \frac{0.4}{30} + \frac{0.8}{45} + \frac{0.7}{60} + \frac{0.4}{75}\right\}$$

on a universe of cubic-feet of sand

Then, 40% by volume of water is added to the mixture, to produce

$$\underset{\sim}{\text{Water}} = \left\{\frac{0.4}{15} + \frac{0.7}{30} + \frac{0.7}{45} + \frac{0.6}{60} + \frac{0.5}{75}\right\}$$

on a universe of volumes of water in cubic feet

a. Using fuzzy Cartesian product, find $\underset{\sim}{P} = \underset{\sim}{C} \times \underset{\sim}{S}$, where $\underset{\sim}{P}$ represents a fuzzy set called the *proportion*
b. Using max-min composition, find $\underset{\sim}{Q} = \underset{\sim}{W} \circ \underset{\sim}{P}$, where $\underset{\sim}{Q}$ represents a fuzzy set called the *overall performance of the concrete.*
c. Using max-product composition, find $\underset{\sim}{Q} = \underset{\sim}{W} \circ \underset{\sim}{P}$.

3.4 In a glass manufacturing industry, there is a need to maintain the pressure and temperature optimum for the production of glass. With changes in the stage of production, the optimum pressure and temperature also varies. Let $\underset{\sim}{T}$ be temperature and $\underset{\sim}{P}$ be pressure at various stages. We can then define a fuzzy relation $\underset{\sim}{R}$ as the required environment for glass production.

$$\underset{\sim}{R} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0.4 & 0.5 & 0.3 & 0.2 \\ 0.3 & 0.8 & 0.6 & 0.3 \\ 0.4 & 0.7 & 0.5 & 0.2 \\ 0.2 & 0.5 & 0.4 & 0.1 \end{bmatrix} \end{array}; \quad \underset{\sim}{T} = \left\{ \frac{0.7}{1} + \frac{0.9}{2} + \frac{0.6}{3} + \frac{0.3}{4} \right\}$$

Find Pressure required using max-min composition given $\underset{\sim}{R}$ and $\underset{\sim}{T}$.

3.5 Given the continuous, noninteractive fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ on universes X and Y, using Zadeh's notation for continuous fuzzy variables, as shown in Figure P3.5(a)

$$\underset{\sim}{A} = \left\{ \int (1 - 0.1|x|)/x \right\}, \quad \text{for } x \in [0, 10].$$

$$\underset{\sim}{B} = \left\{ \int \frac{0.2|y|}{y} \right\}, \quad \text{for } y \in [0, +5].$$

a. Construct a fuzzy relation $\underset{\sim}{R}$ for the Cartesian product of $\underset{\sim}{A}$ and $\underset{\sim}{B}$.
b. Use max–min composition to find $\underset{\sim}{B}'$, given the fuzzy singleton $\underset{\sim}{A}' = \frac{1}{3}$.
(See Figure P3.5(b) and P3.5(c)).
Hint: You can solve this problem graphically by segregating the Cartesian space into various regions according to the min and max operations, or you can approximate the continuous fuzzy variables as discrete variables and use matrix operations. In any case, sketch the solution.

3.6 A Rocket was launched carrying a satellite which will be reaching a distance of 22,300 miles at a speed of 4.9 m/s from Earth. During its path, the average speed ($\underset{\sim}{S}$) for every 10-minute intervals and fuel consumed ($\underset{\sim}{F}$) for every 3000 miles traveled is monitored. Given the fuzzy sets that follow,

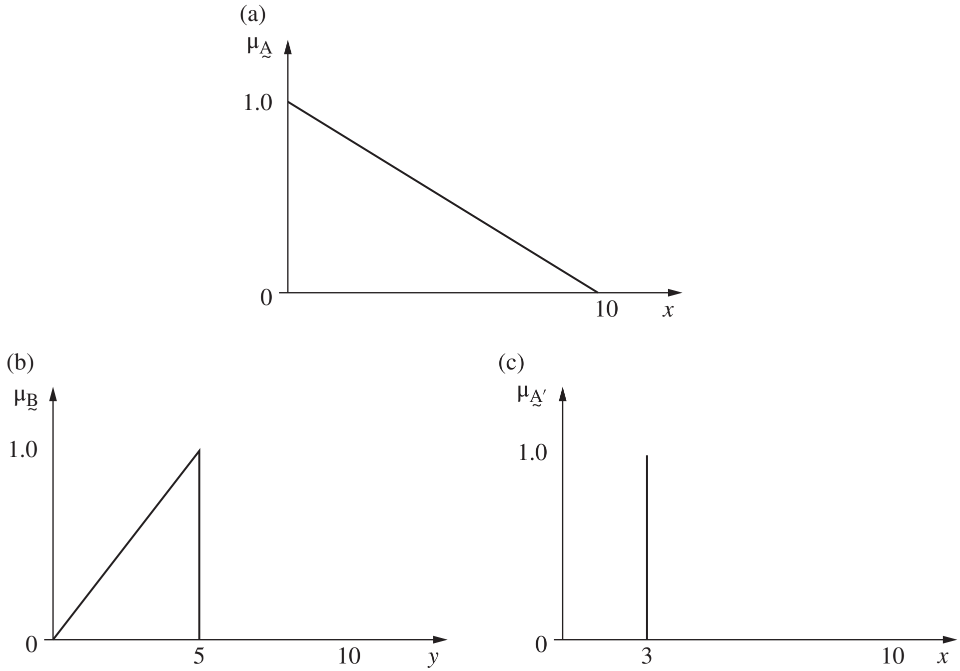

Figure P3.5

$$\underset{\sim}{S} = \left\{\frac{0.6}{10} + \frac{0.7}{20} + \frac{0.9}{30} + \frac{0.8}{40} + \frac{0.8}{50} + \frac{0.7}{60}\right\}$$

$$\underset{\sim}{F} = \left\{\frac{0.9}{3000} + \frac{0.8}{6000} + \frac{0.6}{12000} + \frac{0.3}{15000} + \frac{0.2}{18000}\right\}.$$

find a relation, $\underset{\sim}{R}$, for average speed and fuel consumed.

3.7 There is an imprecise relationship between the flexural strength and volumetric strain of a concrete beam. Let $\underset{\sim}{X}$ be a fuzzy set of flexural strength (units of ksi) and $\underset{\sim}{Y}$ be a fuzzy set of volumetric strain (units of inches/inch per cubic feet of beam) and membership functions are given as follows:

$$\underset{\sim}{X} = \left\{\frac{0.6}{3} + \frac{0.7}{4} + \frac{0.8}{5} + \frac{0.9}{6} + \frac{1}{7}\right\}$$

$$\underset{\sim}{Y} = \left\{\frac{0.3}{0.7} + \frac{0.6}{0.8} + \frac{0.8}{0.9} + \frac{0.7}{1.0} + \frac{0.4}{1.1}\right\}$$

a. Find the Cartesian product represented by the relation $\underset{\sim}{R} = \underset{\sim}{X} \times \underset{\sim}{Y}$. Now, suppose we have a second fuzzy set of flexural strength given as

$$\underset{\sim}{Z} = \left\{ \frac{0.1}{3} + \frac{0.2}{4} + \frac{0.3}{5} + \frac{0.4}{6} + \frac{0.5}{7} \right\}$$

Find the following fuzzy sets:

b. Find $\underset{\sim}{S} = \underset{\sim}{Z} \circ \underset{\sim}{R}$ using max-min composition
c. Find $\underset{\sim}{S} = \underset{\sim}{Z} \circ \underset{\sim}{R}$ using max-product composition.

3.8 High-speed rail monitoring devices sometimes make use of sensitive sensors to measure the deflection of the earth when a rail car passes. These deflections are measured with respect to some distance from the rail car and, hence, are actually small angles measured in micro radians. Let a universe of deflections be $\underset{\sim}{A} = \{1, 2, 3, 4\}$, where the units are angles in micro radians, and let a universe of distances be D = {1, 2, 5, 7}, where the units are measured in feet. Suppose a relation between these two parameters has been determined as follows:

$$\underset{\sim}{R} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{array} \begin{array}{c} \begin{array}{cccc} D_1 & D_2 & D_3 & D_4 \end{array} \\ \begin{bmatrix} 1 & 0.3 & 0.1 & 0 \\ 0.2 & 1 & 0.3 & 0.1 \\ 0 & 0.7 & 1 & 0.2 \\ 0 & 0.1 & 0.4 & 1 \end{bmatrix} \end{array}$$

Now let a universe of rail car weights be W = {1, 2}, where W is the weight in units of 100,000 pounds. Suppose the fuzzy relation for W related to $\underset{\sim}{A}$ is given as

$$\underset{\sim}{S} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{array} \begin{array}{c} \begin{array}{cc} W_1 & W_2 \end{array} \\ \begin{bmatrix} 1 & 0.4 \\ 0.5 & 1 \\ 0.3 & 0.1 \\ 0 & 0 \end{bmatrix} \end{array}$$

Using these two relations ($\underset{\sim}{R}$ and $\underset{\sim}{S}$), find the relation, $\underset{\sim}{R}^{T} \circ \underset{\sim}{S} = \underset{\sim}{T}$ (note the matrix transposition here)

a. Using max–min composition;
b. Using max–product composition.

Value Assignment and Similarity

3.9 A regular inspection for a pipe network is conducted at five different regions (denoted 1, 2, 3, 4, and 5). An engineer needs to develop a maintenance schedule for the pipes based on the severity of leaking: no leaking, moderate leaking, and serious leaking. The strategy will be to replace the serious leaking pipes, and to repair the moderately leaking pipes, and to do nothing for the pipes that don't leak. For this problem our parameters are $n = 5$ and $m = 3$. An actual inspection for each region gave the following percentages for the category of leaking:

Regions	**1**	**2**	**3**	**4**	**5**
No Leaking	0.3	0.7	0.4	0.1	0.0
Moderately Leaking	0.4	0.3	0.1	0.8	0.3
Serious Leaking	0.3	0.0	0.5	0.1	0.7

a. Use the max-min membership method to find the similarity matrix $\underset{\sim}{\mathrm{R}}$.
b. Use the cosine method to find the similarity matrix $\underset{\sim}{\mathrm{R}}$.

3.10 Tests were performed on four different grades of concrete materials to see their failure mode. The failure modes were tensile, compressive and shear failures. The values in the table show the fractions of each grade that failed in each of the three modes. It was found that Tensile, Compressive and Shear strength varies from one another ($n = 4$, $m = 3$)

Material	**1**	**2**	**3**	**4**
Tensile	0.3	0.6	0.5	0.7
Compressive	0.6	0.3	0.8	0.7
Shear	0.8	0.5	0.7	0.3

a. Use the cosine amplitude method to find the similarity, $\underset{\sim}{\mathrm{R}}$.
b. The relation found in (a) is a tolerance relation. Find the associated equivalence relation.

Equivalence Relations

3.11 The accompanying sagittal diagrams shows two relations on the universe, X = [1, 2, 3]. Are the relations shown in Figure P3.11 equivalence relations (Gill, 1976)? (See Figure P3.11(a) and P3.11(b).)

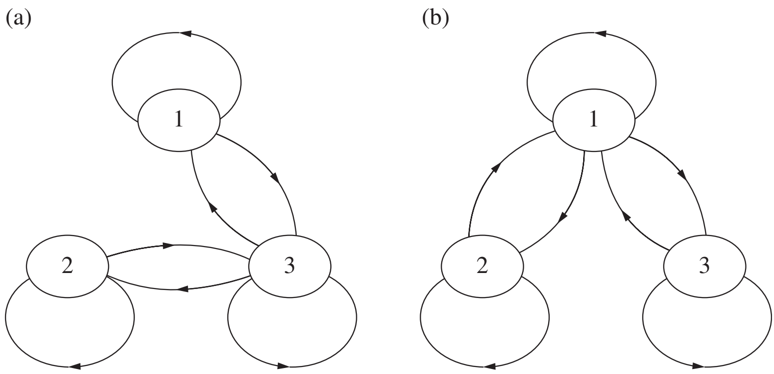

Figure P3.11

Other Composition Operations

3.12 Fill in the following table using the Equations (3.25) to (3.29) to determine values of the composition $\underset{\sim}{B} = \underset{\sim}{A} \circ \underset{\sim}{R}$ for the fuzzy relation given here.

$$\underset{\sim}{R} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 0.7 & 0.6 & 0.5 \\ 0.3 & 0.2 & 0.5 \\ 0.2 & 0.4 & 0.7 \end{bmatrix} \end{array}$$

Compare the similarities and dissimilarities of the various composition methods with respect to the various antecedents of:

	$\underset{\sim}{\mathbf{A}}$	$\underset{\sim}{\mathbf{B}}$
a)	$[0.4 \;\; 0.5 \;\; 0.3]$	?
b)	$[0.3 \;\; 0.2 \;\; 0.3]$	?

4

Properties of Membership Functions, Fuzzification, and Defuzzification

> *"Let's consider your age, to begin with—how old are you?" "I'm seven and a half, exactly." "You needn't say 'exactually,'" the Queen remarked; "I can believe it without that. Now I'll give you something to believe. I'm just one hundred and one, five months, and a day." "I can't believe that!" said Alice.*
>
> *"Can't you?" the Queen said in a pitying tone. "Try again; draw a long breath, and shut your eyes." Alice laughed. "There's no use trying," she said; "one can't believe impossible things."*
>
> *Lewis Carroll, Through the Looking Glass*, 1871

It is one thing to compute, to reason, and to model with fuzzy information; it is another to apply the fuzzy results to the world around us. Despite the fact that the bulk of the information we assimilate every day is fuzzy, such as the age of people in the Lewis Carroll quotation, most of the actions or decisions implemented by humans or machines are crisp or binary. The decisions we make that require an action are binary, the hardware we use is binary, and certainly the computers we use are based on binary digital instructions. For example, in making a decision about developing a new engineering product, the eventual decision is to go forward with development or not; the fuzzy choice to "maybe go forward" might be acceptable in planning stages, but eventually funds are released for development or they are not. In giving instructions to an aircraft autopilot, it is not possible to turn the plane "slightly to the west"; an autopilot device does not understand the natural language of a human. We have to turn the plane by 15 degrees, for example, a crisp number. An electrical circuit typically is either on or off, not partially on.

The bulk of this text illustrates procedures to "fuzzify" the mathematical and engineering principles we have so long considered to be deterministic. But, in various applications and engineering scenarios, there will be a need to "defuzzify" the fuzzy results we generate through a fuzzy systems analysis. In other words, we may eventually find a need to convert the fuzzy results to crisp results. For example, in classification and pattern recognition (Chapter 10) we may want to transform a fuzzy partition or pattern into a crisp partition or pattern; in control (Chapter 11), we may want to give a single-valued input to a semiconductor device instead of a fuzzy input command. This "defuzzification" has the result of reducing a fuzzy set to a crisp single-valued quantity, or to a crisp set; of converting a fuzzy matrix to a crisp matrix; or of making a fuzzy number a crisp number.

Mathematically, the defuzzification of a fuzzy set is the process of "rounding it off" from its location in the unit hypercube to the nearest (in a geometric sense) vertex (Chapter 1). If one thinks of a fuzzy set as a collection of membership values, or a vector of values on the unit interval, defuzzification reduces this vector to a single scalar quantity, presumably to the most typical (prototype) or representative value. Various popular forms of converting fuzzy sets to crisp sets or to single scalar values are introduced later in this chapter.

Features of the Membership Function

Because all information contained in a fuzzy set is described by its membership function, it is useful to develop a lexicon of terms to describe various special features of this function. For purposes of simplicity, the functions shown in the figures will all be continuous, but the terms apply equally for both discrete and continuous fuzzy sets. Figure 4.1 assists in this description.

The *core* of a membership function for some fuzzy set $\underset{\sim}{A}$ is defined as that region of the universe that is characterized by complete and full membership in the set $\underset{\sim}{A}$. That is, the core comprises those elements x of the universe such that $\mu_{\underset{\sim}{A}}(x) = 1$.

The *support* of a membership function for some fuzzy set $\underset{\sim}{A}$ is defined as that region of the universe that is characterized by nonzero membership in the set $\underset{\sim}{A}$. That is, the support comprises those elements x of the universe such that $\mu_{\underset{\sim}{A}}(x) > 0$.

The *boundaries* of a membership function for some fuzzy set $\underset{\sim}{A}$ are defined as that region of the universe containing elements that have a nonzero membership but not complete membership. That is, the boundaries comprise those elements x of the universe such that $0 < \mu_{\underset{\sim}{A}}(x) < 1$. These elements of the universe are those with some *degree* of fuzziness, or only partial membership in the fuzzy set $\underset{\sim}{A}$. Figure 4.1 illustrates the regions in the universe comprising the core, support, and boundaries of a typical fuzzy set.

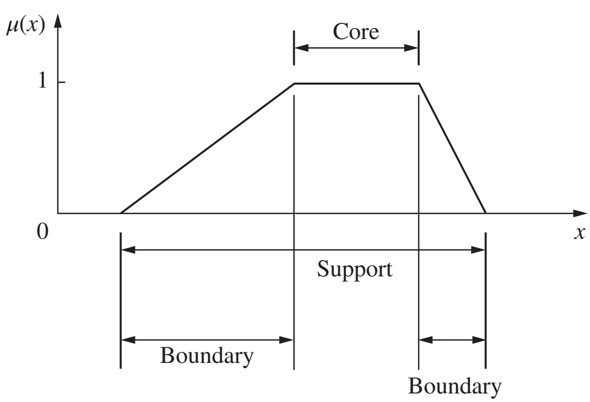

Figure 4.1 Core, support, and boundaries of a fuzzy set.

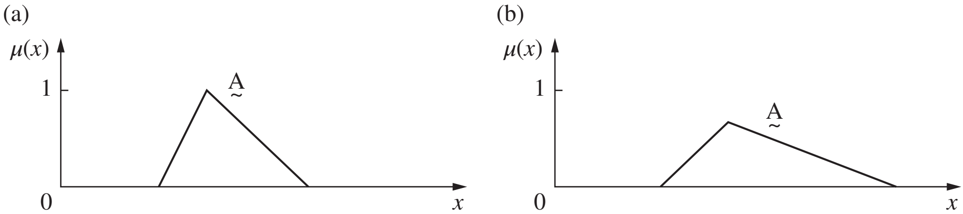

Figure 4.2 Fuzzy sets that are normal (a) and subnormal (b).

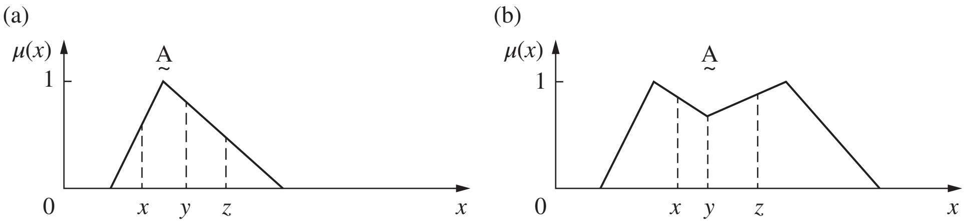

Figure 4.3 Convex, normal fuzzy set (a) and nonconvex, normal fuzzy set (b).

A *normal* fuzzy set is one whose membership function has at least one element x in the universe whose membership value is unity. In fuzzy sets, where one and only one element has a membership equal to one, the element is typically referred to as the *prototype* of the set, or the *prototypical element*. Figure 4.2 illustrates typical normal and subnormal fuzzy sets.

A *convex* fuzzy set is described by a membership function whose membership values are strictly monotonically increasing, or whose membership values are strictly monotonically decreasing, or whose membership values are strictly monotonically increasing then strictly monotonically decreasing with increasing values for elements in the universe. Said another way, if, for any elements x, y, and z in a fuzzy set $\underset{\sim}{A}$, the relation $x < y < z$ implies that $\mu_{\underset{\sim}{A}}(y) \geq \min[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{A}}(Z)]$, then $\underset{\sim}{A}$ is said to be a convex fuzzy set (Ross, 1995). Figure 4.3 shows a typical convex fuzzy set and a typical nonconvex fuzzy set. It is important to remark here that this definition of convexity is *different* from some definitions of the same term in mathematics. In some areas of mathematics, convexity of shape has to do with whether a straight line through any part of the shape goes outside the boundaries of that shape. This definition of convexity is *not* used here; Figure 4.3 succinctly summarizes our definition of convexity.

A special property of two convex fuzzy sets, say $\underset{\sim}{A}$ and B, is that the intersection of these two convex fuzzy sets is also a convex fuzzy set, as shown in Figure 4.4. That is, for $\underset{\sim}{A}$ and B, which are both convex, $\underset{\sim}{A} \cap \underset{\sim}{B}$ is also convex.

The *crossover points* of a membership function are defined as the elements in the universe for which a particular fuzzy set $\underset{\sim}{A}$ has values equal to 0.5, that is, for which $\mu_{\underset{\sim}{A}}(x) = 0.5$.

The *height* of a fuzzy set $\underset{\sim}{A}$ is the maximum value of the membership function, that is, $\text{hgt}(\underset{\sim}{A}) = \max\{\mu_{\underset{\sim}{A}}(x)\}$. If the $\text{hgt}(\underset{\sim}{A}) < 1$, the fuzzy set is said to be subnormal. The $\text{hgt}(\underset{\sim}{A})$ may be viewed as the degree of validity or credibility of information expressed by $\underset{\sim}{A}$ (Klir and Yuan, 1995).

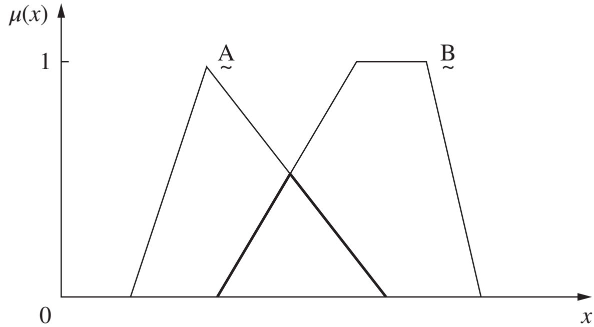

Figure 4.4 The intersection of two convex fuzzy sets produces a convex fuzzy set.

If $\underset{\sim}{A}$ is a convex single-point normal fuzzy set defined on the real line, then $\underset{\sim}{A}$ is often termed a *fuzzy number*.

Various Forms

The most common forms of membership functions are those that are normal and convex. However, many operations on fuzzy sets, hence operations on membership functions, result in fuzzy sets that are subnormal and nonconvex. For example, the extension principle to be discussed in Chapter 12 and the union operator both can produce subnormal or nonconvex fuzzy sets.

Membership functions can be symmetrical or asymmetrical. They are typically defined on one-dimensional universes, but they certainly can be described on multidimensional (or n-dimensional) universes. For example, the membership functions shown in this chapter are one-dimensional curves. In two dimensions, these curves become surfaces and for three or more dimensions these surfaces become hypersurfaces. These hypersurfaces, or curves, are simple mappings from combinations of the parameters in the n-dimensional space to a membership value on the interval [0, 1]. Again, this membership value expresses the degree of membership that the specific combination of parameters in the n-dimensional space has in a particular fuzzy set defined on the n-dimensional universe of discourse. The hypersurfaces for an n-dimensional universe are analogous to joint probability density functions, but, of course, the mapping for the membership function is to membership in a particular set and not to relative frequencies, as it is for probability density functions.

Fuzzy sets of the types depicted in Figure 4.2 are by far the most common ones encountered in practice; they are described as *ordinary membership functions*. However, several other types of fuzzy membership functions have been proposed (Klir and Yuan, 1995) as *generalized membership functions*. The primary reason for considering other types of membership functions is that the values used in developing ordinary membership functions are often overly precise. They require that each element x of the universe on which the fuzzy set $\underset{\sim}{A}$ is defined be assigned a specific membership value, $\mu_{\underset{\sim}{A}}(x)$. Suppose the level of information is not adequate to specify membership functions with this precision. For example, we may know only the upper and lower bounds of membership grades for each element of the universe for a fuzzy set. Such a fuzzy set would be described by an *interval-valued membership function*, such as the one shown in

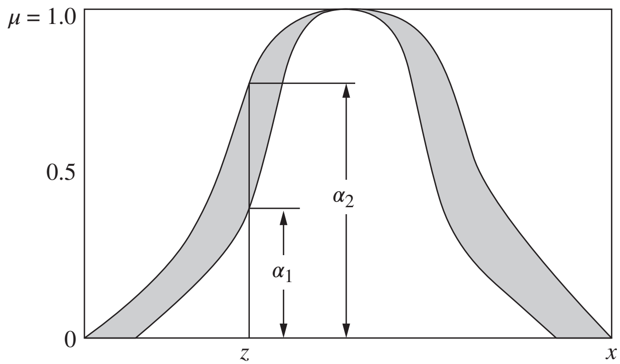

Figure 4.5 An interval-valued membership function.

Figure 4.5. In this figure, for a particular element, $x = z$, the membership in a fuzzy set A, that is, $\mu_{\tilde{A}}(z)$, would be expressed by the membership interval $[\alpha_1, \alpha_2]$. Interval-valued fuzzy sets can be generalized further by allowing their intervals to become fuzzy. Each membership interval then becomes an ordinary fuzzy set. This type of membership function is referred to in the literature as a *type-2 fuzzy set*. Other generalizations of the fuzzy membership functions are available as well (see Klir and Yuan, 1995).

Fuzzification

Fuzzification is the process of making a crisp quantity fuzzy. We do this by simply recognizing that many of the quantities that we consider to be crisp and deterministic are actually not deterministic at all; they carry considerable uncertainty. If the form of uncertainty happens to arise because of imprecision, ambiguity, or vagueness, then the variable is probably fuzzy and can be represented by a membership function.

In the real world, hardware such as a digital voltmeter generates crisp data, but these data are subject to experimental error. The information shown in Figure 4.6 shows one possible range of errors for a typical voltage reading and the associated membership function that might represent such imprecision.

The representation of imprecise data as fuzzy sets is a useful but not mandatory step when those data are used in fuzzy systems. This idea is shown in Figure 4.7, where we consider the data as a crisp reading, Figure 4.7a, or as a fuzzy reading, as shown in Figure 4.7b. In Figure 4.7a, we might want to compare a crisp voltage reading to a fuzzy set, say “low voltage.” In the figure, we see that the crisp reading intersects the fuzzy set “low voltage” at a membership of 0.3; that is, the fuzzy set and the reading can be said to agree at a membership value of 0.3. In Figure 4.7b, the intersection of the fuzzy set “medium voltage” and a fuzzified voltage reading occurs at a membership of 0.4. We can see in Figure 4.7b that the set intersection of the two fuzzy sets is a small triangle, whose largest membership occurs at the membership value of 0.4.

More is discussed about the importance of fuzzification of crisp variables in Chapters 8 and 11. In Chapter 8, the topic is simulation, and the inputs for any nonlinear or complex simulation

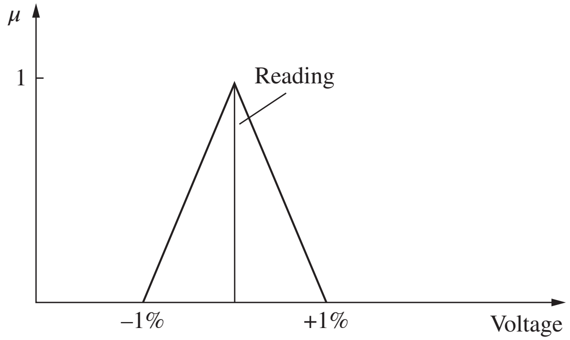

Figure 4.6 Membership function representing imprecision in "crisp voltage reading."

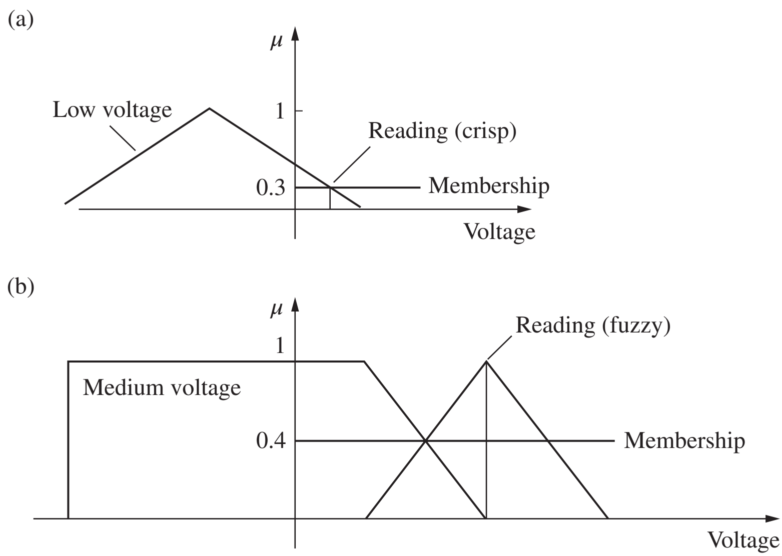

Figure 4.7 Comparisons of fuzzy sets and crisp or fuzzy readings: (a) fuzzy set and crisp reading; (b) fuzzy set and fuzzy reading.

will be expressed as fuzzy sets. If the process is inherently quantitative or the inputs derive from sensor measurements, then these crisp numerical inputs could be fuzzified for them to be used in a fuzzy inference system (discussed in Chapter 5). In Chapter 11, the topic is fuzzy control, and, again, this is a discipline where the inputs generally originate from a piece of hardware, or a sensor and the measured input could be fuzzified for utility in the rule-based system that describes the fuzzy controller. If the system to be controlled is not hardware based, for example, the control of an economic system or the control of an ecosystem subjected to a toxic chemical, then the inputs could be scalar quantities arising from statistical sampling, or other derived numerical quantities. Again, for utility in fuzzy systems, these scalar quantities should first be fuzzified, that is, translated into a membership function, and then they can then be used to form the input structure necessary for a fuzzy system.

Defuzzification to Crisp Sets

We begin by considering a fuzzy set A, then define a lambda-cut set, A_λ, where $0 \leq \lambda \leq 1$. The set A_λ is a crisp set called the *lambda (λ)-cut* (or *alpha-cut*) set of the fuzzy set A, where $A_\lambda = \{x|\mu_{\underset{\sim}{A}}(x) \geq \lambda\}$. Note that the λ-cut set A_λ does not have a tilde underscore; it is a crisp set derived from its parent fuzzy set, A. Any particular fuzzy set $\underset{\sim}{A}$ can be transformed into an infinite number of λ-cut sets, because there are an infinite number of values λ on the interval [0, 1].

Any element $x \in A_\lambda$ belongs to $\underset{\sim}{A}$ with a grade of membership that is greater than or equal to the value λ. The following example illustrates this idea.

Example 4.1

Let us consider the discrete fuzzy set, using Zadeh's notation, defined on universe X = {*a*, *b*, *c*, *d*, *e*, *f*},

$$\underset{\sim}{A} = \left\{\frac{1}{a} + \frac{0.9}{b} + \frac{0.6}{c} + \frac{0.3}{d} + \frac{0.01}{e} + \frac{0}{f}\right\}.$$

This fuzzy set is shown schematically in Figure 4.8. We can reduce this fuzzy set into several λ-cut sets, all of which are crisp. For example, we can define λ-cut sets for the values of λ = 1, 0.9, 0.6, 0.3, 0^+, and 0.

$$A_1 = \{a\}, \quad A_{0.9} = \{a,b\},$$

$$A_{0.6} = \{a,b,c\}, \quad A_{0.3} = \{a,b,c,d\},$$

$$A_{0^+} = \{a,b,c,d,e\}, \quad A_0 = X.$$

The quantity $\lambda = 0^+$ is defined as a small "ε" value >0, that is, a value just greater than zero. By definition, λ = 0 produces the universe X because all elements in the universe have at least a 0 membership value in any set on the universe. Since all A_λ are crisp sets, all the elements just shown in the example λ-cut sets have unit membership in the particular λ-cut set. For example, for λ = 0.3, the elements *a*, *b*, *c*, and *d* of the universe have a membership of 1 in the λ-cut set, $A_{0.3}$, and the elements *e* and *f* of the universe have a membership of 0 in the λ-cut set $A_{0.3}$. Figure 4.9 shows schematically the crisp λ-cut sets for the values λ = 1, 0.9, 0.6, 0.3, 0^+, and 0. Note, in these plots of membership value versus the universe X, that the effect of a λ-cut is to rescale the membership values to one for all elements of the fuzzy set $\underset{\sim}{A}$ having membership values greater than or equal to λ and to zero for all elements of the fuzzy set $\underset{\sim}{A}$ having membership values less than λ.

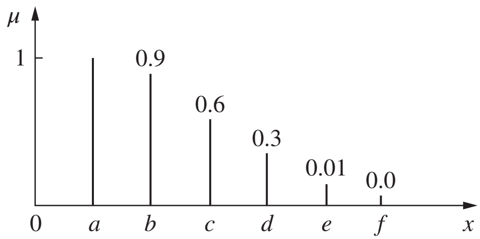

Figure 4.8 A discrete fuzzy set $\underset{\sim}{A}$.

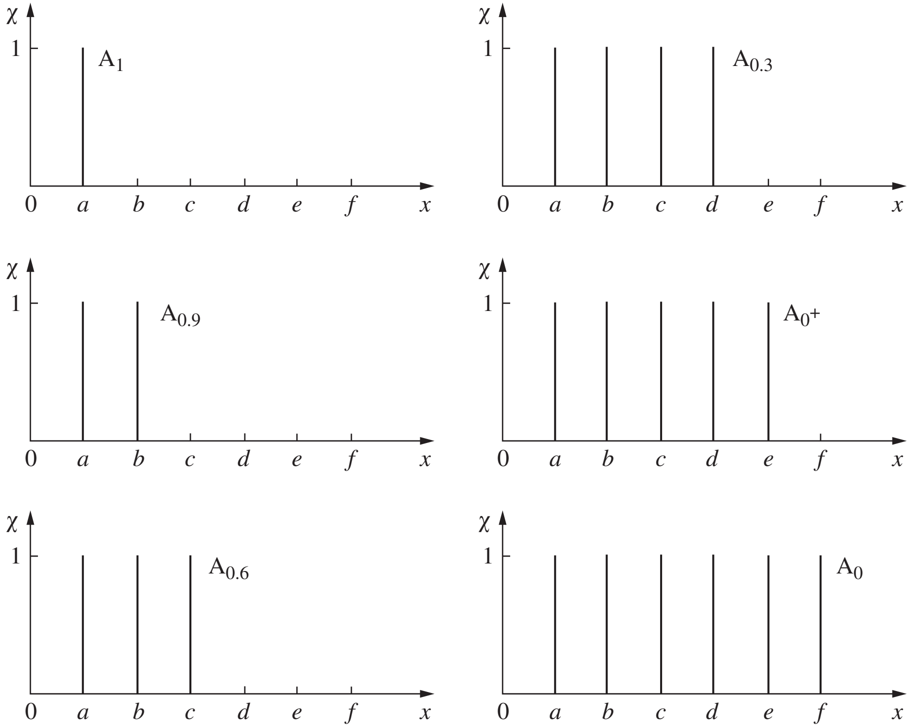

Figure 4.9 Lambda-cut sets for $\lambda = 1, 0.9, 0.6, 0.3, 0^{+}, 0$.

We can express λ-cut sets using Zadeh's notation. For example, λ-cut sets for the values $\lambda = 0.9$ and 0.25 are given here:

$$A_{0.9} = \left\{\frac{1}{a} + \frac{1}{b} + \frac{0}{c} + \frac{0}{d} + \frac{0}{e} + \frac{0}{f}\right\} \quad \text{and} \quad A_{0.25} = \left\{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} + \frac{0}{e} + \frac{0}{f}\right\}.$$

λ-cut sets obey the following four very special properties:

1. $(\underset{\sim}{A} \cup \underset{\sim}{B})_{\lambda} = A_{\lambda} \cup B_{\lambda}$ (4.1a)

2. $(\underset{\sim}{A} \cap \underset{\sim}{B})_{\lambda} = A_{\lambda} \cap B_{\lambda}$ (4.1b)

3. $(\overline{\underset{\sim}{A}})_{\lambda} \neq \overline{A}_{\lambda}$ except for a value of $\lambda = 0.5$ (4.1c)

4. For any $\lambda \leq \alpha$, where $0 \leq \alpha \leq 1$, it is true that $A_{\alpha} \subseteq A_{\lambda}$, where $A_0 = X$ (4.1d)

These properties show that λ-cuts on standard operations on fuzzy sets are equivalent with standard set operations on λ-cut sets. The last operation, Equation (4.1d), can be shown more conveniently using graphics. Figure 4.10 shows a continuous-valued fuzzy set with two λ-cut

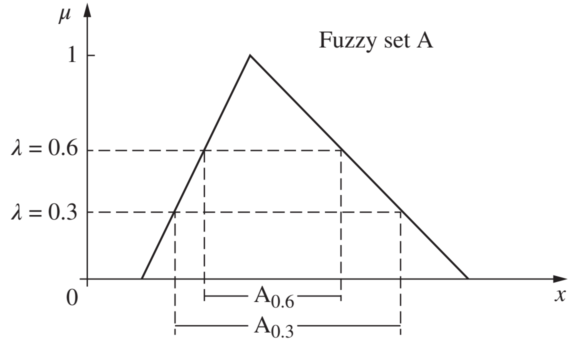

Figure 4.10 Two different λ-cut sets for a continuous-valued fuzzy set.

values (λ and α). Notice in the graphic that for $\lambda = 0.3$ and $\alpha = 0.6$, $A_{0.3}$ has a greater domain than $A_{0.6}$, that is, for $\lambda \le \alpha (0.3 \le 0.6)$, $A_{0.6} \subseteq A_{0.3}$.

In this chapter, various definitions of a membership function are discussed and illustrated. Many of these same definitions arise through the use of λ-cut sets. As seen in Figure 4.1, we can provide the following definitions for a convex fuzzy set $\underset{\sim}{A}$. The core of $\underset{\sim}{A}$ is the $\lambda = 1$ cut set, A_1. The support of $\underset{\sim}{A}$ is the λ-cut set A_{0^+}, where $\lambda = 0^+$, or symbolically, $A_{0^+} = \{x | \mu_{\underset{\sim}{A}}(x) > 0\}$. The intervals $[A_{0^+}, A_1]$ form the boundaries of the fuzzy set $\underset{\sim}{A}$, that is, those regions that have membership values between 0 and 1 (exclusive of 0 and 1): that is, for $0 < \lambda < 1$.

λ-Cuts for Fuzzy Relations

In Chapter 3, a biotechnology example, Example 3.11, was developed using a fuzzy relation that was reflexive and symmetric. Recall this matrix,

$$\underset{\sim}{R} = \begin{bmatrix} 1 & 0.8 & 0 & 0.1 & 0.2 \\ 0.8 & 1 & 0.4 & 0 & 0.9 \\ 0 & 0.4 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0.5 \\ 0.2 & 0.9 & 0 & 0.5 & 1 \end{bmatrix}.$$

We can define a λ-cut procedure for relations similar to the one developed for sets. Consider a fuzzy relation $\underset{\sim}{R}$, where each row of the relational matrix is considered a fuzzy set, that is, the jth row in $\underset{\sim}{R}$ represents a discrete membership function for a fuzzy set, $\underset{\sim}{R}_j$. Hence, a fuzzy relation can be converted to a crisp relation in the following manner. Let us define $R_\lambda = \{(x, y) | \mu_{\underset{\sim}{R}}(x, y) \ge \lambda\}$ as a λ-cut relation of the fuzzy relation $\underset{\sim}{R}$. Because in this case $\underset{\sim}{R}$ is a two-dimensional array defined on the universes X and Y, any pair $(x, y) \in R_\lambda$ belongs to $\underset{\sim}{R}$ with a "strength" of relation greater than or equal to λ. These ideas for relations can be illustrated with an example.

Example 4.2
Suppose we take the fuzzy relation from the biotechnology example in Chapter 3 (Example 3.11), and perform λ-cut operations for the values of $\lambda = 1$, 0.9, 0. These crisp relations are given as follows:

$$\lambda = 1, \mathrm{R}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\lambda = 0.9, \mathrm{R}_{0.9} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix},$$

$$\lambda = 0, \mathrm{R}_0 = \underset{\sim}{\mathbf{E}}\,(\text{whole relation; see Chapter 3}).$$

λ-cuts on fuzzy relations obey certain properties, just as λ-cuts on fuzzy sets do (Equation (4.1)), as given in Equation (4.2):

1. $(\underset{\sim}{\mathrm{R}} \cup \underset{\sim}{\mathrm{S}})_\lambda = \mathrm{R}_\lambda \cup \mathrm{S}_\lambda$ (4.2a)

2. $(\underset{\sim}{\mathrm{R}} \cap \underset{\sim}{\mathrm{S}})_\lambda = \mathrm{R}_\lambda \cap \mathrm{S}_\lambda$ (4.2b)

3. $(\overline{\underset{\sim}{\mathrm{R}}})_\lambda \neq \overline{\underset{\sim}{\mathrm{R}}}_\lambda$ (4.2c)

4. For any $\lambda \leq \alpha, 0 \leq \alpha \leq 1$, then $\mathrm{R}_\alpha \subseteq \mathrm{R}_\lambda$ (4.2d)

Defuzzification to Scalars

As mentioned in the introduction, there may be situations in which the output of a fuzzy process needs to be a single scalar quantity as opposed to a fuzzy set. Defuzzification is the conversion of a fuzzy quantity to a precise quantity, just as fuzzification is the conversion of a precise quantity to a fuzzy quantity. The output of a fuzzy process can be the logical union of two or more fuzzy membership functions defined on the universe of discourse of the output variable. For example, suppose a fuzzy output comprises two parts: (1) $\underset{\sim}{\mathrm{C}}_1$, a trapezoidal shape (Figure 4.11a) and (2) $\underset{\sim}{\mathrm{C}}_2$, a triangular membership shape (Figure 4.11b). The union of these

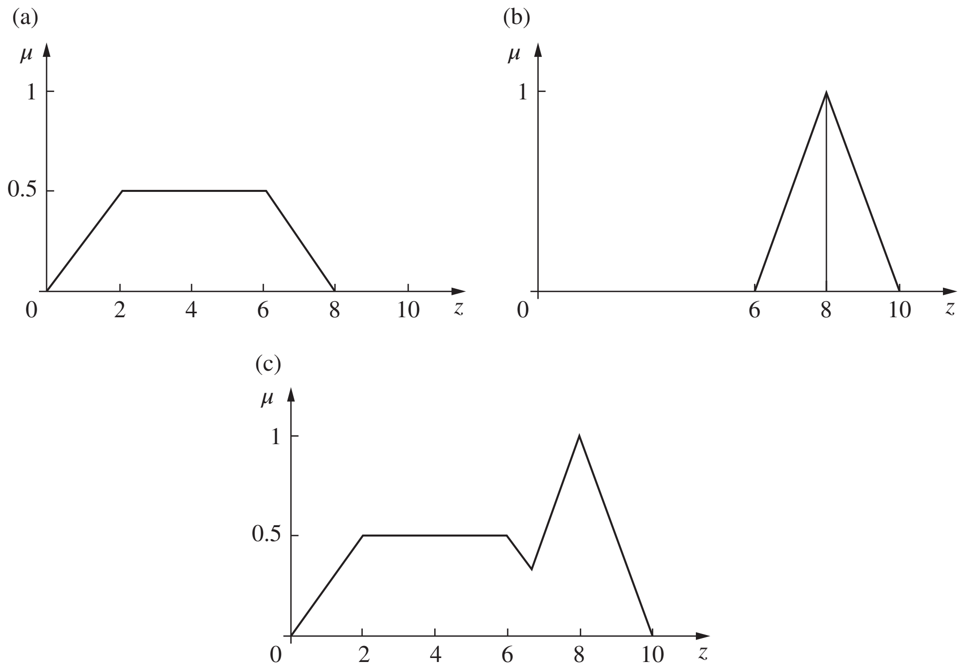

Figure 4.11 Typical fuzzy process output: (a) first part of fuzzy output; (b) second part of fuzzy output; and (c) union of both parts.

two membership functions, that is, $\underset{\sim}{C} = \underset{\sim}{C}_1 \cup \underset{\sim}{C}_2$, involves the max operator, which graphically is the outer envelope of the two shapes shown in Figure 4.11a and b; the resulting shape is shown in Figure 4.11c. Of course, a general fuzzy output process can involve many output parts (more than two), and the membership function representing each part of the output can have shapes other than triangles and trapezoids. Further, as Figure 4.11a shows, the membership functions may not always be normal. In general, we can have

$$\underset{\sim}{C}_k = \bigcup_{i=1}^{k} \underset{\sim}{C}_i = \underset{\sim}{C}. \tag{4.3}$$

Among the many methods that have been proposed in the literature in recent years, four are described here for defuzzifying fuzzy output functions (membership functions) (Hellendoorn and Thomas, 1993). These methods are summarized and illustrated in two examples.

1. *Max membership principle*: Also known as the *height method*, this scheme is limited to peaked output functions. This method is given by the algebraic expression

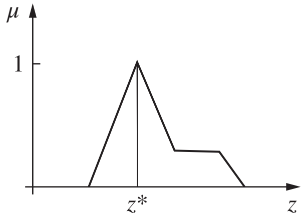

Figure 4.12 Max membership defuzzification method.

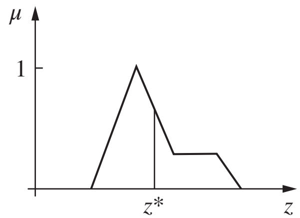

Figure 4.13 Centroid defuzzification method.

$$\mu_{\underset{\sim}{C}}(z^*) \geq \mu_{\underset{\sim}{C}}(z), \quad \text{for all } z \in Z, \tag{4.4}$$

where z^* is the defuzzified value, and is shown graphically in Figure 4.12.

2. *Centroid method*: This procedure (also called *center of area* or *center of gravity*) is the most prevalent and physically appealing of all the defuzzification methods (Sugeno, 1985; Lee, 1990); it is given by the algebraic expression

$$z^* = \frac{\int \mu_{\underset{\sim}{C}}(z) \cdot z \mathrm{d}z}{\int \mu_{\underset{\sim}{C}}}(z)\,\mathrm{d}z, \tag{4.5}$$

where $\int$ denotes an algebraic integration. This method is shown in Figure 4.13.

3. *Weighted average method*: The weighted average method is the most frequently used in fuzzy applications because it is one of the more computationally efficient methods. Unfortunately, it is *usually* restricted to symmetrical output membership functions. It is given by the algebraic expression

$$z^* = \frac{\sum \mu_{\underset{\sim}{C}}(\bar{z}) \cdot \bar{z}}{\int \mu_{\underset{\sim}{C}}}(\bar{z}), \tag{4.6}$$

where $\sum$ denotes the algebraic sum and where $\bar{z}$ is the centroid of each symmetric membership function. This method is shown in Figure 4.14. The weighted average method is

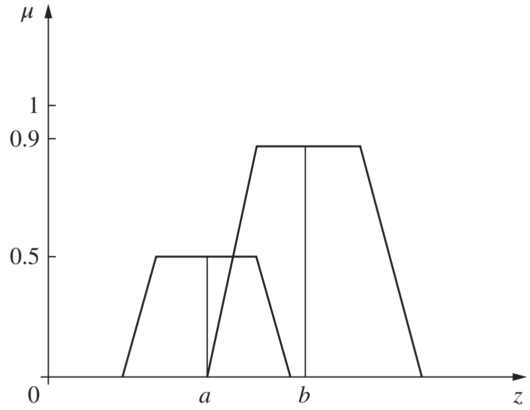

Figure 4.14 Weighted average method of defuzzification.

formed by weighting each membership function in the output by its respective maximum membership value. As an example, the two functions shown in Figure 4.14 would result in the following general form for the defuzzified value:

$$z^* = \frac{a(0.5) + b(0.9)}{0.5 + 0.9}.$$

Because the method can be limited to symmetrical membership functions, the values a and b are the means (centroids) of their respective shapes. This method is sometimes applied to unsymmetrical functions and various scalar outputs (see Sugeno, 1985).

4. *Mean max membership*: This method (also called *middle-of-maxima*) is closely related to the first method, except that the locations of the maximum membership can be nonunique (i.e., the maximum membership can be a plateau rather than a single point). This method is given by the expression (Sugeno, 1985; Lee, 1990)

$$z^* = \frac{a + b}{2} \tag{4.7}$$

where a and b are as defined in Figure 4.15.

Example 4.3

A railroad company intends to lay a new rail line in a particular part of a county. The whole area through which the new line is passing must be purchased for right-of-way considerations. It is surveyed in three stretches, and the data are collected for analysis. The surveyed data for the road are given by the sets, $\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, and $\underset{\sim}{B}_3$, where the sets are defined on the universe of right-of-way widths, in meters. For the railroad to purchase the land, it must have an assessment of the amount of land to be bought. The three surveys on right-of-way width are ambiguous, however, because some of the land along the proposed railway route is already public domain and will not need to be purchased. Additionally, the original surveys are so old (*circa* 1860) that some

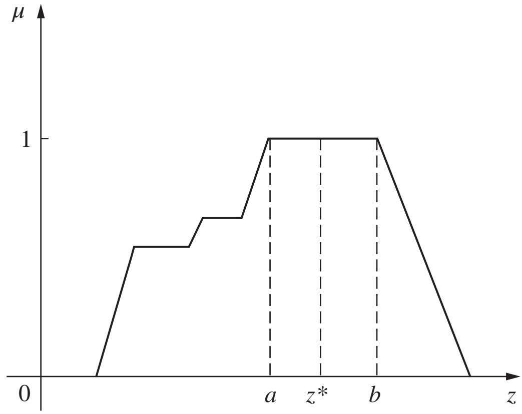

Figure 4.15 Mean max membership defuzzification method.

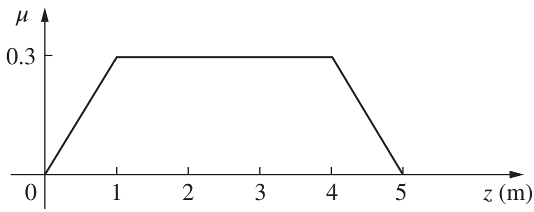

Figure 4.16 Fuzzy set $\underset{\sim}{B}_1$: public right-of-way width (z) for survey 1.

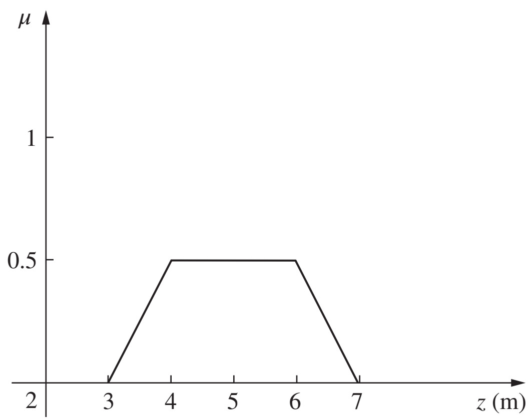

Figure 4.17 Fuzzy set $\underset{\sim}{B}_2$: public right-of-way width (z) for survey 2.

ambiguity exists on boundaries and public right-of-way for old utility lines and old roads. The three fuzzy sets, $\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, and $\underset{\sim}{B}_3$ shown in Figures 4.16 to 4.18, respectively, represent the uncertainty in each survey as to the membership of right-of-way width, in meters, in privately owned land.

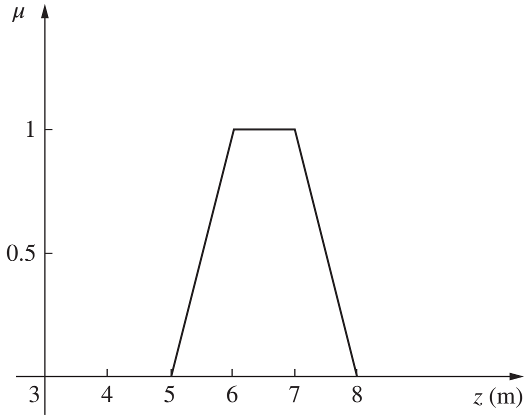

Figure 4.18 Fuzzy set $\underset{\sim}{B}_3$: public right-of-way width (z) for survey 3.

We now want to aggregate these three survey results to find the single-most nearly representative right-of-way width (z) to allow the railroad to make its initial estimate of the right-of-way purchasing cost. Using Equations (4.5) to (4.7) and the preceding three fuzzy sets, we want to find z^*.

According to the centroid method, Equation (4.5), z^* can be found using

$$
\begin{aligned}
z^* &= \frac{\int \mu_{\underset{\sim}{B}}(z)\cdot z\mathrm{d}z}{\int \mu_{\underset{\sim}{B}}}(z)\,\mathrm{d}z \\
&= \left[\int_0^1 (0.3z)z\mathrm{d}z + \int_1^{3.6} (0.3)z\mathrm{d}z + \int_{3.6}^4 \left(\frac{z-3.0}{2}\right)z\mathrm{d}z + \int_4^{5.5} (0.5)z\mathrm{d}z \right.\\
&\quad \left. + \int_{5.5}^6 (z-5)z\mathrm{d}z + \int_6^7 z\mathrm{d}z + \int_7^8 (8-z)z\mathrm{d}z\right] \\
&\quad \div \left[\int_0^1 (0.3z)\mathrm{d}z + \int_1^{3.6} (0.3)\mathrm{d}z + \int_{3.6}^4 \left(\frac{z-3.6}{2}\right)\mathrm{d}z + \int_4^{5.5} (0.5)\mathrm{d}z \right.\\
&\quad \left. + \int_{5.5}^6 \left(\frac{z-5.5}{2}\right)\mathrm{d}z + \int_6^7 \mathrm{d}z + \int_7^8 \left(\frac{7-z}{2}\right)\mathrm{d}z\right] \\
&= 4.9\,\mathrm{m},
\end{aligned}
$$

where z^* is shown in Figure 4.19. According to the weighted average method, Equation (4.6),

$$z^* = \frac{(0.3\times 2.5)+(0.5\times 5)+(1\times 6.5)}{0.3+0.5+1} = 5.41\,\mathrm{m},$$

and is shown in Figure 4.20. According to the mean max membership method, Equation (4.7), z^* is given by $(6+7)/2 = 6.5$ m, and is shown in Figure 4.21.

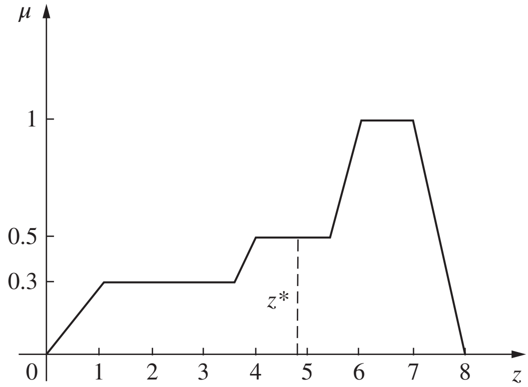

Figure 4.19 The centroid method for finding z^*.

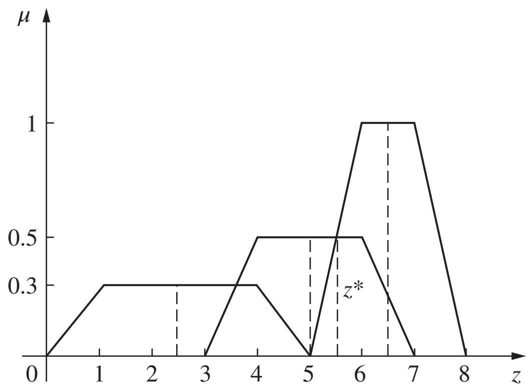

Figure 4.20 The weighted average method for finding z^*.

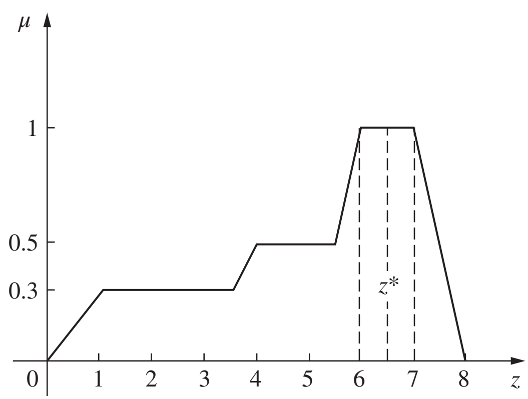

Figure 4.21 The mean max membership method for finding z^*.

Example 4.4

Many products, such as tar, petroleum jelly, and petroleum, are extracted from crude oil. In a newly drilled oil well, three sets of oil samples are taken and tested for their viscosity. The results are given in the form of the three fuzzy sets $\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, and $\underset{\sim}{B}_3$, all defined on a universe of normalized viscosity, as shown in Figures 4.22 to 4.24. Using Equations (4.4) to (4.6), we want to find the most nearly representative viscosity value for all three oil samples, and hence find z^* for the three fuzzy viscosity sets.

To find z^* using the centroid method, we first need to find the logical union of the three fuzzy sets. This is shown in Figure 4.25. Also shown in Figure 4.25 is the result of the max membership method, Equation (4.4). For this method, we see that $\mu_{\underset{\sim}{B}}(z^*)$ has three locations where the membership equals unity. This result is ambiguous, and in this case, the selection of the intermediate point is arbitrary, but it is closer to the centroid of the area shown in Figure 4.25. There could be other compelling reasons to select another value in this case; perhaps, max membership is not a good metric for this problem.

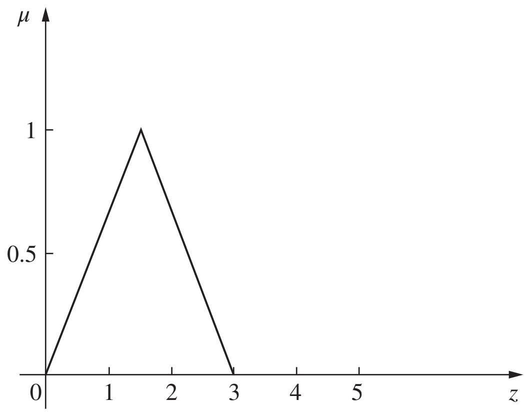

Figure 4.22 Membership in viscosity of oil sample 1, $\underset{\sim}{B}_1$.

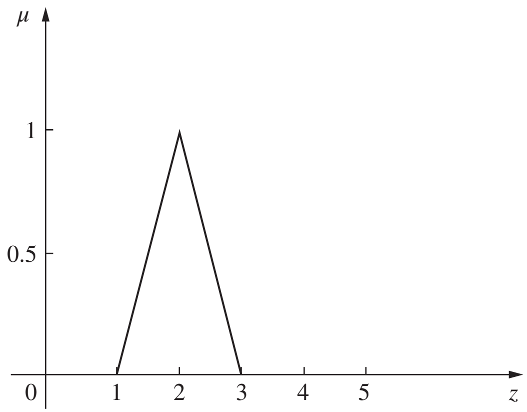

Figure 4.23 Membership in viscosity of oil sample 2, $\underset{\sim}{B}_2$.

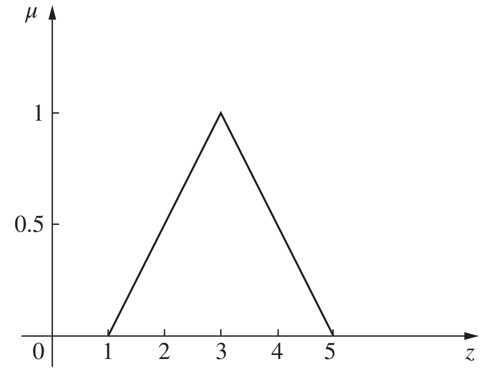

Figure 4.24 Membership in viscosity of oil sample 3, $\underset{\sim}{\mathrm{B}}_3$.

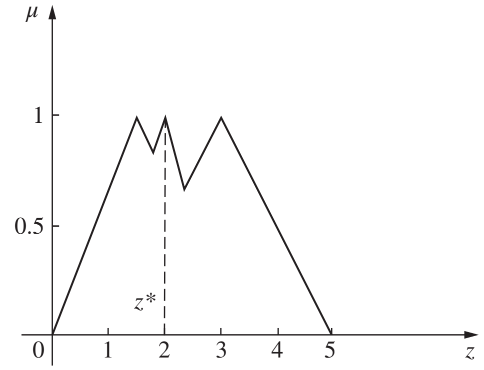

Figure 4.25 Logical union of three fuzzy sets $\underset{\sim}{\mathrm{B}}_1, \underset{\sim}{\mathrm{B}}_2$, and $\underset{\sim}{\mathrm{B}}_3$.

According to the centroid method, Equation (4.5),

$$
\begin{aligned}
z^* &= \frac{\int \mu_{\underset{\sim}{\mathrm{B}}}(z)\, z\, \mathrm{d}z}{\int \mu_{\underset{\sim}{\mathrm{B}}}(z)\, \mathrm{d}z} \\
&= \left[\int_0^{1.5} (0.67z) z\, \mathrm{d}z + \int_{1.5}^{1.8} (2-0.67z) z\, \mathrm{d}z + \int_{1.8}^{2} (z-1) z\, \mathrm{d}z \right. \\
&\quad \left. + \int_2^{2.33} (3-z) z\, \mathrm{d}z. + \int_{2.33}^{3} (0.5z-0.5) z\, \mathrm{d}z + \int_3^5 (2.5-0.5z) z\, \mathrm{d}z \right] \\
&\quad \div \left[\int_0^{1.5} (0.67z)\, \mathrm{d}z + \int_{1.5}^{1.8} (2-0.67z)\, \mathrm{d}z + \int_{1.8}^{2} (z-1)\, \mathrm{d}z + \int_2^{2.33} (3-z)\, \mathrm{d}z \right. \\
&\quad \left. + \int_{2.33}^{3} (0.5z-0.5)\, \mathrm{d}z + \int_3^5 (2.5-0.5z)\, \mathrm{d}z \right] \\
&= 2.5\, \mathrm{m}.
\end{aligned}
$$

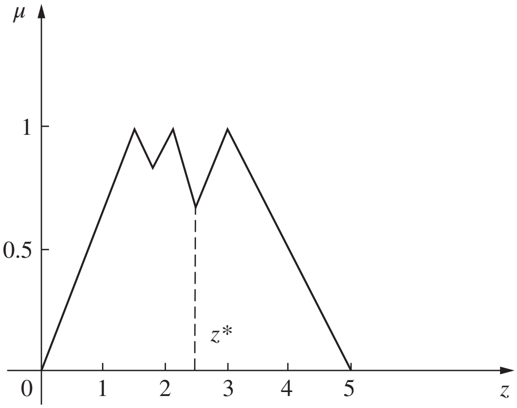

Figure 4.26 Centroid value z^* for three fuzzy oil samples.

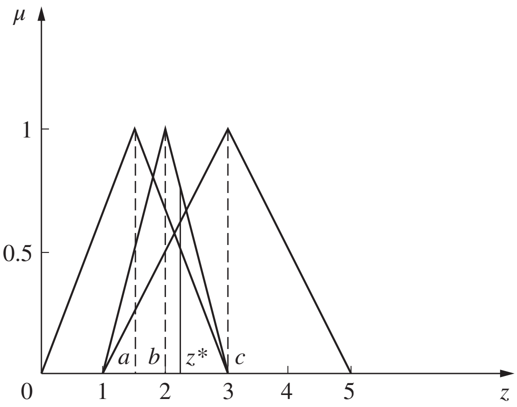

Figure 4.27 Weighted average method for z^*.

The centroid value obtained, z^*, is shown in Figure 4.26.
According to the weighted average method, Equation (4.6),

$$z^* = \frac{(1 \times 1.5) + (1 \times 2) + (1 \times 3)}{1 + 1 + 1} = 2.25\,\mathrm{m},$$

and is shown in Figure 4.27.

There are other popular defuzzification methods in the literature (see Hellendoorn and Thomas, 1993), and three in particular are illustrated in examples in Ross (2010).

Summary

This chapter has introduced the various features and forms of a membership function and the idea of fuzzyifying scalar quantities to make them fuzzy sets. The primary focus of the chapter, however, has been to explain the process of converting from fuzzy membership functions to crisp formats, a process called *defuzzification*. Defuzzification is necessary because, for example, we cannot instruct the voltage going into a machine to increase "slightly," even if this instruction comes from a fuzzy controller; we must alter its voltage by a specific amount. Defuzzification is a natural and necessary process. In fact, there is an analogous form of defuzzification in mathematics where we solve a complicated problem in the complex plane: find the real and imaginary parts of the solution and then *decomplexify* the imaginary solution back to the real numbers space (Bezdek, 1993). There are numerous other methods for defuzzification that have not been presented here. A review of the literature will provide the details on some of these (see, for example, Filev and Yager, 1991; Yager and Filev, 1993; and Ross, 2010).

The following is a natural question to ask: Of the four defuzzification methods presented, which is the best? One obvious answer to the question is that it is context or problem dependent. To answer this question in more depth, Hellendoorn and Thomas (1993) have specified five criteria against which to measure the methods. These criteria will be repeated here for the benefit of the reader who also ponders the question just given in terms of the advantages and disadvantages of the various methods. The first criterion is *continuity*. A small change in the input of a fuzzy process should not produce a large change in the output. Second, a criterion known as *disambiguity* simply points out that a defuzzification method should always result in a unique value for z^*, that is, no ambiguity in the defuzzified value. The third criterion is called *plausibility*. To be plausible, z^* should lie approximately in the middle of the support region of $\underset{\sim}{C}_k$ and have a high degree of membership in $\underset{\sim}{C}_k$. The centroid method, Equation (4.5), does not exhibit plausibility in some situations when it lies in a region of the output that has a low degree of membership. The fourth criterion is that of *computational simplicity*, which suggests that the more time consuming a method is, the less value it should have in a computation system. The height method, Equation (4.4), and the mean max method, Equation (4.7), are computationally faster than the centroid, Equation (4.5), for example. The fifth criterion is called the *weighting method*, which weights the output fuzzy sets. This criterion constitutes the difference between the centroid method, Equation (4.5), and the weighted average method, Equation (4.6). The problem with the fifth criterion is that it is problem dependent, as there is little by which to judge the best weighting method; the weighted average method involves less computation for some problems, but that attribute falls under the fourth criterion, computational simplicity.

As with many issues in fuzzy logic, the method of defuzzification should be assessed in terms of the goodness of the answer in the context of the data available. Other methods that are available purport to be superior to the simple methods presented here (Hellendoorn and Thomas, 1993).

References

Bezdek, J. (1993). Editorial: Fuzzy models—what are they, and why? *IEEE Trans. Fuzzy Syst.* **1**, 1–5.

Filev, D., and Yager, R. (1991). A generalized defuzzification method under BAD distributions. *Int. J. Intell. Syst.*, **6**, 689–697.

Hellendoorn, H., and Thomas, C. (1993). Defuzzification in fuzzy controllers. *J. Intell. Fuzzy Syst.*, **1**, 109–123.

Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications* (Upper Saddle River, NJ: Prentice Hall).

Lee, C. (1990). Fuzzy logic in control systems: Fuzzy logic controller, Parts I and II. *IEEE Trans. Syst. Man Cybern.*, **20**, 404–435.

Ross, T. (1995). *Fuzzy Logic with Engineering Applications* (New York: McGraw-Hill).
Ross, T. (2010). *Fuzzy Logic with Engineering Applications* (Chichester, UK: John Wiley & Sons).
Sugeno, M. (1985). An introductory survey of fuzzy control. *Inf. Sci.*, **36**, 59–83.
Yager, R., and Filev, D. (1993). SLIDE: A simple adaptive defuzzification method. *IEEE Trans. Fuzzy Syst.*, **1**, 69–78.
Zadeh, L. (1965). Fuzzy Sets. *Inform. Contr.* **8**, 338–353.

Problems

4.1 Two fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$, both defined on X, are as follows

$$\underset{\sim}{A} = \left\{\frac{0.1}{x1} + \frac{0.3}{x2} + \frac{0.5}{x3} + \frac{0.7}{x4} + \frac{0.9}{x5}\right\}$$

$$\underset{\sim}{B} = \left\{\frac{0.2}{x1} + \frac{0.4}{x2} + \frac{0.6}{x3} + \frac{0.8}{x4} + \frac{1}{x5}\right\}$$

Express the following λ-cut sets using Zadeh's notation (1965):

a. $\overline{\underset{\sim}{A}}$
b. $\overline{\underset{\sim}{B}}$
c. $\overline{\underset{\sim}{A}}$, for $\lambda = 0.5$
d. $(\underset{\sim}{A} \cup \underset{\sim}{B})$, for $\lambda = 0.5$
e. for $\lambda = 0.2$;
f. $(\overline{\underset{\sim}{A} \cap \underset{\sim}{B}})$, for $\lambda = 0.6$;
g. $(\underset{\sim}{A} \cup \overline{\underset{\sim}{A}})$, for $\lambda = 0.5$

4.2 Show that all λ-cuts of any fuzzy set $\underset{\sim}{A}$ defined in R^n space ($n \geq 1$) are convex if and only if $\mu_{\underset{\sim}{A}}[\lambda r + (1-\lambda)s] \geq \min[\mu_{\underset{\sim}{A}}(r), \mu_{\underset{\sim}{A}}(s)]$, for all r, $s \in R^n$, and all $\lambda \in [0, 1]$ (Klir and Folger, 1988).

4.3 The fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are all defined on the universe $X = [0, 5]$ with the following membership functions: $\mu_{\underset{\sim}{A}}(x) = \dfrac{1}{1+2(x-5)^2}$, $\mu_{\underset{\sim}{B}}(x) = \dfrac{1}{3^x}$

a. Sketch the membership functions
b. Express Zadeh notation for the following λ-cut sets for each of fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ (i) $\lambda = 0.3$ (ii) $\lambda = 0.5$

4.4 Determine the Crisp λ-cut relations for $\lambda = 0.1j$ for $j = 2, 4, 6, 8, 10$ for the following fuzzy relation matrix R

$$\underset{\sim}{R} = \begin{bmatrix} 0.7 & 0.4 & 0.9 & 1 \\ 0.1 & 0.5 & 0.3 & 0.8 \\ 0.3 & 0.9 & 0.8 & 0.2 \\ 0.4 & 0.7 & 0.7 & 0.6 \end{bmatrix}$$

4.5 For the fuzzy relation $\underset{\sim}{R}$ in Problem 3.5(a), sketch (in 3D) the λ-cut relations for the following values of λ:

a. $\lambda = 0+$
b. $\lambda = 0.3$
c. $\lambda = 0.4$
d. $\lambda = 0.9$
e. $\lambda = 1.0$

4.6 Show that any λ-cut relation (for $\lambda > 0$) of a fuzzy tolerance relation results in a crisp tolerance relation.

4.7 Show that any λ-cut relation (for $\lambda > 0$) of a fuzzy equivalence relation results in a crisp equivalence relation.

4.8 When Sulphur reacts with oxygen it gives either sulfur dioxide (SO_2) or sulfur trioxide (SO_3) depending on the number of sulfur and oxygen atoms reacting like:

$S + O_2 \rightarrow SO_2$ (sulfur dioxide)
$2S + 3O_2 \rightarrow 2SO_3$ (sulfur trioxide)

The samples obtained from different proportions are placed on a normalized scale as shown in Figure P4.8(a and b) and are represented by fuzzy sets $\underset{\sim}{A}_1$ and $\underset{\sim}{A}_2$. For the logical union of the membership functions, find the defuzzified quantities for "average" oxygen proportion. Assess,

a. Whether each is applicable and
b. If so, calculate the defuzzified value z^* using the max membership, Centroid method, weighted average method, and mean-max membership method.

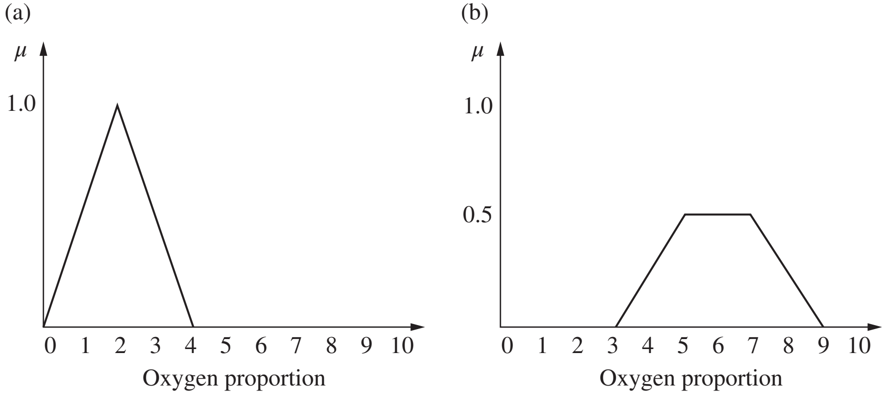

Figure P4.8

4.9 A committee has to review the bidding estimates made by three companies for a contract. The chairperson is interested in the lowest bid, as well as a metric to measure the combined "best" score. For the logical union of the membership functions shown in Figure P4.9(a, b, and c), we want to find the defuzzified quantity. Assess:

a. Whether each is applicable, if so
b. Calculate the defuzzified value, z^* using the four techniques.

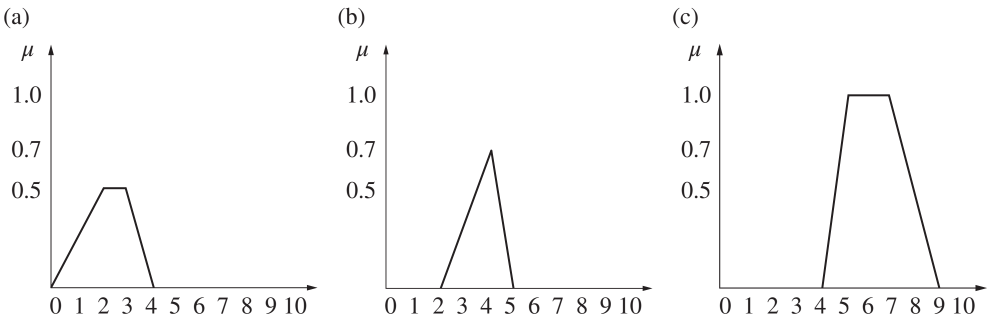

Figure P4.9

4.10 Two different voltmeter devices are recording the overall voltage consumed by the different appliances in a house. It was found that two devices have some slight variation in recording the measurement. To decide the voltage of electric current being used, take the union and employ a defuzzification; V_1 and V_2 are the measurements of two voltmeters shown in Figure P4.10. Assess:

a. whether each of the four methods is applicable, and if so,
b. find the defuzzified value z^* using for those methods.

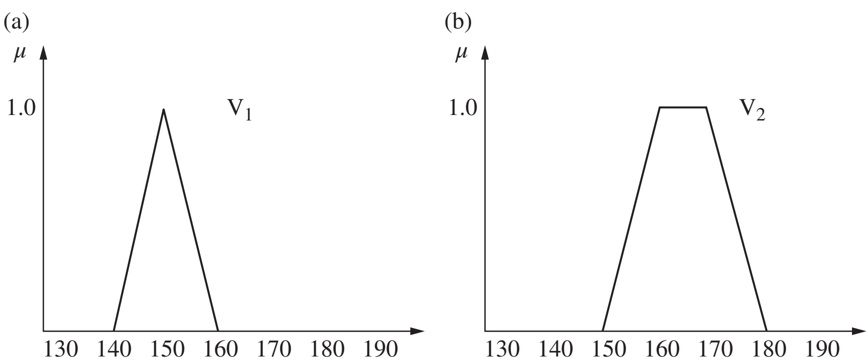

Figure P4.10

5

Logic and Fuzzy Systems

> *"I know what you're thinking about," said Tweedledum; "but it isn't so, no how." "Contrariwise," continued Tweedledee, "if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic."*
>
> Lewis Carroll, *Through the Looking Glass*, 1871

Part I: Logic

Logic is but a small part of the human capacity to reason. Logic can be a means to compel us to infer correct answers, but it cannot by itself be responsible for our creativity or for our ability to remember. In other words, logic can assist us in organizing words to make clear sentences, but it cannot help us determine what sentences to use in various contexts. Consider the passage from the nineteenth-century mathematician Lewis Carroll in his classic *Through the Looking Glass*. How many of us can see the logical context in the discourse of these fictional characters? Logic for humans is a way to quantitatively develop a reasoning process that can be replicated and manipulated with mathematical precepts. The interest in logic is the study of truth in logical propositions; in classical logic this truth is binary, a proposition is either true or false.

From this perspective, fuzzy logic is a method to formalize the human capacity of imprecise reasoning, or as will be shown later in this chapter, approximate reasoning. Such reasoning represents the human ability to reason approximately and judge under uncertainty. In fuzzy logic all truths are partial or approximate. In this sense, this reasoning has also been termed *interpolative reasoning*, where the process of interpolating between the binary extremes of true and false is represented by the ability of fuzzy logic to encapsulate partial truths.

Part I of this chapter introduces the reader to fuzzy logic with a review of classical logic and its operations, logical implications, and certain classical inference mechanisms such as tautologies. The concept of a proposition is introduced as are associated concepts of truth sets, tautologies, and contradictions. The operations of disjunction, conjunction, and negation as well as classical implication and equivalence are introduced; all of these are useful tools to construct compound propositions from single propositions. Operations on propositions are shown to be isomorphic with operations on sets; hence, an algebra of propositions is developed by using the algebra of sets discussed in Chapter 2. Fuzzy logic is then shown to be an extension of classical logic when partial truths are included to extend bivalued logic (true or false) to a multivalued logic (degrees of truth between true and not true).

In Part II of this chapter, we introduce the use of fuzzy sets as a calculus for the interpretation of natural language. Natural language, despite its vagueness and ambiguity, is the vehicle for human communication, and it seems appropriate that a mathematical theory that deals with fuzziness and ambiguity is also the same tool used to express and interpret the linguistic character of our language. The chapter continues with the use of natural language in the expression of a knowledge form known as *rule-based systems*, which shall be referred to generally as *fuzzy systems*. The chapter concludes with a simple graphical interpretation of inference, which is illustrated with some examples.

Classical Logic

In classical logic, a simple proposition P is a linguistic, or declarative, statement contained within a universe of elements, say X, that can be identified as being a collection of elements in X, which are strictly true or strictly false. Hence, a proposition P is a collection of elements, that is, a set, where the truth values for all elements in the set are either all true or all false. The veracity (truth) of an element in the proposition P can be assigned a binary truth value, called T (P), just as an element in a universe is assigned a binary quantity to measure its membership in a particular set. For binary (Boolean) classical logic, T (P) is assigned a value of 1 (truth) or 0 (false). If U is the universe of all propositions, then T is a mapping of the elements, u, in these propositions (sets) to the binary quantities (0, 1), or $T : u \in \mathrm{U} \rightarrow (0, 1)$.

All elements u in the universe U that are true for proposition P are called the truth set of P, denoted T (P). Those elements u in the universe U that are false for proposition P are called the falsity set of P.

In logic we need to postulate the boundary conditions of truth values just as we do for sets; that is, in function-theoretic terms, we need to define the truth value of a universe of discourse. For a universe Y and the null set Ø, we define the following truth values:

$$T(\mathrm{Y}) = 1 \quad \text{and} \quad T(\varnothing) = 0.$$

Now let P and Q be two simple propositions on the same universe of discourse that can be combined using the following five logical connectives

disjunction ($\vee$)
conjunction ($\wedge$)
negation ($-$)
implication ($\rightarrow$)
equivalence ($\leftrightarrow$)

to form logical expressions involving the two simple propositions. These connectives can be used to form new propositions from simple propositions.

The disjunction connective, the *logical or*, is the term used to represent what is commonly referred to as the *inclusive or*. The natural language term *or* and the logical *or* differ in that the former implies exclusion (denoted in the literature as the *exclusive or*; further details are given in this chapter). For example, "soup or salad" on a restaurant menu implies the choice of one or the other option, but not both. The *inclusive or* is the one most often employed in logic; the inclusive or (*logical or* as used here) implies that a compound proposition is true if either of the simple propositions is true or both are true.

The equivalence connective arises from dual implication; that is, for some propositions P and Q, if $P \rightarrow Q$ and $Q \rightarrow P$, then $P \leftrightarrow Q$.

Now define sets A and B from universe X (universe X is isomorphic with universe U), where these sets might represent linguistic ideas or thoughts. A *propositional calculus* (sometimes called the *algebra of propositions*) will exist for the case where proposition P measures the truth of the statement that an element, x, from the universe X is contained in set A and the truth of the statement Q that this element, x, is contained in set B, or more conventionally,

$$\begin{aligned} &\mathrm{P} : \text{truth that } x \in \mathrm{A} \\ &\mathrm{Q} : \text{truth that } x \in \mathrm{B} \end{aligned}$$

where truth is measured in terms of the truth value, that is,

$$\begin{aligned} &\text{if } x \in \mathrm{A}, T(P) = 1; \text{ otherwise, } T(\mathrm{P}) = 0 \\ &\text{if } x \in \mathrm{B}, T(\mathrm{Q}) = 1; \text{ otherwise, } T(\mathrm{Q}) = 0 \end{aligned}$$

or, using the characteristic function to represent truth (1) and falsity (0), the following notation results:

$$\chi_{\mathrm{A}}(x) = \begin{cases} 1, & x \in \mathrm{A} \\ 0, & x \notin \mathrm{A} \end{cases}$$

A notion of *mutual exclusivity* arises in this calculus. For the situation involving two propositions P and Q, where $T(\mathrm{P}) \cap T(\mathrm{Q}) = \emptyset$, we have that the truth of P always implies the falsity of Q and vice versa; hence, P and Q are mutually exclusive propositions.

Example 5.1

Let P be the proposition "The structural beam is an 18WF45" and let Q be the proposition "The structural beam is made of steel." Let X be the universe of structural members comprising girders, beams, and columns; x is an element (beam); A is the set of all wide-flange (WF) beams; and B is the set of all steel beams. Hence,

$$\begin{aligned} &\mathrm{P} : x \text{ is in A} \\ &\mathrm{Q} : x \text{ is in B} \end{aligned}$$

The five logical connectives already defined can be used to create compound propositions, where a compound proposition is defined as a logical proposition formed by logically connecting two or more simple propositions. Just as we are interested in the truth of a simple proposition, classical logic also involves the assessment of the truth of compound propositions. For the case of two simple propositions, the resulting compound propositions are defined next in terms of their binary truth values.

Given a proposition $\mathrm{P} : x \in \mathrm{A}, \overline{\mathrm{P}} : x \notin \mathrm{A}$, we have the following for the logical connectives:

Disjunction

$$\begin{aligned} &\mathrm{P} \vee \mathrm{Q}: x \in \mathrm{A} \text{ or } x \in \mathrm{B}; \\ &\text{hence, } T(\mathrm{P} \vee \mathrm{Q}) = \max(T(P), T(Q)). \end{aligned} \tag{5.1a}$$

Conjunction

$$\begin{aligned} &\mathrm{P} \wedge \mathrm{Q}: x \in \mathrm{A} \text{ and } x \in \mathrm{B}; \\ &\text{hence, } T(\mathrm{P} \wedge \mathrm{Q}) = \min(T(\mathrm{P}), T(\mathrm{Q})). \end{aligned} \tag{5.1b}$$

Negation

$$\text{If } T(\mathrm{P}) = 1, \text{then } T\left(\overline{\mathrm{P}}\right) = 0; \text{ if } T(\mathrm{P}) = 0, \text{ then } T\left(\overline{\mathrm{P}}\right) = 1. \tag{5.1c}$$

Implication

$$\begin{aligned} &(\mathrm{P} \rightarrow \mathrm{Q}): x \notin \mathrm{A} \text{ or } x \in \mathrm{B}; \\ &\text{hence, } T(\mathrm{P} \rightarrow \mathrm{Q}) = T\left(\overline{\mathrm{P}} \cup \mathrm{Q}\right). \end{aligned} \tag{5.1d}$$

Equivalence

$$(\mathrm{P} \leftrightarrow \mathrm{Q}): T(\mathrm{P} \leftrightarrow \mathrm{Q}) = \begin{cases} 1, \text{ for } T(\mathrm{P}) = T(\mathrm{Q}) \\ 0, \text{ for } T(\mathrm{P}) \neq T(Q) \end{cases}. \tag{5.1e}$$

The logical connective *implication*, that is, $\mathrm{P} \rightarrow \mathrm{Q}$ (P implies Q), presented here is also known as the *classical implication* to distinguish it from an alternative form devised in the 1930s by Lukasiewicz, a Polish mathematician, who was first credited with exploring logics other than Aristotelian (classical or binary logic) (Rescher, 1969), and from several other forms (see end of this chapter). In this implication, the proposition P is also referred to as the *hypothesis* or the *antecedent* and the proposition Q is also referred to as the *conclusion* or the *consequent*. The compound proposition $\mathrm{P} \rightarrow \mathrm{Q}$ is true in all cases except where a true antecedent P appears with a false consequent, Q, that is, a true hypothesis cannot imply a false conclusion.

Example 5.2 (Similar to Gill, 1976).
Consider the following four propositions:

1. if $1 + 1 = 2$, then $4 > 0$;
2. if $1 + 1 = 3$, then $4 > 0$;

3. if $1 + 1 = 3$, then $4 < 0$;
4. if $1 + 1 = 2$, then $4 < 0$.

The first three propositions are all true; the fourth is false. In the first two, the conclusion $4 > 0$ is true regardless of the truth of the hypothesis; in the third case both propositions are false, but this does not disprove the implication; finally, in the fourth case, a true hypothesis cannot produce a false conclusion.

Hence, the classical form of the implication is true for all propositions of P and Q except for those propositions that are in both the truth set of P and the false set of Q, that is,

$$T(\mathrm{P} \rightarrow \mathrm{Q}) = \overline{T(\mathrm{P}) \cap T(\overline{\mathrm{Q}})}. \tag{5.2}$$

This classical form of the implication operation requires some explanation. For a proposition P defined on set A and a proposition Q defined on set B, the implication "P implies Q" is equivalent to taking the union of elements in the complement of set A with the elements in the set B (this result can also be derived by using De Morgan's principles on Equation (5.2)). That is, the logical implication is analogous to the set-theoretic form

$$(\mathrm{P} \rightarrow \mathrm{Q}) \equiv (\overline{\mathrm{A}} \cup \mathrm{B} \text{ is true}) \equiv (\text{either "not in A" or "in B"}),$$

so that

$$T(\mathrm{P} \rightarrow \mathrm{Q}) = T(\overline{\mathrm{P}} \vee \mathrm{Q}) = \max(T(\overline{\mathrm{P}}),\ T(\mathrm{Q})). \tag{5.3}$$

This expression is linguistically equivalent to the statement "$\mathrm{P} \rightarrow \mathrm{Q}$ is true" when either "not A" or "B" is true (logical or). Graphically, this implication and the analogous set operation are represented by the Venn diagram in Figure 5.1. As noted in the diagram, the region represented by the difference A|B is the set region where the implication $\mathrm{P} \rightarrow \mathrm{Q}$ is false (the implication "fails"). The shaded region in Figure 5.1 represents the collection of elements in the universe where the implication is true; that is, the set

$$\overline{\mathrm{A|B}} = \overline{\mathrm{A}} \cup \overline{\overline{\mathrm{B}}} = \overline{\mathrm{A} \cap \overline{\mathrm{B}}}.$$

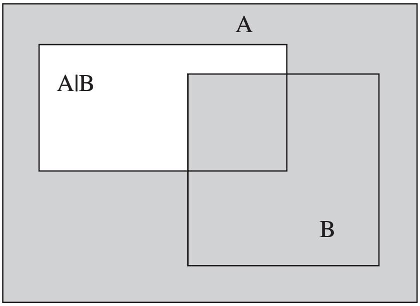

Figure 5.1 Graphical analog of the classical implication operation; gray area is where implication holds.

If x is in A *and* x is not in B, then

$$\mathrm{A} \rightarrow \mathrm{B} \text{ fails} \equiv \mathrm{A}|\mathrm{B} \text{ (difference)}.$$

Now, with two propositions (P and Q) each being able to take on one of two truth values (true or false, 1 or 0), there will be a total of $2^2 = 4$ propositional situations. These situations are illustrated, along with the appropriate truth values, for the propositions P and Q, and the various logical connectives between them are shown in Table 5.1. The values in the last five columns of the table are calculated using the expressions in Equations (5.1) and (5.3). In Table 5.1, T (or 1) denotes true and F (or 0) denotes false.

Suppose the implication operation involves two different universes of discourse: P is a proposition described by set A, which is defined on universe X, and Q is a proposition described by set B, which is defined on universe Y. Then, the implication $\mathrm{P} \rightarrow \mathrm{Q}$ can be represented in set-theoretic terms by the relation R, where R is defined as

$$\begin{aligned} &\mathrm{R} = (\mathrm{A} \times \mathrm{B}) \cup \left(\overline{\mathrm{A}} \times \mathrm{Y}\right) \equiv \text{IF A, THEN B} \\ &\text{IF } x \in \mathrm{A} \text{ where } x \in \mathrm{X} \text{ and } \mathrm{A} \subset \mathrm{X} \\ &\text{THEN y} \in \mathrm{B} \text{ where } y \in \mathrm{Y} \text{ and } \mathrm{B} \subset \mathrm{Y} \end{aligned} \tag{5.4}$$

This implication, Equation (5.4), is also equivalent to the linguistic rule form, IF A, THEN B. The graphic shown in Figure 5.2 represents the space of the Cartesian product $\mathrm{X} \times \mathrm{Y}$, showing

Table 5.1 Truth table for various compound propositions.

P	Q	$\overline{\mathbf{P}}$	$\mathbf{P} \vee \mathbf{Q}$	$\mathbf{P} \wedge \mathbf{Q}$	$\mathbf{P} \rightarrow \mathbf{Q}$	$\mathbf{P} \leftrightarrow \mathbf{Q}$
T (1)	T (1)	F (0)	T (1)	T (1)	T (1)	T (1)
T (1)	F (0)	F (0)	T (1)	F (0)	F (0)	F (0)
F (0)	T (1)	T (1)	T (1)	F (0)	T (1)	F (0)
F (0)	F (0)	T (1)	F (0)	F (0)	T (1)	T (1)

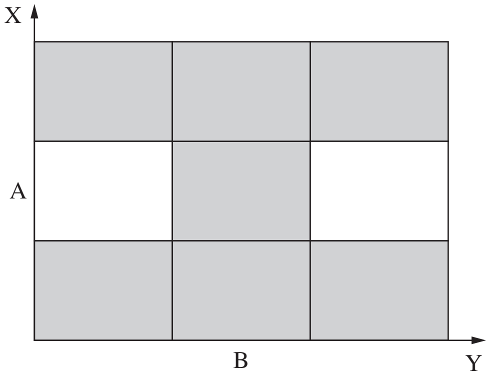

Figure 5.2 The Cartesian space showing the implication IF A, THEN B.

typical sets A and B; superposed on this space is the set-theoretic equivalent of the implication. That is,

$$\mathrm{P} \rightarrow \mathrm{Q}:\ \text{IF } x \in \mathrm{A},\ \text{THEN } \mathrm{y} \in \mathrm{B} \text{ or } \mathrm{P} \rightarrow \mathrm{Q} \equiv \overline{\mathrm{A}} \cup \mathrm{B}.$$

The shaded regions of the compound Venn diagram in Figure 5.2 represent the truth domain of the implication, IF A, THEN B ($\mathrm{P} \rightarrow \mathrm{Q}$).

Another compound proposition in linguistic rule form is the expression

$$\text{IF A, THEN B, ELSE C.}$$

Linguistically, this compound proposition could be expressed as

$$\text{IF A, THEN B and IF } \overline{\mathrm{A}}\text{, THEN C.}$$

In classical logic, this rule has the form

$$\begin{aligned} &(\mathrm{P} \rightarrow \mathrm{Q}) \wedge (\overline{\mathrm{P}} \rightarrow \mathrm{S}); \\ &\mathrm{P}:\ x \in \mathrm{A},\ \mathrm{A} \subset \mathrm{X}, \\ &\mathrm{Q}:\ y \in \mathrm{B},\ \mathrm{B} \subset \mathrm{Y}, \\ &\mathrm{S}:\ y \in \mathrm{C},\ \mathrm{C} \subset \mathrm{Y}. \end{aligned} \tag{5.5}$$

The set-theoretic equivalent of this compound proposition is given as

$$\text{IF A, THEN B, ELSE C} \equiv (\mathrm{A} \times \mathrm{B}) \cup (\overline{\mathrm{A}} \times \mathrm{C}) = \mathrm{R} \text{ is a relation on } \mathrm{X} \times \mathrm{Y}. \tag{5.6}$$

Then the shaded region in Figure 5.3 represents the truth domain for this compound proposition for the particular case where $\mathrm{B} \cap \mathrm{C} = \varnothing$.

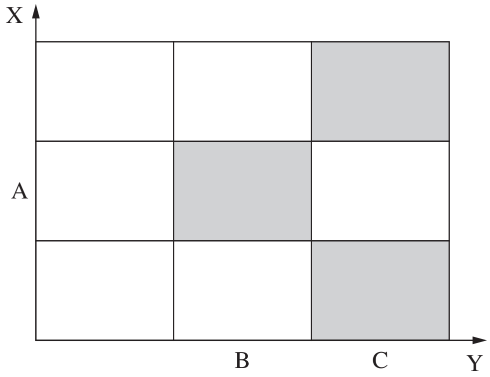

Figure 5.3 Truth domain for IF A, THEN B, ELSE C.

Tautologies

In classical logic it is useful to consider compound propositions that are always true, irrespective of the truth values of the individual simple propositions. Classical logical compound propositions with this property are called *tautologies*. Tautologies are useful for reasoning, for proving theorems, and for making deductive inferences. So, if a compound proposition can be expressed in the form of a tautology, the truth value of that compound proposition is known to be true. Inference schemes in expert systems often employ tautologies because tautologies are formulas that are true on logical grounds alone. For example, if A is the set of all prime numbers ($A_1 = 1$, $A_2 = 2$, $A_3 = 3$, $A_4 = 5, \ldots$) on the real line universe, X, then the proposition "A_i is not divisible by 6" is a tautology.

One tautology, known as *modus ponens* deduction, is a common inference scheme used in forward-chaining rule-based expert systems. It is an operation whose task is to find the truth value of a consequent in a production rule, given the truth value of the antecedent in the rule. *Modus ponens* deduction concludes that given two propositions, P and $P \rightarrow Q$, if both of which are true, then the truth of the simple proposition Q is automatically inferred. Another useful tautology is the *modus tollens* inference, which is used in backward-chaining expert systems. In *modus tollens*, an implication between two propositions is combined with a second proposition and both are used to imply a third proposition. Some common tautologies are as follows:

$$\overline{\mathrm{B}} \cup \mathrm{B} \leftrightarrow \mathrm{X}.$$

$$\mathrm{A} \cup \mathrm{X};\ \ \overline{\mathrm{A}} \cup \mathrm{X} \leftrightarrow \mathrm{X}.$$

$$(\mathrm{A} \wedge (\mathrm{A} \rightarrow \mathrm{B})) \rightarrow \mathrm{B} \quad (\textit{modus ponens}). \tag{5.7}$$

$$\left(\overline{\mathrm{B}} \wedge (\mathrm{A} \rightarrow \mathrm{B})\right) \rightarrow \overline{A} \quad (\textit{modus tollens}). \tag{5.8}$$

A simple proof of the truth value of the *modus ponens* deduction is provided here, along with the various properties for each step of the proof, for purposes of illustrating the utility of a tautology in classical reasoning.

Proof

$(\mathrm{A} \wedge (\mathrm{A} \rightarrow \mathrm{B}) \rightarrow \mathrm{B}$	
$\left(\mathrm{A} \wedge \left(\overline{\mathrm{A}} \cup \mathrm{B}\right)\right) \rightarrow \mathrm{B}$	*Implication*
$\left(\left(\mathrm{A} \wedge \overline{\mathrm{A}}\right) \cup (\mathrm{A} \wedge \mathrm{B})\right) \rightarrow \mathrm{B}$	*Distributivity*
$(\emptyset \cup (\mathrm{A} \wedge \mathrm{B})) \rightarrow \mathrm{B}$	*Excluded middle axioms*
$(\mathrm{A} \wedge \mathrm{B}) \rightarrow \mathrm{B}$	*Identity*
$\left(\overline{\mathrm{A} \wedge \mathrm{B}}\right) \cup \mathrm{B}$	*Implication*
$\left(\overline{\mathrm{A}} \vee \overline{\mathrm{B}}\right) \cup \mathrm{B}$	*De Morgan's principles*
$\overline{\mathrm{A}} \vee \left(\overline{\mathrm{B}} \cup \mathrm{B}\right)$	*Associativity*
$\overline{\mathrm{A}} \cup \mathrm{X}$	*Excluded middle axioms*
$\mathrm{X} \Rightarrow T(\mathrm{X}) = 1$	*Identity; QED*

Table 5.2 Truth table (*modus ponens*).

A	B	$\mathbf{A} \rightarrow \mathbf{B}$	$(\mathbf{A} \wedge (\mathbf{A} \rightarrow \mathbf{B}))$	$(\mathbf{A} \wedge (\mathbf{A} \rightarrow \mathbf{B})) \rightarrow \mathbf{B}$	
0	0	1	0	1	
0	1	1	0	1	Tautology
1	0	0	0	1	
1	1	1	1	1	

Table 5.3 Truth table (*modus tollens*).

A	B	$\overline{\mathbf{A}}$	$\overline{\mathbf{B}}$	$\mathbf{A} \rightarrow \mathbf{B}$	$(\overline{\mathbf{B}} \wedge (\mathbf{A} \rightarrow \mathbf{B}))$	$(\overline{\mathbf{B}} \wedge (\mathbf{A} \rightarrow \mathbf{B})) \rightarrow \overline{\mathbf{A}}$	
0	0	1	1	1	1	1	
0	1	1	0	1	0	1	Tautology
1	0	0	1	0	0	1	
1	1	0	0	1	0	1	

A simpler manifestation of the truth value of this tautology is shown in Table 5.2 in truth table form, where a column of all ones for the result shows a tautology.

Similarly, a simple proof of the truth value of the *modus tollens* inference is listed here.

Proof

$$\left(\overline{\mathrm{B}} \wedge (\mathrm{A} \rightarrow \mathrm{B})\right) \rightarrow \overline{\mathrm{A}}$$

$$\left(\overline{\mathrm{B}} \wedge \left(\overline{\mathrm{A}} \cup \mathrm{B}\right)\right) \rightarrow \overline{\mathrm{A}}$$

$$\left(\left(\overline{\mathrm{B}} \wedge \overline{\mathrm{A}}\right) \cup \left(\overline{\mathrm{B}} \wedge \mathrm{B}\right)\right) \rightarrow \overline{\mathrm{A}}$$

$$\left(\left(\overline{\mathrm{B}} \wedge \overline{\mathrm{A}}\right) \cup \varnothing\right) \rightarrow \overline{\mathrm{A}}$$

$$\left(\overline{\mathrm{B}} \wedge \overline{\mathrm{A}}\right) \rightarrow \overline{\mathrm{A}}$$

$$\left(\overline{\overline{\mathrm{B}} \wedge \overline{\mathrm{A}}}\right) \cup \overline{\mathrm{A}}$$

$$\left(\overline{\overline{\mathrm{B}}} \vee \overline{\overline{\mathrm{A}}}\right) \cup \overline{\mathrm{A}}$$

$$\mathrm{B} \cup \left(\mathrm{A} \cup \overline{\mathrm{A}}\right)$$

$$\mathrm{B} \cup \mathrm{X} = \mathrm{X} \Rightarrow T(\mathrm{X}) = 1 \qquad \text{QED}$$

The truth table form of this result is shown in Table 5.3.

Contradictions

Compound propositions that are always false, regardless of the truth value of the individual simple propositions constituting the compound proposition, are called *contradictions*. For example, if A is the set of all prime numbers ($A_1 = 1$, $A_2 = 2$, $A_3 = 3$, $A_4 = 5$, . . .) on the real line universe, X, then the proposition "A_i is a multiple of 4" is a contradiction. Some simple contradictions are listed here:

$$\overline{B} \cap B,$$
$$A \cap \varnothing; \quad \overline{A} \cap \varnothing.$$

Equivalence

As mentioned, propositions P and Q are equivalent, that is, $P \leftrightarrow Q$, is true only when both P and Q are true or when both P and Q are false. For example, the propositions P "triangle is equilateral" and Q "triangle is equiangular" are equivalent because they are either both true or both false for some triangle. This condition of equivalence is shown in Figure 5.4, where the shaded region is the region of equivalence.

It can be easily proved that the statement $P \leftrightarrow Q$ is a tautology if P is identical to Q, that is, if and only if $T(P) = T(Q)$.

Example 5.3

Suppose we consider the universe of positive integers, $X = \{1 \le n \le 8\}$. Let P = "n is an even number" and let Q = "$(3 \le n \le 7) \wedge (n \ne 6)$." Then $T(P) = \{2, 4, 6, 8\}$ and $T(Q) = \{3, 4, 5, 7\}$. The equivalence $P \leftrightarrow Q$ has the truth set

$$T(P \leftrightarrow Q) = (T(P) \cap T(Q)) \cup \left(\overline{T(P)} \cap \overline{T(Q)}\right) = \{4\} \cup \{1\} = \{1,4\}$$

One can see that "1 is an even number" and "$(3 \le 1 \le 7) \wedge (1 \ne 6)$" are both false, and "4 is an even number" and "$(3 \le 4 \le 7) \wedge (4 \ne 6)$" are both true.

Example 5.4

Prove that $P \leftrightarrow Q$ if P = "n is an integer power of 2 less than 7 and greater than zero" and Q = "$n^2 - 6n + 8 = 0$." Since $T(P) = \{2, 4\}$ and $T(Q) = \{2, 4\}$, it follows that $P \leftrightarrow Q$ is an equivalence.

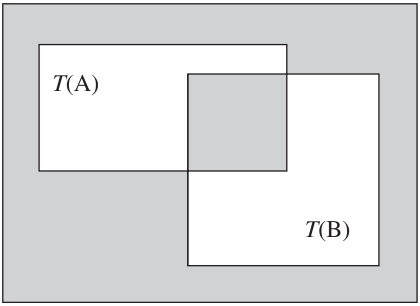

Figure 5.4 Venn diagram for equivalence (darkened areas), that is, for $T(A \leftrightarrow B)$.

Suppose a proposition R has the form $P \rightarrow Q$. Then the proposition $\overline{Q} \rightarrow \overline{P}$ is called the *contrapositive* of R; the proposition $Q \rightarrow P$ is called the *converse* of R; and the proposition $\overline{P} \rightarrow \overline{Q}$ is called the *inverse* of R. Interesting properties of these propositions can be shown (see Problem 5.3 at the end of this chapter).

The *dual* of a compound proposition *that does not involve implication* is the same proposition with false (0) replacing true (1) (i.e., a set being replaced by its complement), true replacing false, conjunction ($\wedge$) replacing disjunction ($\vee$), and disjunction replacing conjunction. If a proposition is true, then its *dual* is also true (Problems 5.4 and 5.5).

Exclusive Or and Exclusive Nor

Two more interesting compound propositions are worthy of discussion. These are the *exclusive or* and the *exclusive nor*. The exclusive or is of interest because it arises in many situations involving natural language and human reasoning. For example, when you are going to travel by plane or boat to some destination, the implication is that you can travel by air or sea, but not both, that is, one or the other. This situation involves the exclusive or; it does not involve the intersection, as does the logical or (union in Equation (2.1) and Figure 2.2 and disjunction in Equation (5.1a)). For two propositions, P and Q, the exclusive or, denoted here as *XOR*, is given in Table 5.4 and Figure 5.5.

The exclusive nor is the complement of the exclusive or (Mano, 1988). A look at its truth table, Table 5.5, shows that it is an equivalence operation, that is,

Table 5.4 Truth table for exclusive or, *XOR*.

P	Q	P *XOR* Q
1	1	0
1	0	1
0	1	1
0	0	0

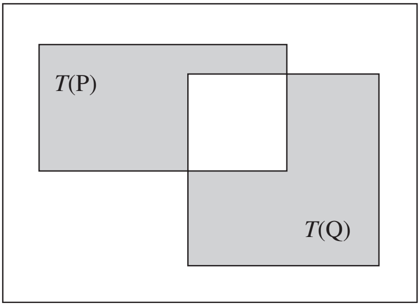

Figure 5.5 *Exclusive or* shown in gray areas.

Table 5.5 Truth table for exclusive nor.

P	Q	$\overline{\text{P } XOR \text{ Q}}$
1	1	1
1	0	0
0	1	0
0	0	1

$$\overline{\text{P } XOR \text{ Q}} \leftrightarrow (\text{P} \leftrightarrow \text{Q}),$$

and, hence, it is graphically equivalent to the Venn diagram shown in Figure 5.4.

Logical Proofs

Logic involves the use of inference in everyday life, as well as in mathematics. In the latter, we often want to prove theorems to form foundations for solution procedures. In natural language, if we are given some hypotheses, it is often useful to make certain conclusions from them, the so-called process of inference (inferring new facts from established facts). In the terminology we have been using, we want to know if the proposition $(\text{P}_1 \wedge \text{P}_2 \wedge \cdots \wedge \text{P}_n) \rightarrow \text{Q}$ is true. That is, is the statement a tautology?

The process works as follows. First, the linguistic statement (compound proposition) is made. Second, the statement is decomposed into its respective single propositions. Third, the statement is expressed algebraically with all pertinent logical connectives in place. Fourth, a truth table is used to establish the veracity of the statement.

Example 5.5

Hypotheses: Engineers are mathematicians. Logical thinkers do not believe in magic. Mathematicians are logical thinkers.
Conclusion: Engineers do not believe in magic.

Let us decompose this information into individual propositions.

$P:$ a person is an engineer.

$Q:$ a person is a mathematician.

$R:$ a person is a logical thinker.

$S:$ a person believes in magic.

The statements can now be expressed as algebraic propositions as

$$\left((\text{P} \rightarrow \text{Q}) \wedge \left(\text{R} \rightarrow \overline{\text{S}}\right) \wedge (\text{Q} \rightarrow \text{R})\right) \rightarrow \left(\text{P} \rightarrow \overline{\text{S}}\right).$$

It can be shown that this compound proposition is a tautology (Problem 5.6).

Sometimes it might be difficult to prove a proposition by a direct proof (i.e., verify that it is true), so an alternative is to use an indirect proof. For example, the popular *proof by contradiction* (*reductio ad absurdum*) exploits the fact that $P \to Q$ is true if and only if $P \wedge \overline{Q}$ is false. Hence, if we want to prove that the compound statement $(P_1 \wedge P_2 \wedge \cdots \wedge P_n) \to Q$ is a tautology, we can alternatively show that the alternative statement $P_1 \wedge P_2 \wedge \cdots \wedge P_n \wedge \overline{Q}$ is a contradiction.

Example 5.6

Hypotheses: If an arch-dam fails, the failure is the result of a poor subgrade. An arch-dam fails.
Conclusion: The arch-dam failed because of a poor subgrade.

This information can be shown to be algebraically equivalent to the expression

$$((P \to Q) \wedge P) \to Q.$$

To prove this by contradiction, we need to show that the algebraic expression

$$\left((P \to Q) \wedge P \wedge \overline{Q}\right)$$

is a contradiction. We can do this by constructing the truth table in Table 5.6. Recall that a contradiction is indicated when the last column of a truth table is filled with zeros.

Deductive Inferences

The *modus ponens* deduction is used as a tool for making inferences in rule-based systems. A typical if–then rule is used to determine whether an antecedent (cause or action) infers a consequent (effect or reaction). Suppose we have a rule of the form IF A, THEN B, where A is a set defined on universe X and B is a set defined on universe Y. As discussed before, this rule can be translated into a relation between sets A and B; that is, recalling Equation (5.4), $R = (A \times B) \cup (\overline{A} \times Y)$. Now suppose a new antecedent, say A', is known. Can we use *modus ponens* deduction, Equation (5.7), to infer a new consequent, say B', resulting from the new antecedent? That is, can we deduce, in rule form, IF A', THEN B'? The answer, of course, is yes, through the use of the composition operation (defined initially in Chapter 3). Because "A implies B" is defined

Table 5.6 Truth table for dam failure problem.

P	Q	$\overline{P}$	$\overline{Q}$	$\overline{P} \vee Q$	$(\overline{P} \vee Q) \wedge P \wedge \overline{Q}$
0	0	1	1	1	0
0	1	1	0	1	0
1	0	0	1	0	0
1	1	0	0	1	0

on the Cartesian space $X \times Y$, B' can be found through the following set-theoretic formulation, again from Equation (5.4):

$$B' = A' \circ R = A' \circ \left((A \times B) \cup \left(\overline{A} \times Y \right) \right),$$

where the symbol $^\circ$ denotes the composition operation. *Modus ponens* deduction can also be used for the compound rule IF A, THEN B, ELSE C, where this compound rule is equivalent to the relation defined in Equation (5.6) as $R = (A \times B) \cup \left(\overline{A} \times C \right)$. For this compound rule, if we define another antecedent A', the following possibilities exist, depending on whether (1) A' is fully contained in the original antecedent A, (2) A' is contained only in the complement of A, or (3) A' and A overlap to some extent as described next:

$$\begin{aligned}
&\text{IF } A' \subset A, \text{ THEN } y = B \\
&\text{IF } A' \subset \overline{A}, \text{ THEN } y = C \\
&\text{IF } A' \cap A \neq \varnothing,\ A' \cap \overline{A} \neq \varnothing, \text{ THEN} y = B \cup C
\end{aligned} \quad .$$

The rule IF A, THEN B (proposition P is defined on set A in universe X, and proposition Q is defined on set B in universe Y), that is, $(P \rightarrow Q) = R = (A \times B) \cup \left(\overline{A} \times Y \right)$, is then defined in function-theoretic terms as

$$\chi_R(x,y) = \max[(\chi_A(x) \wedge \chi_B(y)), ((1 - \chi_A(x)) \wedge 1], \tag{5.9}$$

where $\chi()$ is the characteristic function as defined before.

Example 5.7

Suppose we have two universes of discourse for a heat exchanger problem described by the following collection of elements: $X = \{1, 2, 3, 4\}$ and $Y = \{1, 2, 3, 4, 5, 6\}$. Suppose X is a universe of normalized temperatures and Y is a universe of normalized pressures. Define crisp set A on universe X and crisp set B on universe Y as follows: $A = \{2, 3\}$ and $B = \{3, 4\}$. The deductive inference IF A, THEN B (i.e., IF temperature is A, THEN pressure is B) will yield a matrix describing the membership values of the relation R, that is, $\chi_R(x, y)$, through the use of Equation (5.9). That is, the matrix R represents the rule IF A, THEN B as a matrix of characteristic (crisp membership) values.

Crisp sets A and B can be written using Zadeh's notation,

$$A = \left\{ \frac{0}{1} + \frac{1}{2} + \frac{1}{3} + \frac{0}{4} \right\}.$$

$$B = \left\{ \frac{0}{1} + \frac{0}{2} + \frac{1}{3} + \frac{1}{4} + \frac{0}{5} + \frac{0}{6} \right\}.$$

If we treat set A as a column vector and set B as a row vector, the following matrix results from the Cartesian product of $A \times B$, using Equation (3.16):

$$A \times B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The Cartesian product $\overline{A} \times Y$ can be determined using Equation (3.16) by arranging $\overline{A}$ as a column vector and the universe Y as a row vector (sets $\overline{A}$ and Y can be written using Zadeh's notation):

$$\overline{A} = \left\{ \frac{1}{1} + \frac{0}{2} + \frac{0}{3} + \frac{1}{4} \right\}.$$

$$Y = \left\{ \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \right\}.$$

$$\overline{A} \times Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then, the full relation R describing the implication IF A, THEN B is the maximum of the two matrices $A \times B$ and $\overline{A} \times Y$ or, using Equation (5.9),

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

The compound rule IF A, THEN B, ELSE C can also be defined in terms of a matrix relation as $R = (A \times B) \cup (\overline{A} \times C)$ with sets or, with propositions, $(P \rightarrow Q) \wedge (\overline{P} \rightarrow S)$, as given by Equations (5.5) and (5.6), where the membership function is determined as

$$\chi_R(x,y) = \max[(\chi_A(x) \wedge \chi_B(y)),\ ((1-\chi_A(x)) \wedge \chi_C(y))]. \tag{5.10}$$

Example 5.8

Continuing with the previous heat exchanger example, suppose we define a crisp set C on the universe of normalized temperatures Y as $C = \{5, 6\}$, or, using Zadeh's notation,

$$C = \left\{ \frac{0}{1} + \frac{0}{2} + \frac{0}{3} + \frac{0}{4} + \frac{1}{5} + \frac{1}{6} \right\},$$

The deductive inference IF A, THEN B, ELSE C (i.e., IF pressure is A, THEN temperature is B, ELSE temperature is C) will yield a relational matrix R, with characteristic values $\chi_R(x, y)$ obtained using Equation (5.10). The first half of the expression in Equation (5.10) (i.e., $A \times B$) has already been determined in the previous example. The Cartesian product $\overline{A} \times C$ can be determined using Equation (3.16) by arranging the set $\overline{A}$ as a column vector and the set C as a row vector (see set $\overline{A}$ in Example 5.7), or

$$\overline{A} \times C = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

Then, the full relation R describing the implication IF A, THEN B, ELSE C is the maximum of the two matrices $A \times B$ and $\overline{A} \times C$ (Equation (5.10)):

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

Fuzzy Logic

The restriction of classical propositional calculus to a two-valued logic has created many interesting paradoxes over the ages. For example, the Barber of Seville is a classic paradox (also termed *Russell's barber*). In the small Spanish town of Seville, there is a rule that all and only those men who do not shave themselves are shaved by the barber. Who shaves the barber? Another example comes from ancient Greece. Does the liar from Crete lie when he claims, "All Cretians are liars?" If he is telling the truth, his statement is false. But, if his statement is false, he is not telling the truth. A simpler form of this paradox is the two-word proposition, "I lie." The statement cannot be both true and false.

Returning to the Barber of Seville, we conclude that the only way for this paradox (or any classic paradox for that matter) to work is if the statement is both true and false simultaneously. This can be shown using set notation (Kosko, 1992). Let S be the proposition that the barber shaves himself and $\overline{S}$ (not S) that he does not. Then because $S \to \overline{S}$ (S implies not S), and $\overline{S} \to S$, the two propositions are logically equivalent: $S \leftrightarrow \overline{S}$. Equivalent propositions have the same truth value; hence,

$$T(S) = T(\overline{S}) = 1 - T(S),$$

which yields the expression

$$T(S) = \frac{1}{2}.$$

As seen, paradoxes reduce to half-truths (or half-falsities) mathematically. In classical binary (bivalued) logic, however, such conditions are not allowed, that is, only $T(S) = 1$ or 0 is valid;

this is a manifestation of the constraints placed on classical logic by the excluded middle axioms.

A subtler form of paradox can also be addressed by a multivalued logic. Consider the paradoxes represented by the classical *sorites* (literally, a heap of syllogisms); for example, the case of a liter-full glass of water. Often this example is called the Optimist's conclusion (is the glass half-full or half-empty when the volume is at 500 milliliters?). Is the liter-full glass still full if we remove 1 mL of water? Is the glass still full if we remove 2 mL of water, 3, 4, or 100 mL? If we continue to answer yes, then eventually we will have removed all the water, and an empty glass will still be characterized as full! At what point did the liter-full glass of water become empty? Perhaps at 500 mL full? Unfortunately, no single milliliter of liquid provides for a transition between full and empty. This transition is gradual, so that as each milliliter of water is removed, the truth value of the glass being full gradually diminishes from a value of 1 at 1000 ml to 0 at 0 mL. Hence, for many problems we have need for a multivalued logic other than the classic binary logic that is so prevalent today.

A relatively recent debate involving similar ideas to those in paradoxes stems from a paper by psychologists Osherson and Smith (1981), in which they claim (incorrectly) that fuzzy set theory is not expressive enough to represent strong intuitionistic concepts. This idea can be described as the logically empty and logically universal concepts. The authors argued that the concept *apple that is not an apple is logically empty*, and that the concept *fruit that either is or is not an apple is logically universal*. Never mind that the argument is flawed, recognizing an apple is a crisp concept, not fuzzy. Perhaps the authors should have asked "fruit that is not a vegetable . . . " because in this case the concepts "fruit" and "vegetable" can be fuzzy. For example, tomatoes and cucumbers are technically fruits, but most people (and cooks) label them vegetables. The concepts argued by Osherson and Smith are correct for classical logic; the logically empty idea and the logically universal idea are the axiom of contradiction and the axiom of the excluded middle, respectively. The authors argued that fuzzy logic also should adhere to these axioms to correctly represent concepts in natural language but, of course, there is a compelling reason why they should not. Several authorities have disputed this argument (Belohlavek, Klir, Lewis, and Way, 2002). Although the *standard fuzzy operations* (e.g., *min and max*) do not follow the excluded middle axioms, there are other operations for intersection (t-norms), union (t-conorms), and complement that do conform to these axioms if such a confirmation is required by empirical evidence. Belohlavek and colleagues (2002) suggest that fuzzy set theory has an expressive power that is far beyond that of classical set theory. The fuzzy set operations for complement, intersection, and union are not unique; there are an infinite number from which to choose. Morever, fuzzy set theory has aggregation operations that have no counterparts in classical set theory. Based on this plethora of representations, fuzzy sets are capable of emulating many relationships among natural categories revealed by empirical data or that is required by human intuition. When it is suggested that this makes fuzzy set theory too flexible, the reality is that this flexibility provides for a powerful representational formalism. In another paper, the original authors, Smith and Osherson (1984) finally conclude after their own additional study that "fuzzy set theory is of interest to cognitive scientists because it offers a calculus for combining prototype concepts."

A fuzzy logic proposition, $\underset{\sim}{P}$ is a statement involving some concept without clearly defined boundaries. Linguistic statements that tend to express subjective ideas and that can be interpreted slightly differently by various individuals typically involve fuzzy propositions. Most natural language is fuzzy, in that it involves vague and imprecise terms. Statements describing a person's height or weight or assessments of people's preferences about colors or menus can be used as examples of fuzzy propositions. The truth value assigned to $\underset{\sim}{P}$ can be any value on the interval [0, 1]. The assignment of the truth value to a proposition is actually a mapping from the interval [0, 1] to the universe U of truth values, T, as indicated in Equation (5.11):

$$T : u \in \mathrm{U} \rightarrow (0,1). \tag{5.11}$$

As in classical binary logic, we assign a logical proposition to a set in the universe of discourse. Fuzzy propositions are assigned to fuzzy sets. Suppose proposition $\underset{\sim}{P}$ is assigned to fuzzy set $\underset{\sim}{A}$; then, the truth value of a proposition, denoted $T(\underset{\sim}{P})$, is given by

$$T(\underset{\sim}{P}) = \mu_{\underset{\sim}{A}}(x), \quad \text{where } 0 \le \mu_{\underset{\sim}{A}} \le 1. \tag{5.12}$$

Equation (5.12) indicates that the degree of truth for the proposition $\underset{\sim}{P}$: $x \in \underset{\sim}{A}$ is equal to the membership grade of x in the fuzzy set $\underset{\sim}{A}$.

The logical connectives of negation, disjunction, conjunction, and implication are also defined for a fuzzy logic. These connectives are given in Equations (5.13) to (5.16) for two simple propositions: proposition $\underset{\sim}{P}$ defined on fuzzy set $\underset{\sim}{A}$ and proposition $\underset{\sim}{Q}$ defined on fuzzy set $\underset{\sim}{B}$.

Negation

$$T(\overline{\underset{\sim}{P}}) = 1 - T(\underset{\sim}{P}). \tag{5.13}$$

Disjunction

$$\underset{\sim}{P} \vee \underset{\sim}{Q} : x \text{ is } \underset{\sim}{A} \text{ or } \underset{\sim}{B} \quad T(\underset{\sim}{P} \vee \underset{\sim}{Q}) = \max(T(\underset{\sim}{P}), T(\underset{\sim}{Q})). \tag{5.14}$$

Conjunction

$$\underset{\sim}{P} \wedge \underset{\sim}{Q} : x \text{ is } \underset{\sim}{A} \text{ and } \underset{\sim}{B} \quad T(\underset{\sim}{P} \wedge \underset{\sim}{Q}) = \min(T(\underset{\sim}{P}), T(\underset{\sim}{Q})). \tag{5.15}$$

Implication (Zadeh, 1973)

$$\begin{aligned} &\underset{\sim}{P} \rightarrow \underset{\sim}{Q} : x \text{ is } \underset{\sim}{A}, \text{ then } x \text{ is } \underset{\sim}{B} \\ &T(\underset{\sim}{P} \rightarrow \underset{\sim}{Q}) = T(\overline{\underset{\sim}{P}} \vee \underset{\sim}{Q}) = \max(T(\overline{\underset{\sim}{P}}), T(\underset{\sim}{Q})). \end{aligned} \tag{5.16}$$

As before in binary logic, the implication connective can be modeled in rule-based form; $\underset{\sim}{P} \rightarrow \underset{\sim}{Q}$ is IF x is $\underset{\sim}{A}$, THEN y is $\underset{\sim}{B}$ and it is equivalent to the fuzzy relation $\underset{\sim}{R} = (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\overline{\underset{\sim}{A}} \times \mathrm{Y})$ (recall Equation (5.4)), just as it is in classical logic. The membership function of $\underset{\sim}{R}$ is expressed by the following formula:

$$\mu_{\underset{\sim}{R}}(x,y) = \max[(\mu_{\underset{\sim}{A}}(x) \wedge \mu_{\underset{\sim}{B}}(y)), (1 - \mu_{\underset{\sim}{A}}(x))]. \tag{5.17}$$

Example 5.9

Suppose we are evaluating a new invention to determine its commercial potential. We will use two metrics to make our decisions regarding the innovation of the idea. Our metrics are the "uniqueness" of the invention, denoted by a universe of novelty scales, $X = \{1, 2, 3, 4\}$, and the "market size" of the invention's commercial market, denoted on a universe of scaled market sizes, $Y = \{1, 2, 3, 4, 5, 6\}$. In both universes, the lowest numbers are the "highest uniqueness" and the "largest market," respectively. A new invention in your group, say a compressible liquid of useful temperature and viscosity conditions, has just received scores of "medium uniqueness," denoted by fuzzy set $\underset{\sim}{A}$, and "medium market size," denoted fuzzy set $\underset{\sim}{B}$. We wish to determine the implication of such a result, that is, IF $\underset{\sim}{A}$, THEN $\underset{\sim}{B}$. We assign the invention the following fuzzy sets to represent its ratings:

$$\underset{\sim}{A} = \text{medium uniqueness} = \left\{\frac{0.6}{2} + \frac{1}{3} + \frac{0.2}{4}\right\}.$$

$$\underset{\sim}{B} = \text{medium market size} = \left\{\frac{0.4}{2} + \frac{1}{3} + \frac{0.8}{4} + \frac{0.3}{5}\right\}.$$

$$\underset{\sim}{C} = \text{diffuse market size} = \left\{\frac{0.3}{1} + \frac{0.5}{2} + \frac{0.6}{3} + \frac{0.6}{4} + \frac{0.5}{5} + \frac{0.3}{6}\right\}.$$

The following matrices are then determined in developing the membership function of the implication, $\mu_{\underset{\sim}{R}}(x,y)$, illustrated in Equation (5.17):

$$\underset{\sim}{A} \times \underset{\sim}{B} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0.6 & 0.3 & 0 \\ 0 & 0.4 & 1 & 0.8 & 0.3 & 0 \\ 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0 \end{bmatrix} \end{array},$$

$$\overline{\underset{\sim}{A}} \times Y = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix} \end{array},$$

and finally, $\underset{\sim}{R} = \max(\underset{\sim}{A} \times \underset{\sim}{B}, \overline{\underset{\sim}{A}} \times Y)$

$$\underset{\sim}{R} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0.4 & 0.4 & 0.6 & 0.6 & 0.4 & 0.4 \\ 0 & 0.4 & 1 & 0.8 & 0.3 & 0 \\ 0.8 & 0.8 & 0.8 & 0.8 & 0.8 & 0.8 \end{bmatrix} \end{array}.$$

When the logical conditional implication is of the compound form

$$\text{IF } x \text{ is } \underset{\sim}{A}, \quad \text{THEN } y \text{ is } \underset{\sim}{B}, \quad \text{ELSE } y \text{ is } \underset{\sim}{C},$$

then the equivalent fuzzy relation, $\underset{\sim}{R}$, is expressed as $\underset{\sim}{R} = (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\overline{\underset{\sim}{A}} \times \underset{\sim}{C})$ in a form just as Equation (5.6), whose membership function is expressed by the following formula:

$$\mu_{\underset{\sim}{R}}(x,y) = \max[(\mu_{\underset{\sim}{A}}(x) \wedge \mu_{\underset{\sim}{B}}(y)), ((1-\mu_{\underset{\sim}{A}}(x)) \wedge \mu_{\underset{\sim}{C}}(y))]. \tag{5.18}$$

Hence, using the result of Equation (5.18), the new relation is

$$\underset{\sim}{R} = \overline{\underset{\sim}{A}} \times \underset{\sim}{C} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0.3 & 0.5 & 0.6 & 0.6 & 0.5 & 0.3 \\ 0.3 & 0.4 & 0.4 & 0.4 & 0.4 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0.5 & 0.6 & 0.6 & 0.5 & 0.3 \end{bmatrix} \end{array},$$

and finally,

$$\underset{\sim}{R} = (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\overline{\underset{\sim}{A}} \times \underset{\sim}{C}) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 0.3 & 0.5 & 0.6 & 0.6 & 0.5 & 0.3 \\ 0.3 & 0.4 & 0.6 & 0.6 & 0.4 & 0.3 \\ 0 & 0.4 & 1 & 0.8 & 0.3 & 0 \\ 0.3 & 0.5 & 0.6 & 0.6 & 0.5 & 0.3 \end{bmatrix} \end{array},$$

Approximate Reasoning

The ultimate goal of fuzzy logic is to form the theoretical foundation for reasoning about imprecise propositions; such reasoning has been referred to as *approximate reasoning* (Zadeh, 1976, 1979). Approximate reasoning is analogous to classical logic for reasoning with precise propositions, and hence is an extension of classical propositional calculus that deals with partial truths.

Suppose we have a rule-based format to represent fuzzy information. These rules are expressed in conventional antecedent–consequent form, such as

Rule 1: IF x is $\underset{\sim}{A}$, THEN y is $\underset{\sim}{B}$, where $\underset{\sim}{A}$ and $\underset{\sim}{B}$ represent fuzzy propositions (sets).

Now suppose we introduce a new antecedent, say $\underset{\sim}{A}'$, and we consider the following rule:

Rule 2: IF x *is* $\underset{\sim}{A}'$, THEN y is $\underset{\sim}{B}'$.

From information derived from Rule 1, is it possible to derive the consequent in Rule 2, $\underset{\sim}{B}'$? The answer is yes, and the procedure is fuzzy composition. The consequent $\underset{\sim}{B}'$ can be found from the composition operation, $\underset{\sim}{B}' = \underset{\sim}{A}' \circ \underset{\sim}{R}$.

The two most common forms of the composition operator are the max–min and the max–product compositions, as initially defined in Chapter 3.

Example 5.10

Continuing with the invention example, Example 5.9, suppose that the fuzzy relation just developed, that is, R, describes the invention's commercial potential. We wish to know what market size would be associated with a uniqueness score of "almost high uniqueness." That is, with a new antecedent, $\underset{\sim}{A}'$, the consequent, $\underset{\sim}{B}'$, can be determined using composition. Let

$$\underset{\sim}{A}' = \text{almost high uniqueness} = \left\{\frac{0.5}{1} + \frac{1}{2} + \frac{0.3}{3} + \frac{0}{4}\right\}.$$

Then, using the following max–min composition

$$\underset{\sim}{B}' = \underset{\sim}{A}' \circ \underset{\sim}{R} = \left\{\frac{0.3}{1} + \frac{0.5}{2} + \frac{0.6}{3} + \frac{0.6}{4} + \frac{0.5}{5} + \frac{0.3}{6}\right\},$$

we get the fuzzy set describing the associated market size. In other words, the consequent is fairly diffuse, where there is no strong (or weak) membership value for any of the market size scores (i.e., no membership values near 0 or 1).

This power of fuzzy logic and approximate reasoning to assess qualitative knowledge can be illustrated in more familiar terms to engineers in the context of the following example in the field of biophysics.

Example 5.11

For research on the human visual system, it is sometimes necessary to characterize the strength of response to a visual stimulus based on a magnetic field measurement or on an electrical potential measurement. When using magnetic field measurements, a typical experiment will require nearly 100 off/on presentations of the stimulus at one location to obtain useful data. If the researcher is attempting to map the visual cortex of the brain, several stimulus locations must be used in the experiments. When working with a new subject, a researcher will make preliminary measurements to determine if the type of stimulus being used evokes a good response in the subject. The magnetic measurements are in units of femtotesla (10^{-15} tesla). Therefore, the inputs and outputs are both measured in terms of magnetic units.

We will define inputs on the universe X = [0, 50, 100, 150, 200] femtotesla, and outputs on the universe Y = [0, 50, 100, 150, 200] femtotesla. We will define two fuzzy sets, two different stimuli, on universe X:

$$\underset{\sim}{W} = \text{“weak stimulus”} = \left\{\frac{1}{0} + \frac{0.9}{50} + \frac{0.3}{100} + \frac{0}{150} + \frac{0}{100}\right\} \subset X.$$

$$\underset{\sim}{M} = \text{“medium stimulus”} = \left\{\frac{0}{0} + \frac{0.4}{50} + \frac{1}{100} + \frac{0.4}{150} + \frac{0}{100}\right\} \subset X.$$

and one fuzzy set on the output universe Y,

$$\underset{\sim}{S} = \text{“severe response”} = \left\{\frac{0}{0} + \frac{0}{50} + \frac{0.5}{100} + \frac{0.9}{150} + \frac{1}{200}\right\} \subset Y.$$

The complement of $\underset{\sim}{S}$ will then be

$$\overline{\underset{\sim}{S}} = \left\{\frac{1}{0} + \frac{1}{50} + \frac{0.5}{100} + \frac{0.1}{150} + \frac{0}{200}\right\}.$$

We will construct the proposition IF “weak stimulus” THEN not “severe response,” using classical implication.

$$\text{IF } \underset{\sim}{W} \text{ THEN } \overline{\underset{\sim}{S}} = \underset{\sim}{W} \rightarrow \overline{\underset{\sim}{S}} = (\underset{\sim}{W} \times \overline{\underset{\sim}{S}}) \cup (\overline{\underset{\sim}{W}} \times Y).$$

$$\underset{\sim}{W} \times \overline{\underset{\sim}{S}} = \begin{bmatrix} 1 \\ 0.9 \\ 0.3 \\ 0 \\ 0 \end{bmatrix} [1 \;\; 1 \;\; 0.5 \;\; 0.1 \;\; 0] = \begin{array}{cc} & \begin{array}{ccccc} 0 & 50 & 100 & 150 & 200 \end{array} \\ \begin{array}{c} 0 \\ 50 \\ 100 \\ 150 \\ 200 \end{array} & \begin{bmatrix} 1 & 1 & 0.5 & 0.1 & 0 \\ 0.9 & 0.9 & 0.5 & 0.1 & 0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array},$$

$$\overline{\underset{\sim}{W}} \times Y = \begin{bmatrix} 0 \\ .1 \\ .7 \\ 1 \\ 1 \end{bmatrix} [1 \;\; 1 \;\; 1 \;\; 1 \;\; 1] = \begin{array}{cc} & \begin{array}{ccccc} 0 & 50 & 100 & 150 & 200 \end{array} \\ \begin{array}{c} 0 \\ 50 \\ 100 \\ 150 \\ 200 \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array},$$

$$\underset{\sim}{R} = (\underset{\sim}{W} \times \overline{\underset{\sim}{S}}) \cup (\overline{\underset{\sim}{W}} \times Y) = \begin{array}{c|c} & \begin{array}{ccccc} 0 & 50 & 100 & 150 & 200 \end{array} \\ \begin{array}{c} 0 \\ 50 \\ 100 \\ 150 \\ 200 \end{array} & \begin{bmatrix} 1 & 1 & 0.5 & 0.1 & 0 \\ 0.9 & 0.9 & 0.5 & 0.1 & 0.1 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}.$$

This relation $\underset{\sim}{R}$, then, expresses the knowledge embedded in the rule IF "weak stimuli" THEN not "severe response." Now, using a new antecedent (IF part) for the input $\underset{\sim}{M}$ = "medium stimuli" and a max–min composition, we can find another response on the Y universe to relate approximately to the new stimulus $\underset{\sim}{M}$, that is, to find $\underset{\sim}{M} \circ \underset{\sim}{R}$:

$$\underset{\sim}{M} \circ \underset{\sim}{R} = [0 \;\; 0.4 \;\; 1 \;\; 0.4 \;\; 0] \begin{array}{c} \begin{array}{ccccc} 0 & 50 & 100 & 150 & 200 \end{array} \\ \begin{bmatrix} 1 & 1 & 0.5 & 0.1 & 0 \\ 0.9 & 0.9 & 0.5 & 0.1 & 0.1 \\ 0.7 & 0.7 & 0.7 & 0.7 & 0.7 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array} = [0.7 \;\; 0.7 \;\; 0.7 \;\; 0.7 \;\; 0.7]$$

This result might be labeled linguistically as "no measurable response."

An interesting issue in approximate reasoning is the idea of an inverse relationship between fuzzy antecedents and fuzzy consequences arising from the composition operation. Consider the following problem. Suppose we use the original antecedent, A, in the fuzzy composition. Do we get the original fuzzy consequent, B, as a result of the operation? That is, does the composition operation have a unique inverse, that is, $\underset{\sim}{B} = \underset{\sim}{A} \circ \underset{\sim}{R}$? The answer is an unqualified no, and one should not expect an inverse to exist for fuzzy composition.

Example 5.12

Again, continuing with the invention example, Examples 5.9 and 5.10, suppose that $\underset{\sim}{A}' = \underset{\sim}{A}$ = medium uniqueness. Then,

$$\underset{\sim}{B}' = \underset{\sim}{A}' \circ \underset{\sim}{R} = \underset{\sim}{A} \circ \underset{\sim}{R} = \left\{ \frac{0.4}{1} + \frac{0.4}{2} + \frac{1}{3} + \frac{0.8}{4} + \frac{0.4}{5} + \frac{0.4}{6} \right\} \neq \underset{\sim}{B}.$$

That is, the new consequent does not yield the original consequent ($\underset{\sim}{B}$ = medium market size) because the inverse is not guaranteed with fuzzy composition.

In classical binary logic this inverse does exist; that is, crisp *modus ponens* would give

$$B' = A' \circ R = A \circ R = B,$$

where the sets A and B are crisp, and the relation R is also crisp. In the case of approximate reasoning, the fuzzy inference is not precise but rather is approximate. However, the inference does represent an approximate linguistic characteristic of the relation between two universes of discourse, X and Y.

Example 5.13

Suppose you are a soils engineer and you wish to track the movement of soil particles under applied loading in an experimental apparatus that allows viewing of the soil motion. You are building pattern recognition software to enable a computer to monitor and detect the motions. However, there are some difficulties in "teaching" your software to view the motion. The tracked particle can be occluded by another particle. The occlusion can occur when a tracked particle is behind another particle, behind a mark on the camera's lens, or partially out of sight of the camera. We want to establish a relationship between particle occlusion, which is a poorly known phenomenon, and lens occlusion, which is quite well-known in photography. Let the membership functions

$$\underset{\sim}{A} = \left\{ \frac{0.1}{x_1} + \frac{0.9}{x_2} + \frac{0.0}{x_3} \right\} \quad \text{and} \quad \underset{\sim}{B} = \left\{ \frac{0}{y_1} + \frac{1}{y_2} + \frac{0}{y_3} \right\}$$

describe fuzzy sets for a *tracked particle moderately occluded* behind another particle and a *lens mark associated with moderate image quality*, respectively. Fuzzy set $\underset{\sim}{A}$ is defined on a universe $X = \{x_1, x_2, x_3\}$ of tracked particle indicators and fuzzy set $\underset{\sim}{B}$ (note in this case that $\underset{\sim}{B}$ is a crisp singleton) is defined on a universe $Y = \{y_1, y_2, y_3\}$ of lens obstruction indices. A typical rule might be IF occlusion due to particle occlusion is moderate, THEN image quality will be similar to a moderate lens obstruction, or symbolically,

$$\text{IF } x \text{ is } \underset{\sim}{A}, \text{ THEN } y \text{ is } \underset{\sim}{B} \text{ or } (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\overline{\underset{\sim}{A}} \times Y) = \underset{\sim}{R}$$

We can find the relation, $\underset{\sim}{R}$, as follows:

$$\underset{\sim}{A} \times \underset{\sim}{B} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 0 & 0.1 & 0 \\ 0 & 0.9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}, \quad \overline{\underset{\sim}{A}} \times Y = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 0.9 & 0.9 & 0.9 \\ 0.1 & 0.1 & 0.1 \\ 1 & 1 & 1 \end{bmatrix} \end{array},$$

$$\underset{\sim}{R} = (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\overline{\underset{\sim}{A}} \times Y) = \begin{bmatrix} 0.9 & 0.9 & 0.9 \\ 0.1 & 0.9 & 0.1 \\ 1 & 1 & 1 \end{bmatrix}.$$

This relation expresses in matrix form all the knowledge embedded in the implication. Let $\underset{\sim}{A}'$ be a fuzzy set, in which a tracked particle is behind a particle with *slightly more occlusion* than the particle expressed in the original antecedent $\underset{\sim}{A}$, which is given as

$$\underset{\sim}{A}' = \left\{ \frac{0.3}{x_1} + \frac{1.0}{x_2} + \frac{0.0}{x_3} \right\}.$$

We can find the associated membership of the image quality using max–min composition. For example, approximate reasoning will provide

$$\text{IF } x \text{ is } \underset{\sim}{A}', \text{ THEN } \underset{\sim}{B}' = \underset{\sim}{A}' \circ \underset{\sim}{R},$$

and we get

$$\underset{\sim}{B}' = [0.3 \quad 1 \quad 0] \circ \begin{bmatrix} 0.9 & 0.9 & 0.9 \\ 0.1 & 0.9 & 0.1 \\ 1 & 1 & 1 \end{bmatrix} = \left\{ \frac{0.3}{y_1} + \frac{0.9}{y_2} + \frac{0.3}{y_3} \right\}.$$

This image quality, $\underset{\sim}{B}'$, is more fuzzy than $\underset{\sim}{B}$, as indicated by the former's membership function.

Other Forms of the Implication Operation

There are other techniques for obtaining the fuzzy relation $\underset{\sim}{R}$ based on the IF $\underset{\sim}{A}$, THEN $\underset{\sim}{B}$, or $\underset{\sim}{R} = \underset{\sim}{A} \to \underset{\sim}{B}$. These are known as fuzzy implication operations, and they are valid for all values of $x \in X$ and $y \in Y$. The following forms of the implication operator show different techniques for obtaining the membership function values of fuzzy relation $\underset{\sim}{R}$ defined on the Cartesian product space $X \times Y$:

$$\mu_{\underset{\sim}{R}}(x,y) = \max\left[\mu_{\underset{\sim}{B}}(y), 1 - \mu_{\underset{\sim}{A}}(x)\right] \tag{5.19}$$

$$\mu_{\underset{\sim}{R}}(x,y) = \min\left[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{B}}(y)\right] \tag{5.20}$$

$$\mu_{\underset{\sim}{R}}(x,y) = \min\{1, [1 - \mu_{\underset{\sim}{A}}(x) + \mu_{\underset{\sim}{B}}(y)]\} \tag{5.21}$$

$$\mu_{\underset{\sim}{R}}(x,y) = \mu_{\underset{\sim}{A}}(x) \cdot \mu_{\underset{\sim}{B}}(y) \tag{5.22}$$

$$\mu_{\underset{\sim}{R}}(x,y) = \begin{cases} 1, & \text{for } \mu_{\underset{\sim}{A}}(x) \le \mu_{\underset{\sim}{B}}(y); \\ \mu_{\underset{\sim}{B}}(y), & \text{otherwise.} \end{cases} \tag{5.23}$$

In situations where the universes are represented by discrete elements the fuzzy relation R is a matrix.

Equation (5.19) is equivalent to classical implication (Equation (5.16)) for $\mu_{\underset{\sim}{B}}(y) \le \mu_{\underset{\sim}{A}}(x)$. Equation (5.20) has been given various terms in the literature; it has been referred to as *correlation minimum* and as *Mamdani's implication*, after British Prof. Mamdani's work in the area of system control (Mamdani, 1976). This formulation for the implication is also equivalent to the fuzzy Cartesian product of fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$, that is, $R = A \times B$. For $\mu_{\underset{\sim}{A}}(x) \geq 0.5$ and $\mu_{\underset{\sim}{B}}(y) \geq 0.5$, classical implication reduces to Mamdani's implication. The implication defined in Equation (5.21) is known as *Lukasiewicz's implication*, after the Polish logician Jan Lukasiewicz (Rescher, 1969). Equation (5.22) describes a form of *correlation-product implication* and is based on the notions of conditioning and reinforcement. This product form tends to dilute the influence of joint membership values that are small and, as such, are related to Hebbian-type learning algorithms in neuropsychology when used in artificial neural network computations. Equation (5.23) is sometimes called *Brouwerian* implication and is discussed in Sanchez (1976). Although the classical implication continues to be the most popular and is valid for fuzzy and crisp applications, these other methods have been introduced as computationally effective under certain conditions of the membership values, $\mu_{\underset{\sim}{A}}(x)$ and $\mu_{\underset{\sim}{B}}(y)$. The appropriate choice of an implication operator is a matter left to the analyst because it is typically context dependent (see Problems 5.17 and 5.18 for comparisons). Ross (1995) gives a few other fuzzy implications.

Part II: Fuzzy Systems

> *It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us.*
>
> Charles Dickens, *A Tale of Two Cities*, 1859

Natural language is perhaps the most powerful form of conveying information that humans possess for any given problem or situation that requires solving or reasoning. This power has largely remained untapped in today's mathematical paradigms; not so anymore with the utility of fuzzy logic. Consider the information contained in the passage from Charles Dickens's *A Tale of Two Cities*. Imagine reducing this passage to a more precise form such that it could be assimilated by a binary computer. First, we will have to remove the fuzziness inherent in the passage, limiting the statements to precise, either–or, Aristotelian logic. Consider the following crisp version of the first few words of the Dickens passage:

> The time interval *x* was the period exhibiting a 100% maximum of possible values as measured along some arbitrary social scale, [and] the interval *x* was also the period of time exhibiting a 100% minimum of these values as measured along the same scale (Clark, 1992).

The crisp version of this passage has established an untenable paradox, identical to that posed by the excluded middle axioms in probability theory. Another example is available from the same classic, the last sentence in Dickens's *A Tale of Two Cities*: "It is a far, far better thing that I do, than I have ever done; it is a far, far better rest that I go to, than I have ever known." It would also be difficult to address this original fuzzy phrase by an intelligent machine using binary logic. Both of these examples demonstrate the power of communication inherent in natural language, and they demonstrate how far we are from enabling intelligent machines to reason the way humans do—a long way!

Natural Language

Cognitive scientists tell us that humans base their thinking primarily on conceptual patterns and mental images rather than on any numerical quantities. In fact, the expert system paradigm known as "frames" is based on the notion of a cognitive picture in one's mind. Furthermore, humans communicate with their own natural language by referring to previous mental images with rather vague but simple terms. Despite the vagueness and ambiguity in natural language, humans communicating in a common language have little trouble in basic understanding. Our language has been termed the *shell of our thoughts* (Zadeh, 1975a). Hence, any attempts to model the human thought process as expressed in our communications with one another must be preceded by models that attempt to emulate our natural language.

Our natural language consists of fundamental terms characterized as atoms in the literature. A collection of these atoms will form the molecules, or phrases, of our natural language. The fundamental terms can be called *atomic* terms. Examples of some atomic terms are *slow, medium, young, beautiful*, and so on. A collection of atomic terms is called a composite, or simply a set of terms. Examples of composite terms are *very slow horse, medium-weight female, young tree, fairly beautiful painting*, and so on. Suppose we define the atomic terms and sets of atomic terms to exist as elements and sets on a universe of natural language terms, say universe X. Furthermore, let us define another universe, called Y, as a universe of cognitive interpretations, or meanings. Although it may seem straightforward to envision a universe of terms, it may be difficult to ponder a universe of *interpretations*. Consider this universe, however, to be a collection of individual elements and sets that represent the cognitive patterns and mental images referred to previously in this chapter. Clearly, then, these interpretations would be rather vague, and they might best be represented as fuzzy sets. Hence, an atomic term, or as Zadeh (1975a) defines it, a linguistic variable, can be interpreted using fuzzy sets.

The need for expressing linguistic variables using the precepts of mathematics is quite well established. Leibniz (1677/1951), who was an early developer of calculus, once claimed, "If we could find characters or signs appropriate for expressing all our thoughts as definitely and as exactly as arithmetic expresses numbers or geometric analysis expresses lines, we could in all subjects, in so far as they are amenable to reasoning, accomplish what is done in arithmetic and geometry" (p. 15). Fuzzy sets are a relatively new quantitative method to accomplish just what Leibniz had suggested.

With these definitions and foundations, we are now in a position to establish a formal model of linguistics using fuzzy sets. Suppose we define a specific atomic term in the universe of natural language, X, as element α, and we define a fuzzy set $\underset{\sim}{A}$ in the universe of interpretations, or meanings, Y, as a specific meaning for the term α. Then, natural language can be expressed as a mapping $\underset{\sim}{M}$ from a set of atomic terms in X to a corresponding set of interpretations defined on universe Y. Each atomic term α in X corresponds to a fuzzy set $\underset{\sim}{A}$ in Y, which is the "interpretation" of α. This mapping, which can be denoted $\underset{\sim}{M}(\alpha, \underset{\sim}{A})$, is shown schematically in Figure 5.6.

The fuzzy set $\underset{\sim}{A}$ represents the fuzziness in the mapping between an atomic term and its interpretation and can be denoted by the membership function $\mu_{\underset{\sim}{M}}(\alpha, y)$ or more simply by

$$\mu_{\underset{\sim}{M}}(\alpha, y) = \mu_{\underset{\sim}{A}}(y). \tag{5.24}$$

As an example, suppose we have the atomic term *young* (α) and we want to interpret this linguistic atom in terms of age, *y*, by a membership function that expresses the term *young*.

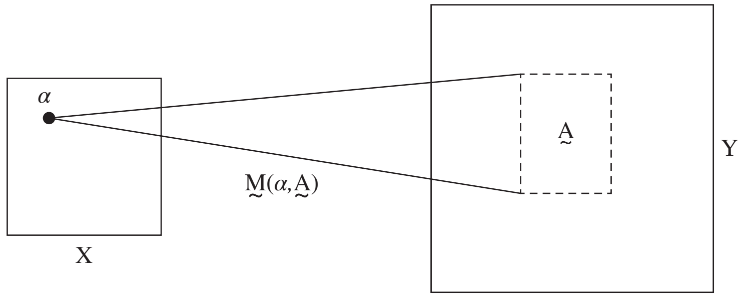

Figure 5.6 Mapping of a linguistic atom, α, to a cognitive interpretation, $\underset{\sim}{A}$.

The membership function given here for the fuzzy set labeled $\underset{\sim}{A}$, might be one interpretation of the term "young" expressed as a function of age,

$$\mu_{\underset{\sim}{M}}(\text{young}, y) = \begin{cases} \left[1 + \left(\dfrac{y-25}{5}\right)^2\right]^{-1}, & y > 25\,\text{years}; \\ 1, & y \leq \text{years}. \end{cases}$$

Similarly, the atomic term *old* might be expressed as another fuzzy set, $\underset{\sim}{O}$, on the universe of interpretation, Y, as

$$\mu_{\underset{\sim}{M}}(\text{old}, y) = 1 - \left[1 + \left(\frac{y-50}{5}\right)^2\right]^{-1}, \text{ for } 50 \leq y \leq 100.$$

On the basis of the foregoing, we can call α a natural language variable whose "value" is defined by the fuzzy set $\mu_\alpha(y)$. Hereinafter, the "value" of a linguistic variable will be synonymous with its *interpretation*.

As suggested before, a composite is a collection, or set, of atomic terms combined by various linguistic connectives such as *and*, *or*, and *not*. Define two atomic terms, α and β, on the universe X. The *interpretation* of the composite, defined on universe Y, can be defined by the following set-theoretic operations (Zadeh, 1975b):

$$\begin{aligned} \alpha \text{ or } \beta&: \mu_{\alpha \text{ or } \beta}(y) = \max\left(\mu_\alpha(y), \mu_\beta(y)\right), \\ \alpha \text{ and } \beta&: \mu_{\alpha \text{ and } \beta}(y) = \min\left(\mu_\alpha(y), \mu_\beta(y)\right), \\ \text{Not } \alpha = \overline{\alpha}&: \mu_{\overline{\alpha}}(y) = 1 - \mu_\alpha(y). \end{aligned} \tag{5.25}$$

These operations are analogous to those proposed earlier in this chapter (*standard fuzzy operations*), where the natural language connectives *and*, *or*, and *not* were logical connectives.

Linguistic Hedges

In linguistics, fundamental atomic terms are often modified with adjectives (nouns) or adverbs (verbs) like *very*, *low*, *slight*, *more or less*, *fairly*, *slightly*, *almost*, *barely*, *mostly*, *roughly*, *approximately*, and so many more that it would be difficult to list them all. We will call these modifiers *linguistic hedges*: that is, the singular meaning of an atomic term is modified, or hedged, from its original interpretation. Using fuzzy sets as the calculus of interpretation, these linguistic hedges have the effect of modifying the membership function for a basic atomic term (Zadeh, 1972). As an example, let us look at the basic linguistic atom, α, and subject it to some hedges. Define $\alpha = \int_{\mathrm{Y}} \mu_\alpha(y)/y$, then

$$\text{“Very” } \alpha = \alpha^2 = \int_{\mathrm{Y}} \frac{[\mu_\alpha(y)]^2}{y}. \tag{5.26}$$

$$\text{“Very, very”} \alpha = \alpha^4. \tag{5.27}$$

$$\text{“Slightly” } \alpha = \sqrt{\alpha} = \int_{\mathrm{Y}} \frac{[\mu_\alpha(y)]^{0.5}}{y}. \tag{5.28}$$

The expressions shown in Equations (5.26) and (5.27) are linguistic hedges known as *concentrations* (Zadeh, 1972). Concentrations tend to concentrate the elements of a fuzzy set by reducing the degree of membership of all elements that are only “partly” in the set. The less an element is in a set (i.e., the lower its original membership value), the more it is reduced in membership through concentration. For example, by using Equation (5.26) for the hedge *very*, a membership value of 0.9 is reduced by 10% to a value of 0.81, but a membership value of 0.1 is reduced by an order of magnitude to 0.01. This decrease is simply a manifestation of the properties of the membership value itself; for $0 \le \mu \le 1$, then $\mu \ge \mu^2$. Alternatively, the expression given in Equation (5.28) is a linguistic hedge known as *dilation* (or dilution in some publications). Dilations stretch or dilate a fuzzy set by increasing the membership of elements that are “partly” in the set (Zadeh, 1972). For example, using Equation (5.28) for the hedge *slightly*, a membership value of 0.81 is increased by 11% to a value of 0.9, whereas a membership value of 0.01 is increased by an order of magnitude to 0.1.

Another operation on linguistic fuzzy sets is known as *intensification*. This operation acts in a combination of concentration and dilation. It increases the degree of membership of those elements in the set with original membership values greater than 0.5, and it decreases the degree of membership of those elements in the set with original membership values less than 0.5. This also has the effect of making the boundaries of the membership function (see Figure 4.1) steeper. *Intensification* can be expressed by numerous algorithms; one of which, proposed by Zadeh (1972), is

$$\text{“intensify” } \alpha = \begin{cases} 2\mu_\alpha^2(y), & \text{for } 0 \le \mu_\alpha(y) \le 0.5; \\ 1 - 2[1 - \mu_\alpha(y)]^2, & \text{for } 0.5 \le \mu_\alpha(y) \le 1. \end{cases} \tag{5.29}$$

Intensification increases the contrast between the elements of the set that have more than half-membership and those that have less than half-membership. Figures 5.7 to 5.9 illustrate

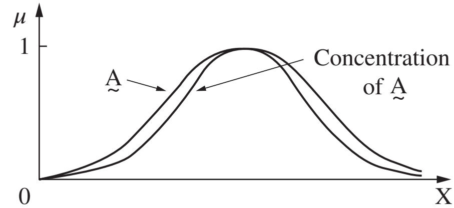

Figure 5.7 Fuzzy concentration.

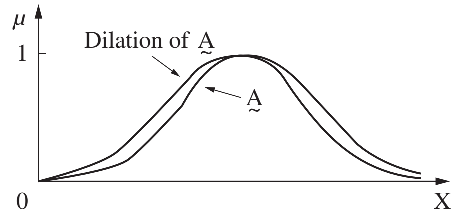

Figure 5.8 Fuzzy dilation.

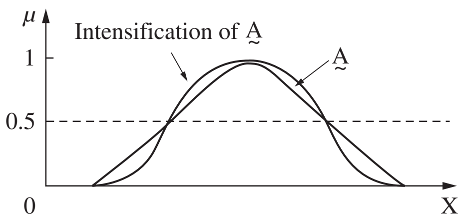

Figure 5.9 Fuzzy intensification.

the operations of concentration, dilation, and intensification, respectively, for fuzzy linguistic hedges on a typical fuzzy set $\underset{\sim}{A}$.

Example 5.14
Suppose we have a universe of integers, Y = {1, 2, 3, 4, 5}. We define the following linguistic terms as a mapping onto Y:

$$\text{“Small”} = \left\{\frac{1}{1} + \frac{0.8}{2} + \frac{0.6}{3} + \frac{0.4}{4} + \frac{0.2}{5}\right\}.$$

$$\text{“Large”} = \left\{\frac{0.2}{1} + \frac{0.4}{2} + \frac{0.6}{3} + \frac{0.8}{4} + \frac{1}{5}\right\}.$$

Now we modify these two linguistic terms with hedges,

$$\text{“Very Small”} = \text{“Small”}^2 \ (\text{Equation}\,(5.26)) = \left\{\frac{1}{1} + \frac{0.64}{2} + \frac{0.36}{3} + \frac{0.16}{4} + \frac{0.04}{5}\right\}.$$

$$\text{“Not Very Small”} = 1 - \text{“Very Small”} = \left\{\frac{0}{1} + \frac{0.36}{2} + \frac{0.64}{3} + \frac{0.84}{4} + \frac{0.96}{5}\right\}.$$

Then we construct a phrase, or a composite term:

$$\alpha = \text{“Not Very Small and not very, very large,”}$$

which involves the following set-theoretic operations:

$$\alpha = \left(\frac{0.36}{2} + \frac{0.64}{3} + \frac{0.84}{4} + \frac{0.96}{5}\right) \cap \left(\frac{1}{1} + \frac{1}{2} + \frac{0.9}{3} + \frac{0.6}{4}\right) = \left(\frac{0.36}{2} + \frac{0.64}{3} + \frac{0.6}{4}\right).$$

Suppose we want to construct a linguistic variable “intensely small” (extremely small); we will make use of Equation (5.29) to modify “small” as follows:

$$\begin{aligned}\text{“Intensely small”} &= \left\{\frac{1-2[1-1]^2}{1} + \frac{1-2[1-0.8]^2}{2}\right\} \\ &\quad + \frac{1-2[1-0.6]^2}{3} + \frac{2[0.4]^2}{4} + \frac{2[0.2]^2}{5} \\ &= \left\{\frac{1}{1} + \frac{0.92}{2} + \frac{0.68}{3} + \frac{0.32}{4} + \frac{0.08}{5}\right\}.\end{aligned}$$

In summary, the foregoing material introduces the idea of a *linguistic variable* (atomic term), which is a variable whose values (interpretation) are natural language expressions referring to the contextual semantics of the variable.

Fuzzy (Rule-Based) Systems

In the field of artificial intelligence (machine intelligence), there are various ways to represent knowledge. Perhaps the most common way to represent human knowledge is to form it into natural language expressions of the type

$$\text{IF premise (antecedent), THEN conclution (consequent).} \tag{5.30}$$

The form in Expression (5.30) is commonly referred to as the IF–THEN *rule-based* form; this form is generally referred to as the *deductive* form. It typically expresses an inference such that if we know a fact (premise, hypothesis, antecedent), then we can infer, or derive, another fact called a conclusion (consequent). This form of knowledge representation, characterized as *shallow knowledge*, is quite appropriate in the context of linguistics because it expresses human empirical and heuristic knowledge in our own language of communication. It does not, however, capture the *deeper* forms of knowledge usually associated with intuition, structure, function, and behavior of the objects around us simply because these latter forms of knowledge are not readily reduced to linguistic phrases or representations; this deeper form, as described in Chapter 1, is referred to as *inductive*. The fuzzy rule-based system is most useful in modeling some complex systems that can be observed by humans because they make use of linguistic variables as their antecedents and consequents; as described here these linguistic variables can be naturally represented by fuzzy sets and logical connectives of these sets.

Aggregation of Fuzzy Rules

Most rule-based systems involve more than one rule. The process of obtaining the overall consequent (conclusion) from the individual consequents contributed by each rule in the rule-base is known as aggregation of rules. In determining an aggregation strategy, two simple extreme cases exist (Ross, 1995)

1. *Conjunctive system of rules*: In the case of a system of rules that must be jointly satisfied, the rules are connected by "and" connectives. In this case, the aggregated output (consequent), y, is found by the fuzzy intersection of all individual rule consequents, y^i, where $i = 1, 2, \ldots r$, as

$$y = y^1 \cap y^2 \cap \ldots \cap y^r,$$

which is defined by the membership function

$$\mu_y(y) = \min\left(\mu_{y^1}(y),\ \mu_{y^2}(y),\ \ldots,\ \mu_{y^r}(y)\right),\ \text{for } y \in \mathrm{Y}. \tag{5.31}$$

2. *Disjunctive system of rules*: For the case of a disjunctive system of rules where the satisfaction of at least one rule is required, the rules are connected by the "or" connectives. In this case, the aggregated output is found by the fuzzy union of all individual rule contributions, as

$$y = y^1 \cup y^2 \cup \ldots \cup y^r,$$

which is defined by the membership function

$$\mu_y(y) = \max\left(\mu_{y^1}(y),\ \mu_{y^2}(y),\ \ldots,\ \mu_{y^r}(y)\right),\ \text{for } y \in \mathrm{Y}. \tag{5.32}$$

Graphical Techniques of Inference

Part I of this chapter illustrates mathematical procedures to conduct deductive inferencing of IF–THEN rules. These procedures can be implemented on a computer for processing speed.

Sometimes, however, it is useful to be able to conduct the inference computation manually with a few rules to check computer programs or to verify the inference operations. Conducting the matrix operations illustrated in this chapter, Part I, for a few rule sets can quickly become quite onerous. Graphical methods that emulate the inference process and that make manual computations involving a few simple rules straightforward have been proposed (Jang, Sun, and Mizutani, 1997). This section describes three common methods of deductive inference for fuzzy systems based on linguistic rules: (1) Mamdani systems, (2) Sugeno models, and (3) Tsukamoto models.

The first inference method, from Mamdani and Assilian (1975), is the most common in practice and in the literature. To begin the general illustration of this idea, we consider a simple two-rule system where each rule comprises two antecedents and one consequent. This is analogous to a dual-input and single-output fuzzy system. The graphical procedures illustrated here can be easily extended and will hold for fuzzy rule-bases (or fuzzy systems) with any number of antecedents (inputs) and consequents (outputs). A fuzzy system with two noninteractive inputs x_1 and x_2 (antecedents) and a single output y (consequent) is described by a collection of r linguistic IF–THEN propositions in the Mamdani form:

$$\text{IF } x_1 \text{ is } \underset{\sim}{\mathrm{A}}_1^k \text{ and } x_2 \text{ is } \underset{\sim}{\mathrm{A}}_2^k \text{ THEN } y^k \text{ is } \underset{\sim}{\mathrm{B}}^k, \quad \text{for } k=1,2,\ldots,r, \tag{5.33}$$

where $\underset{\sim}{\mathrm{A}}_1^k$ and $\underset{\sim}{\mathrm{A}}_2^k$ are the fuzzy sets representing the kth antecedent pairs and $\underset{\sim}{\mathrm{B}}^k$ is the fuzzy set representing the kth consequent.

In the following presentation, we consider two different cases of two-input Mamdani systems: (1) the inputs to the system are scalar values, and we use a max–min inference method, and (2) the inputs to the system are scalar values, and we use a max–product inference method. Of course, the inputs to any fuzzy system can also be a membership function, such as a gauge reading that has been fuzzified, but we shall lose no generality in describing the method by employing fuzzy singletons (scalar values) as the input.

Case 1

Inputs x_1 and x_2 are crisp values, that is, delta functions. The rule-based system is described in Equation (5.33), so membership for the inputs x_1 and x_2 will be described as

$$\mu(x_1)=\delta(x_1-\text{input}(i))=\begin{cases}1, & x_1=\text{input}(i);\\ 0, & \text{otherwise.}\end{cases} \quad \text{and}$$

$$\mu(x_2)=\delta(x_2-\text{input}(j))=\begin{cases}1, & x_2=\text{input}(j);\\ 0, & \text{otherwise.}\end{cases} \tag{5.34}$$

Based on the Mamdani implication method of inference given in this chapter, Equation (5.20), and for a set of disjunctive rules, the aggregated output for the r rules will be given as

$$\mu_{\underset{\sim}{\mathrm{B}}^k}(y) = \max_k \left[\min\left[\mu_{\underset{\sim}{\mathrm{A}}_1^k}(\mathrm{input}(i)),\ \mu_{\underset{\sim}{\mathrm{A}}_2^k}(\mathrm{input}(j))\right]\right],\quad k = 1,\ 2,\ \ldots,\ r. \tag{5.35}$$

Equation (5.35) has a simple graphical interpretation, as seen in Figure 5.10. Figure 5.10 illustrates the graphical analysis of two rules, where the symbols A_{11} and A_{12} refer to the first and second fuzzy antecedents of the first rule, respectively, and the symbol B_1 refers to the fuzzy consequent of the first rule; the symbols A_{21} and A_{22} refer to the first and second fuzzy antecedents, respectively, of the second rule, and the symbol B_2 refers to the fuzzy consequent of the second rule. The minimum function in Equation (5.35) is illustrated in Figure 5.10 and arises because the antecedent pairs given in the general rule structure for this system are connected by a logical "and" connective, as seen in Equation (5.33). The minimum membership value for the antecedents propagates through to the consequent and truncates the membership function for the consequent of each rule. This graphical inference is done for each rule. Then,

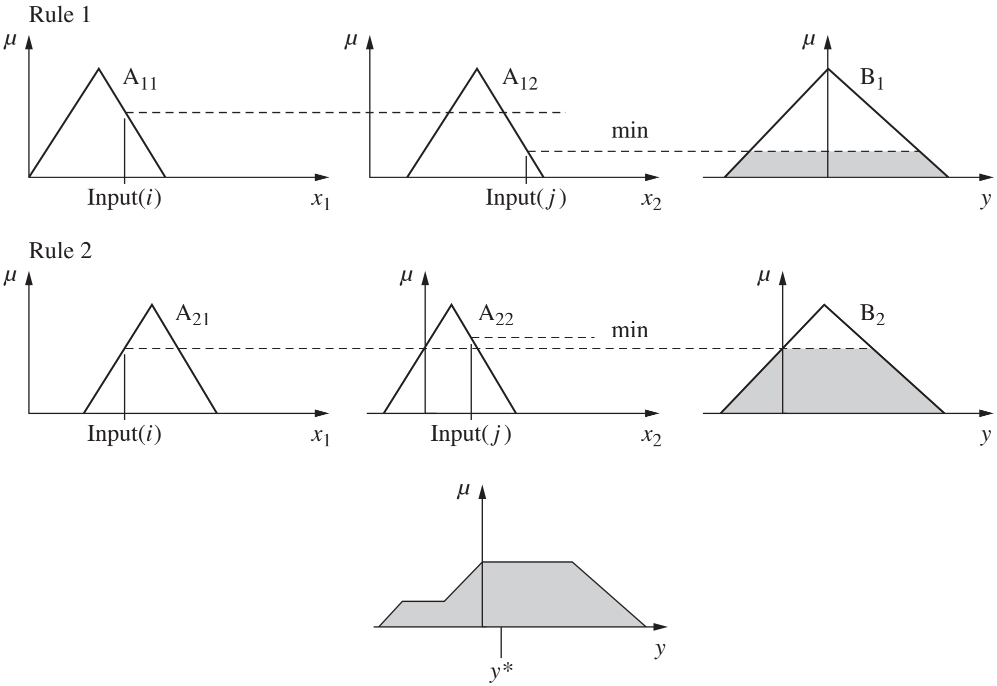

Figure 5.10 Graphical Mamdani (max–min) inference method with crisp inputs.

the truncated membership functions for each rule are aggregated using the graphical equivalent of either Equation (5.31), for conjunction rules, or Equation (5.32), for disjunctive rules; in Figure 5.10 the rules are disjunctive, so the aggregation operation *max* results in an aggregated membership function comprising the outer envelope of the individual truncated membership forms from each rule. If one wishes to find a crisp value for the aggregated output, some appropriate defuzzification technique (Chapter 4) could be employed to the aggregated membership function, and a value such as y^* shown in Figure 5.10 would result.

Case 2

In the preceding example, if we were to use a max–product (or correlation-product) implication technique (Equation (5.22)) for a set of disjunctive rules, the aggregated output for the r rules would be given as

$$\mu_{\underset{\sim}{B}^k}(y) = \max_k \left[\mu_{\underset{\sim}{A}_1^k}(\text{input}(i)) \cdot \mu_{\underset{\sim}{A}_2^k}(\text{input}(j))\right], \; k = 1, 2, \ldots, r, \tag{5.36}$$

and the resulting graphical equivalent of Equation (5.36) would be as shown in Figure 5.11. In Figure 5.11 the effect of the max–product implication is shown by the consequent membership functions remaining as scaled triangles (instead of truncated triangles as in Case 1). Again, Figure 5.11 shows the aggregated consequent resulting from a disjunctive set of rules (the outer envelope of the individual scaled consequents) and a defuzzified value, y^*, resulting from some defuzzification method (Chapter 4).

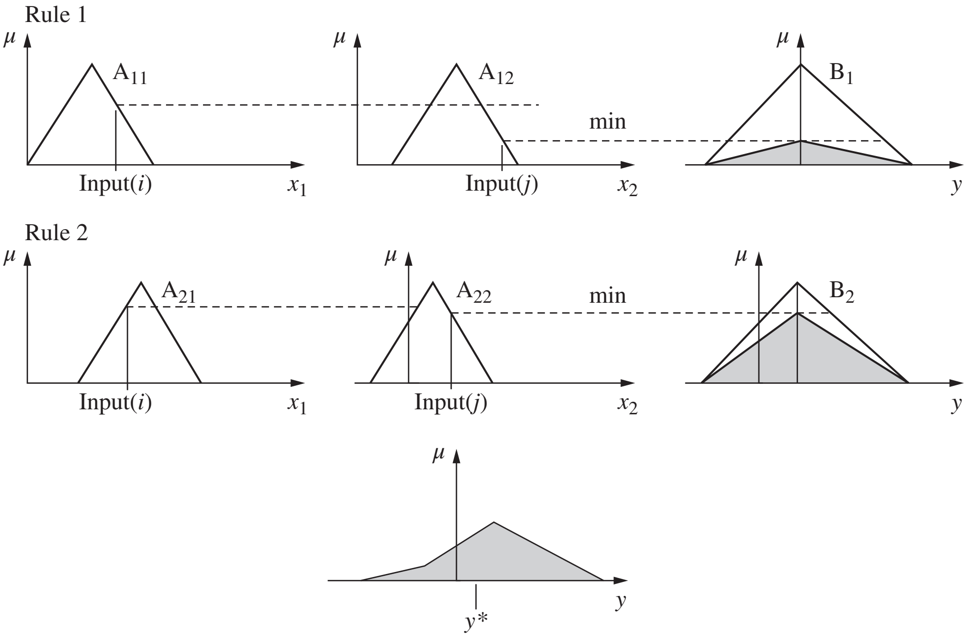

Figure 5.11 Graphical Mamdani (max–product) implication method with crisp inputs.

Example 5.15

In mechanics, the energy of a moving body is called *kinetic energy*. If an object of mass m (kilograms) is moving with a velocity v (meters per second), then the kinetic energy k (in joules) is given by the equation $k = \frac{1}{2}mv^2$. Suppose we model the mass and velocity as inputs to a system (moving body) and the energy as output, then observe the system for a while and deduce the following two disjunctive rules of inference based on our observations:

Rule 1: IF x_1 is $\underset{\sim}{A}_1^1$ (small mass) *and* x_2 is $\underset{\sim}{A}_2^1$ (high velocity),
THEN y is $\underset{\sim}{B}^1$ (medium energy).

Rule 2: IF x_1 is $\underset{\sim}{A}_1^2$ (large mass) *or* x_2 is $\underset{\sim}{A}_2^2$ (*medium* velocity),
THEN y is $\underset{\sim}{B}^2$ (high energy).

We now proceed to describe these two rules in a graphical form and illustrate the two cases of graphical inference presented previously in this section.

Suppose we have made some observations of the system (moving body) and we estimate the values of the two inputs, mass and velocity, as crisp values. For example, let input (i) = 0.35 kg (mass) and input(j) = 55 m s^{-1} (velocity). Case 1 models the inputs as delta functions, Equations (5.34) use a Mamdani implication, Equation (5.35). Graphically, this is illustrated in Figure 5.12, where the output fuzzy membership function is defuzzified using a centroid method.

In Figures 5.12 and 5.13, the two rules governing the behavior of the moving body system are illustrated graphically. The antecedents, mass (kilograms), and velocity (meters per second), for each rule are shown as fuzzy membership functions corresponding to the linguistic values for each antecedent. Moreover, the consequent, energy (joules), for each rule is also shown as a fuzzy membership function corresponding to the linguistic label for that consequent. The inputs for mass and velocity intersect the antecedent membership functions at some membership level. The minimum or maximum of the two membership values is propagated to the consequent depending on whether the "and" or "or" connective, respectively, is used between the two antecedents in the rule. The propagated membership value from operations on the antecedents then truncates (for Mamdani implication) or scales (for max–product implication) the membership function for the consequent for that rule. This truncation or scaling is conducted for each rule, and then the truncated or scaled membership functions from each rule are aggregated according to Equation (5.31) (conjunctive) or Equation (5.32) (disjunctive). In this example, we are using two disjunctive rules.

In Case 2, we only change the method of implication from the first case. Now using a max–product implication method, Equation (5.36), and a centroidal defuzzification method, the graphical result is shown in Figure 5.13.

The Mamdani method has several variations. There are different t-norms to use for the connectives of the antecedents, different aggregation operators for the rules, and numerous defuzzification methods that could be used. As the foregoing example illustrates, the two Mamdani methods yield different shapes for the aggregated fuzzy consequents for the two rules used. However, the defuzzified values for the output energy are both fairly consistent: 244 J and 260 J. The power of fuzzy rule-based systems is their ability to yield "good" results with reasonably simple mathematical operations.

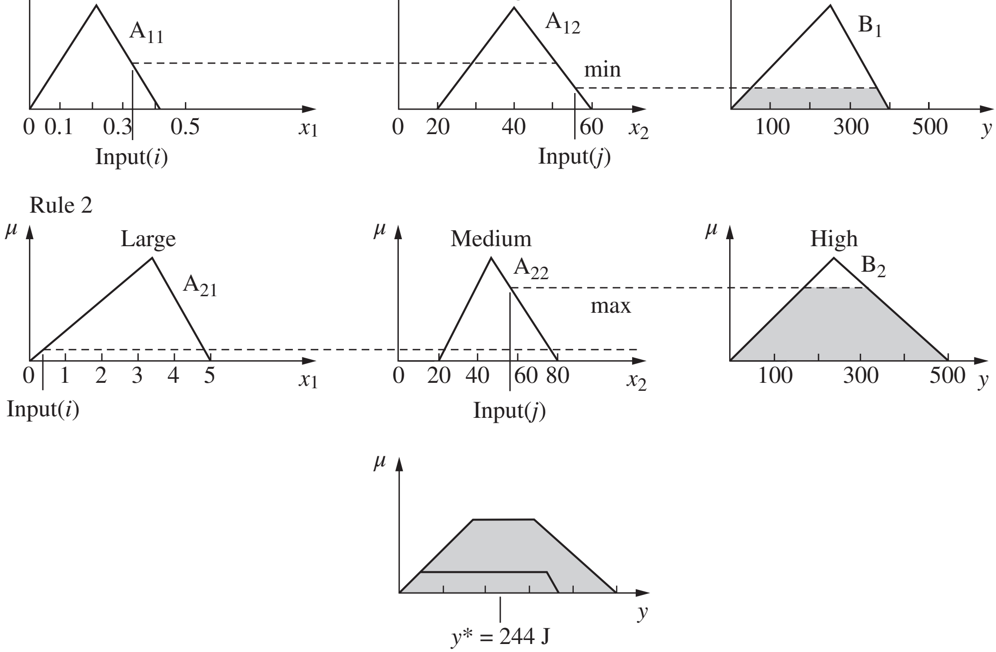

Figure 5.12 Fuzzy inference method using the case 1 graphical approach.

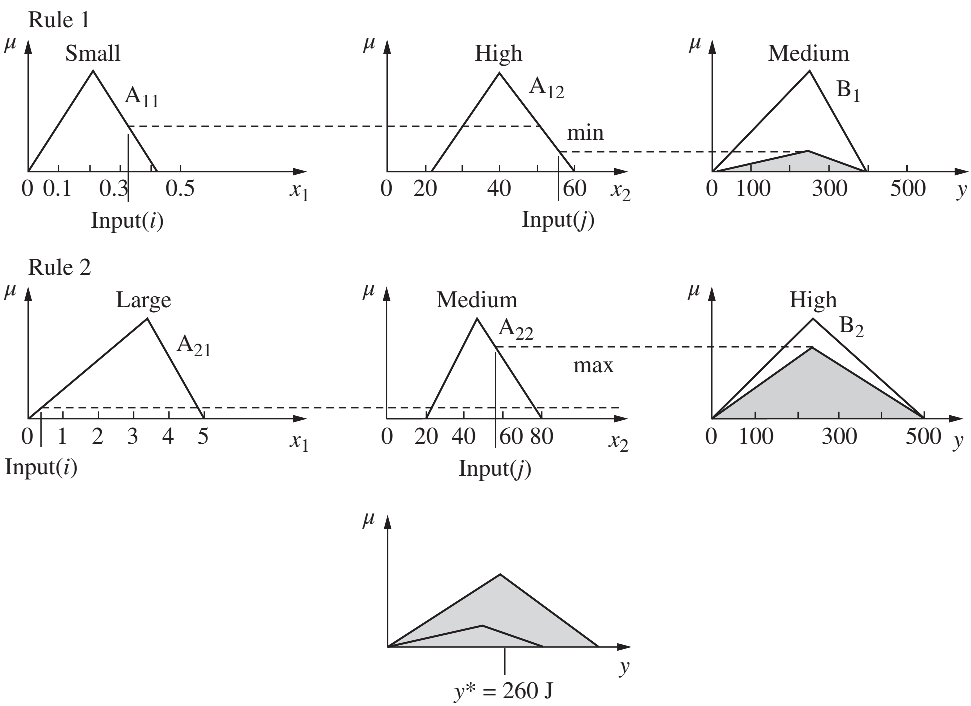

Figure 5.13 Fuzzy inference method using the case 2 graphical approach.

The second inference method, generally referred to as the *Sugeno method*, or the *TSK method* (Takagi and Sugeno, 1985; Sugeno and Kang, 1988), was proposed in an effort to develop a systematic approach to generating fuzzy rules from a given input–output data set. A typical rule in a Sugeno model, which has two inputs x and y and output z, has the form

$$\text{IF } x \text{ is } \underset{\sim}{A} \textit{ and } y \text{ is } \underset{\sim}{B}, \text{ THEN } z \text{ is } z = f(x, y),$$

where $z = f(x, y)$ is a crisp function in the consequent. Usually $f(x, y)$ is a polynomial function in the inputs x and y, but it can be any general function as long as it describes the output of the system within the fuzzy region specified in the antecedent of the rule to which it is applied. When $f(x, y)$ is a constant, the inference system is called a *zero-order Sugeno model*, which is a special case of the Mamdani system in which each rule's consequent is specified as a fuzzy singleton. When $f(x, y)$ is a linear function of x and y, the inference system is called a *first-order Sugeno model*. Jang and colleagues (1997) point out that the output of a zero-order Sugeno model is a smooth function of its input variables as long as the neighboring membership functions in the antecedent have enough overlap. By contrast, the overlap of the membership functions in the consequent of a Mamdani model does not have a decisive effect on the smoothness; it is the overlap of the antecedent membership functions that determines the smoothness of the resulting system behavior.

In a Sugeno model, each rule has a crisp output, given by a function. Because of this the overall output is obtained via a weighted average defuzzification (Equation (4.6)), as shown in Figure 5.14. This process avoids the time-consuming methods of defuzzification necessary in the Mamdani model.

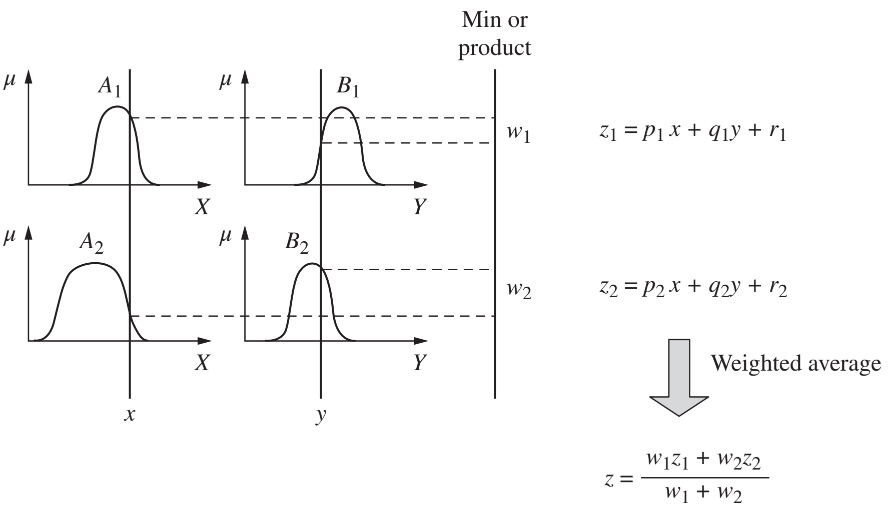

Figure 5.14 The Sugeno fuzzy model. Jang, Jyh-Shing Roger; Sun, Chuen-Tsai; Mizutani, Eiji, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* 1st Edition, © 1997. Reprinted by permission of Pearson Education Inc., Upper Saddle River, New Jersey.

Example 5.16

An example of a two-input, single-output Sugeno model with four rules is repeated from Jang and colleagues (1997):

$$\text{IF X is small and Y is small, THEN } z = -x + y + 1.$$

$$\text{IF X is small and Y is large, THEN } z = -y + 3.$$

$$\text{IF X is large and Y is small, THEN } z = -x + 3.$$

$$\text{IF X is large and Y is large, THEN } z = x + y + 2.$$

Figure 5.15a plots the membership function of inputs X and Y, and Figure 5.15b is the resulting input–output surface of the system. The surface is complex, but it is still obvious that the surface comprises four planes, each of which is specified by the output function of each of the four rules. Figure 5.15b shows that there is a smooth transition between the four output planes. Without the mathematically difficult process of a defuzzification operation, the Sugeno model is a popular method for sample-based fuzzy systems modeling.

The third inference method is from Tsukamoto (1979). In this method, the consequent of each fuzzy rule is represented by a fuzzy set with a monotonic membership function, as shown in Figure 5.16. In a monotonic membership function, sometimes called a *shoulder function*, the inferred output of each rule is defined as a crisp value induced by the membership value coming from the antecedent clause of the rule. The overall output is calculated by the weighted average of each rule's output, as seen in Figure 5.16. Because each rule infers a crisp output, Tsukamoto model's aggregation of the overall output also avoids the time-consuming process of defuzzification. Because of the special nature of the output membership functions required by the method, it is not as useful as a general approach, and must be employed in specific situations.

Example 5.17

An example of a single-input, single-output Tsukamoto fuzzy model is given by the following rules:

$$\text{IF X is small, THEN Y is } C_1,$$

$$\text{IF X is medium, THEN Y is } C_2,$$

$$\text{IF X is large, THEN Y is } C_3,$$

where the antecedent and consequent fuzzy sets are as shown in Figure 5.17a and Figure 5.17b, respectively. If we plot the output of each of the three rules as a function of the input, X, we get the three curves shown in Figure 5.17c (the solid curve is Rule 1, the dashed curve is Rule 2, and the dotted curve is Rule 3). The overall output of the three-rule system is shown in Figure 5.17d. Because the reasoning mechanism of the Tsukamoto fuzzy model does not strictly follow a composition operation in its inference, it always generates a crisp output even when the input and output membership functions are fuzzy membership functions.

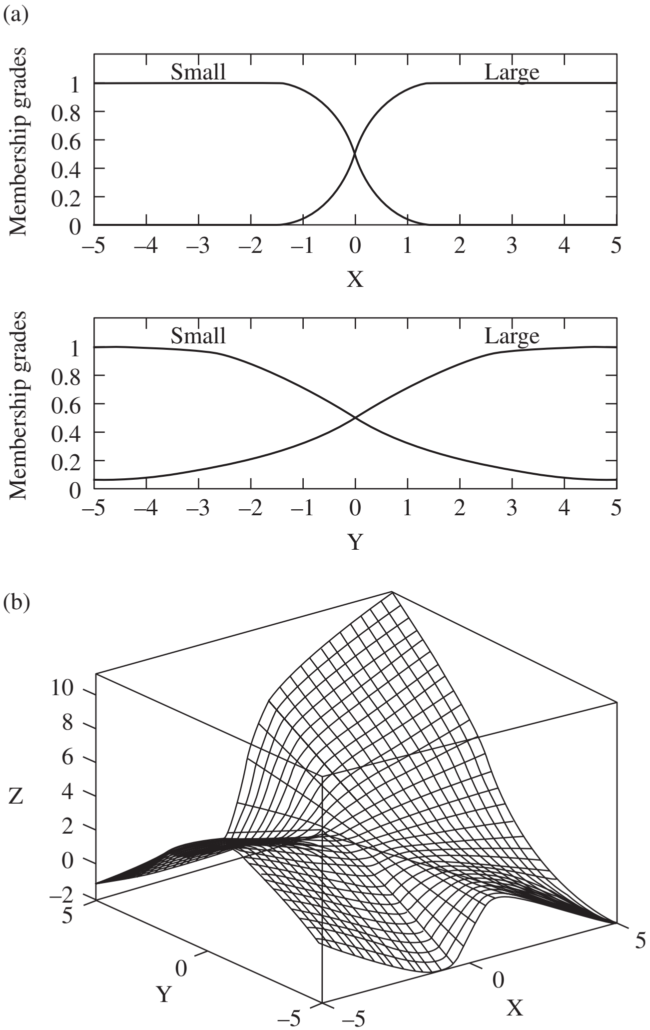

Figure 5.15 Sugeno Model for Example 5.16. (a) antecedent and consequent membership functions; (b) overall system response surface. Jang, Jyh-Shing Roger; Sun, Chuen-Tsai; Mizutani, Eiji, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* 1st Edition, © 1997. Reprinted by permission of Pearson Education Inc., Upper Saddle River, New Jersey.

Example 5.18

In heat exchanger design, a flexibility analysis requires the designer to determine if the size of the heat exchanger is either small or large. To quantify this linguistic vagueness of size, we form the general design equation for a heat exchanger, $Q = AU\Delta T_{\text{log mean}}$, where the heat transfer coefficient U and area A need to be determined. Figure 5.18 show a schematic of this exchanger.

We want to determine the sizes for a heat exchanger in which a stream of benzene is heated using saturated steam at pressure 68.95 kPa and temperature 362.7 K. The initial temperature of

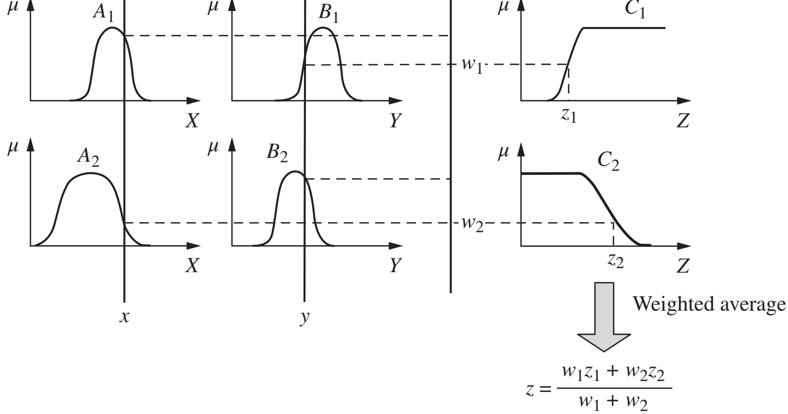

Figure 5.16 The Tsukamoto fuzzy model. Jang, Jyh-Shing Roger; Sun, Chuen-Tsai; Mizutani, Eiji, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* 1st Edition, © 1997. Reprinted by permission of Pearson Education Inc., Upper Saddle River, New Jersey.

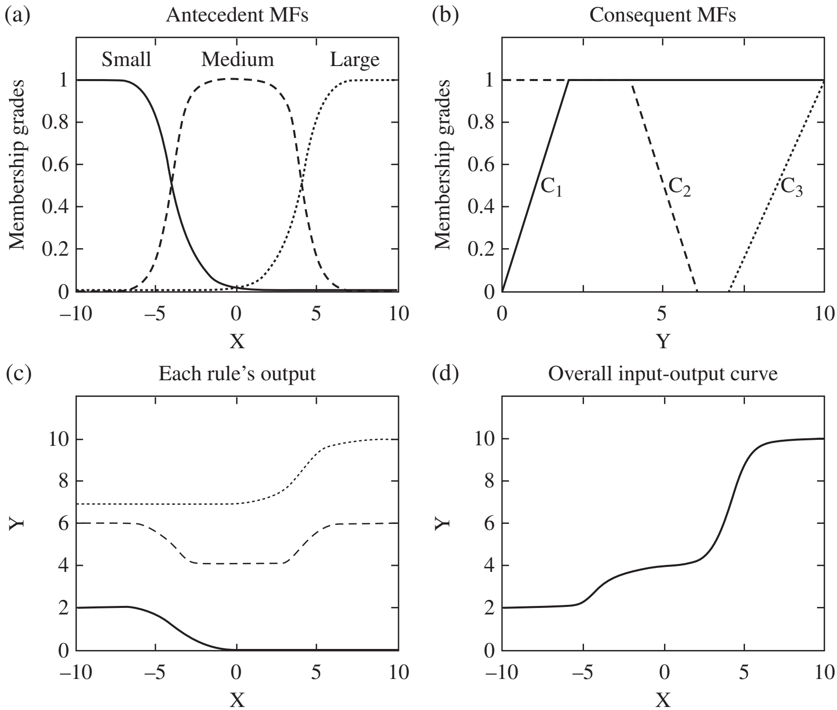

Figure 5.17 Tsukamoto model for Example 5.17. (a) antecedent membership functions; (b) consequent membership functions; (c) each rule's output curve; (d) overall system response curve. Jang, Jyh-Shing Roger; Sun, Chuen-Tsai; Mizutani, Eiji, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* 1st Edition, © 1997. Reprinted by permission of Pearson Education Inc., Upper Saddle River, New Jersey.

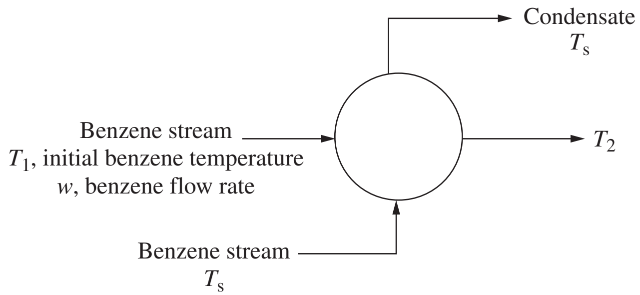

Figure 5.18 Heat exchanger design.

the benzene steam is 17 °C, and the model used to determine the size of the heat exchanger is the following:

$$AU = wC_p \ \ln\left(\frac{T_s - T_1}{\Delta T_{app}}\right),$$

where C_p is the heat capacity of the benzene [1.7543 kJ (K kg)$^{-1}$] and $T_s - T_1 = 72.55$ K.

We will model the benzene flow rate, w, in kilograms per second, and temperature approach (ΔT_{app}) in kelvin, as the inputs, and we will model the size of the heat exchanger as output. We will deduce the following disjunctive rules of inference based on the observations of the model:

Rule 1: IF w is $\underset{\sim}{A}_1^1$ (large flow rate) and ΔT_{app} is $\underset{\sim}{A}_2^1$ (small approach),
THEN AU is $\underset{\sim}{B}^1$ (large heat exchanger).

Rule 2: IF w is $\underset{\sim}{A}_1^2$ (small flow rate) or ΔT_{app} is $\underset{\sim}{A}_2^2$ (large approach),
THEN AU is $\underset{\sim}{B}^2$ (small heat exchanger).

Rule 3: IF w is $\underset{\sim}{A}_1^2$ (small flow rate) and ΔT_{app} is $\underset{\sim}{A}_2^1$ (small approach),
THEN AU is $\underset{\sim}{B}^1$ (large heat exchanger).

The graphical equivalent of these rules is shown in Figure 5.19. A weighted average defuzzification method will be employed to compare the results from one input pair for each of the three following inference methods: Mamdani, Sugeno, and Tsukamoto.

We will input two crisp values of benzene flow rate and temperature approach:

$$w = 1300 \text{ kg s}^{-1} \quad \text{and} \quad \Delta T_{app} = 6.5 \text{ K}.$$

1. Using the max–min Mamdani implication method of inference, we know that

$$\mu_{\underset{\sim}{B}^k}(AU) = \max_k \left\{ \min\left[\mu_{A_1^k}(w), \mu_{A_2^k}(\Delta T_{app})\right]\right\}$$

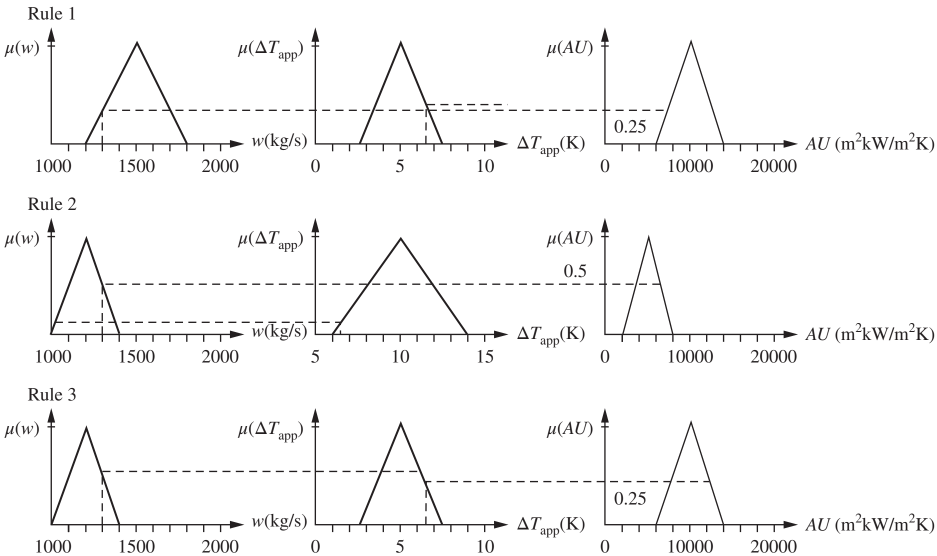

Figure 5.19 Graphical inference using the Mamdani method for three rules.

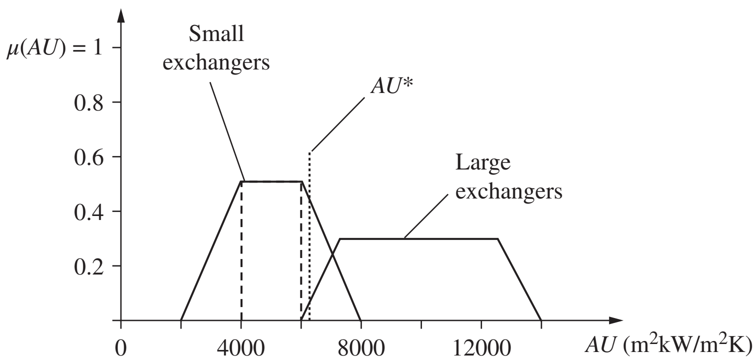

Figure 5.20 Result of the defuzzification step in the Mamdani method.

Using the graphical approach, we get the rules shown in Figure 5.19.

And using a weighted average defuzzification, we get

$$AU^* = \frac{(5000\,\mathrm{m^2kW/m^2K})(0.5) + (10000\,\mathrm{m^2kW/m^2K})(2)(0.25)}{0.5 + 0.25 + 0.25} = 7500\,\mathrm{m^2kW/m^2K},\mathrm{M}$$

which is also shown in Figure 5.20.

2. For the Sugeno fuzzy method of inference, we have experience in heat exchanger design that gives the following expressions in a polynomial form for our two consequents (small and large heat exchangers):

$$AU_{\text{small}} = 3.4765w - 210.5\Delta T_{\text{app}} + 2103.$$
$$AU_{\text{large}} = 4.6925w - 526.2\Delta T_{\text{app}} + 2631.$$

Taking the minimum membership value for the input conjunction "and" of Rules 1 and 3, and the maximum value for the input disjunction "or" in Rule 2, the membership value of each of the consequents will be (Figure 5.19):

$$\text{Rule 1:} \ \mu(AU) = 0.25,$$
$$\text{Rule 2:} \ \mu(AU) = 0.5,$$
$$\text{Rule 3:} \ \mu(AU) = 0.25.$$

Then

$$AU_{\text{small}} = 5256\,\text{m}^2\text{kW/m}^2\text{K} \ \text{ and } \ AU_{\text{large}} = 5311\,\text{m}^2\text{kW/m}^2\text{K}.$$

Finally, the defuzzified value of the heat exchange size is (using the weighted average method of defuzzification; see Figure 5.14)

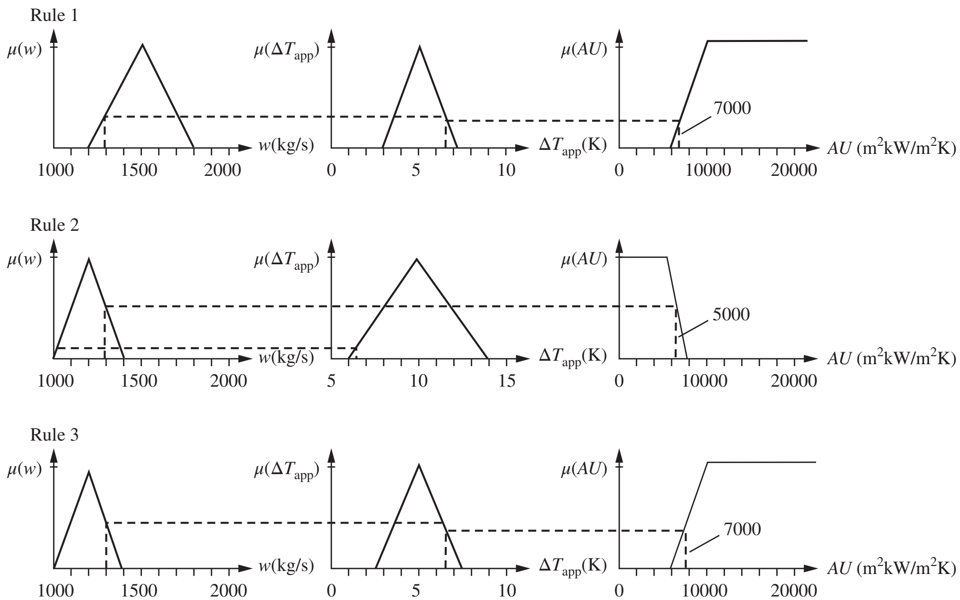

Figure 5.21 Tsukamoto method of inference for the three rules.

$$AU^* = \frac{(5311\,\mathrm{m^2kW/m^2K})(0.25) + (5256\,\mathrm{m^2kW/m^2K})(0.5) + (5311\,\mathrm{m^2kW/m^2K})(0.25)}{0.25 + 0.5 + 0.25}$$

$$= 5283.5\,\mathrm{m^2kW/m^2K}.$$

3. For the Tsukamoto fuzzy method of inference, we modify the output membership functions from the Mamdani case (see Figure 5.19), but we added shoulders to them for Tsukamoto. Using a graphical approach, we get the rules shown in Figure 5.21.

   The defuzzified value of the heat exchanger size is

$$AU^* = \frac{(7000\,\mathrm{m^2kW/m^2K})(0.25) + (5000\,\mathrm{m^2kW/m^2K})(0.5) + (7000\,\mathrm{m^2kW/m^2K})(0.25)}{0.25 + 0.5 + 0.25}$$

$$= 6250\ \mathrm{m^2kW/m^2K}.$$

The Mamdani and Tsukamoto methods yield similar values of AU because they are based on similar membership functions for the output. The difference with the Sugeno method is a function of the accuracy of the polynomials that model the output.

Summary

This chapter has presented the basic axioms, operations, and properties of binary logic and fuzzy logic. Just as in Chapter 2, we find that the only significant difference between a binary logic and a fuzzy logic stems from the logical equivalent of the excluded middle axioms. Examples that illustrate the various operations of a fuzzy logic are provided. An approximate reasoning, proposed by Zadeh (1976, 1979), is presented to illustrate the power of using fuzzy sets in the reasoning process. Other works in the area of fuzzy reasoning and approximate reasoning have been helpful in explaining the theory; for example, a useful comparison study (Mizumoto and Zimmerman, 1982) and a work defining the mathematical foundations (Yager, 1985). From a general point of view, other multivalued logics have been developed (Dubois and Prade, 1980; Klir and Folger, 1988), and these other logics may be viewed as *fuzzy logics* in the sense that they represent more than just the crisp truth values of 0 and 1. In fact, Gaines (1976) has shown that some forms of multivalued logics result from fuzzifying, in the sense of the extension principle, the standard propositional calculus. The illustration of approximate reasoning given here is conducted using fuzzy relations to represent the rules of inference. The chapter concludes by pointing out the rich variety in reasoning possible with fuzzy logic when one considers the vast array of implication and composition operations; an example of this can be found in Yager (1983). The implications can be interpreted as specific chains of reasoning. Giles (1976) gives a nice interpretation of these chains of reasoning in terms of risk: every chain of reasoning is analogous to a dialogue between speakers whose assertions entail a commitment about their truth.

The subjectivity that exists in fuzzy modeling is a blessing rather than a curse. The vagueness present in the definition of terms is consistent with the information contained in the conditional rules developed by the engineer when observing some complex process. Even though the set of linguistic variables and their meanings is compatible and consistent with the set of conditional rules used, the overall outcome of the qualitative process is translated into objective and quantifiable results. Fuzzy mathematical tools and the calculus of fuzzy IF–THEN rules provide a most useful paradigm for the automation and implementation of an extensive body of human knowledge heretofore not embodied in the quantitative modeling process; we call this paradigm *fuzzy systems*. These mathematical tools provide a means of sharing, communicating, and transferring this human subjective knowledge of systems and processes.

This chapter has also summarized the seminal works of Zadeh (1972, 1973, 1975a, b) in the area of linguistic modeling. Modeling in the area of linguistics has reached far beyond the boundaries of engineering. For example, Kickert (1979) used fuzzy linguistic modeling to adapt a factual prototype of Mulder's power theory to a numerical simulation. This is a marvelous illustration of the power of fuzzy sets in a situation where ideals of the *rational man* run contrary to the satisfaction gained simply through the exercise of power.

In a recent work, a Sugeno inference model was developed by Ali, Reda Taha, Thornton, Shrive, and Frank (2005) to quantify the level of recruitment (mobilization) of fibrils in medial collateral ligaments. Creep experiments on New Zealand rabbits were analyzed using the Sugeno system. The proposed model was capable of relating creep compliance to creep time and creep stress through a fuzzy interpretation of ligament mobilization, and was shown necessary for the design of human artificial ligaments.

This chapter has focused on two popular forms of logic, classical and fuzzy. There are other forms, of course. For example, almost a century ago, L. E. J. Brouwer posed a form of logic known as *intuitionism*. This logic has been subject to debate for all of this time (Franchella, 1995). Intuitionism is a branch of logic, which stresses that mathematics has priority over logic, the objects of mathematics are constructed and operated on in the mind by the mathematician, and it is impossible to define the properties of mathematical objects simply by establishing a number of axioms. In particular, intuitionists reject, as they called it then, the *principle of the excluded middle* (we refer to this as the *excluded middle axiom* in this text), which allows proof by contradiction.

Brouwer rejected in mathematical proofs the *principle of the excluded middle*, which states that any mathematical statement is either true or false. In 1918, he published a set theory, in 1919 a measure theory, and in 1923 a theory of functions all developed without using this principle.

In 1928 Brouwer's paper, "Reflections on Formalism," four key differences between formalism and intuitionism have been identified and discussed, all having to do either with the role of the *principle of the excluded middle* or with the relation between mathematics and language. Brouwer emphasizes, as he had done in his dissertation in 1907, that formalism presupposes contextual mathematics at the metalevel. In this paper Brouwer presents his first strong counterexample of the *principle of the excluded middle*, by showing that it is false that every real number is either rational or irrational. An illustration of this is the following: A is a statement: "π has infinitely many 7s in its expansion" and $\overline{\mathrm{A}}$ is a statement: "π has only finitely many 7s in its expansion." We do not know whether A is true or false, so we cannot claim that A or $\overline{\mathrm{A}}$ is true, because that would imply that we either know $\overline{\mathrm{A}}$ or we know A (V. Kreinovich, personal communication, 2003).

Brouwer also considered weak counterexamples to the *principle of the excluded middle*. A still open problem in mathematics, known as *Goldbach's conjecture* (the conjecture that every even number equal to or greater than four is the sum of two prime numbers), is one such counterexample. The conjecture Brouwer states, "we have at present experienced neither its truth nor its falsity, so intuitionistically speaking, it is at present neither true nor false, and hence we cannot assert 'Goldbach's conjecture is true, or it is false'" (Franchella, 1995).

More simple examples where the excluded axiom does not apply are instructive. For example, “The door is not unlocked” does not necessarily imply that the door is locked (the door might not have a lock; therefore it is false that the door can be locked). Or, Schrödinger’s cat that is alive and dead at the same time. Jan Lukawiewicz (1920) argued that the law of the excluded middle belongs to two-valued logics (such as classical predicate logic) and such systems are not expressive enough to model more complex dependencies. One of his examples was that expressions that involve *time* might be indeterminate because some events didn’t happen yet, and so we can’t determine if these are true or false. This was one of the reasons behind his three-valued logic. Aristotle offered the notion that the state of “not-proven” exists between “guilty” and “not guilty”; in this case the intersection is not null.

Another form of logic is termed *linear logic*, where we have two different versions of conjunction. For example, in the phrase *I can buy a snack and a drink*, we can mean that we can only buy one, not both, or that we can buy both. Both forms of the conjunction are allowed (V. Kreinovich, personal communication, 2003).

In Brouwer’s intuitionism there is a single description of the connective, “or.” In intuitionism (also termed *constructive logic*) the meaning of “or” is as follows: the statement “A or B” means that either we know A or we know B. In a nonconstructive logic, the statement “A or B” means that we know that one or the other (A or B) is true, but we do not know which one is true. In classical logic, we have both types of “or.” What Brouwer pointed out is that if we interpret the “or” as a *constructive or*, then the excluded middle axiom is not valid.

The significance of other forms of logic is that we oftentimes intertwine our human intuition with formal logic structures that are likely *layered* in our minds, just like the laws of nature are *layered* in reality. For example, sometimes we use Newton’s laws to describe behavior in mechanics that we can see visually, yet the phenomena might better be described by quantum mechanics laws at a scale that is not known to us through simple observation. The material developed in this chapter provides a good foundation for discussions in Chapter 8 on nonlinear simulation and in Chapter 11 on fuzzy control.

References

Ali, A., Reda Taha, M. M., Thornton, G. M., Shrive, N. G., and Frank, C. B. (2005). A biomechanical study using fuzzy systems to quantify collagen fibre recruitment and predict creep of the rabbit medial collateral ligament. *J. Biomech. Eng.*, **127**, 484–493.

Belohlavek, R., Klir, G., Lewis, H., and Way, E. (2002). On the capability of fuzzy set theory to represent concepts. *Int. J. Gen. Syst.*, **31**, 569–585.

Clark, D. W. (1992). Computer illogic. *Mirage Mag.*, Fall, 12–13.

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications* (New York: Academic Press).

Franchella, M. (1995). L. E. J. Brouwer: Toward intuitionistic logic. *Hist. Math.*, **22**, 304–322.

Gaines, B. (1976). Foundations of fuzzy reasoning. *Int. J. Man Mach. Stud.*, **8**, 623–688.

Giles, R. (1976). Lukasiewicz logic and fuzzy theory. *Int. J. Man Mach. Stud.*, **8**, 313–327.

Gill, A. (1976). *Applied Algebra for the Computer Sciences* (Englewood Cliffs, NJ: Prentice Hall).

Jang, J., Sun, C., and Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, 2nd ed., (Upper Saddle River, NJ: Pearson Education, Inc.).

Kickert, W. (1979). An example of linguistic modelling: The case of Mulder’s theory of power, in M. Gupta, R. Ragade, and R. Yager, eds., *Advances in Fuzzy Set Theory and Applications*, pp. 519–540, (Amsterdam: Elsevier).

Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).

Kosko, B. (1992). *Neural Networks and Fuzzy Systems* (Englewood Cliffs, NJ: Prentice Hall).

Leibnez, G. W. (1677/1951). “Preface to the general science,” in Philip Wiener, *Leibnez: Selections* (New York: Scribner’s).
Łukasiewicz J. (1920). On three-valued logic [Polish], *Ruch filozoficzny* **5**, 170–171. in L. Borkowski, ed., *Selected works by Jan Łukasiewicz*, pp. 87–88 (Amsterdam: North Holland, 1970).
Mamdani, E. H. (1976). Advances in linguistic synthesis of fuzzy controllers. *Int. J. Man Mach. Stud.*, **8**, 669–678.
Mamdani, E., and Assilian, S. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man Mach. Stud.*, **7**, 1–13.
Mano, M. (1988). *Computer Engineering: Hardware Design* (Englewood Cliffs, NJ: Prentice Hall).
Mizumoto, M., and Zimmerman, H. J. (1982). Comparison of fuzzy reasoning methods. *Fuzzy Sets Syst.*, **8**, 253–283.
Osherson, D., and Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, **9**, 35–58.
Rescher, N. (1969). *Many-valued Logic* (New York: McGraw-Hill).
Ross, T. (1995). *Fuzzy Logic with Engineering Applications* (New York: McGraw-Hill).
Sanchez, E. (1976). Resolution of composite fuzzy relation equations. *Inform. Contr.*, **30**, 38–48.
Smith, E., and Osherson, D. (1984). Conceptual combinations with prototpyes concepts, *Cog. Sci.*, **8**, 337–361.
Sugeno, M., and Kang, G. (1988). Structure identification of fuzzy model. *Fuzzy Sets Syst.*, **28**, 15–33.
Takagi, T., and Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst., Man, Cybern.*, **15**, 116–132.
Tsukamoto, Y. (1979). An approach to fuzzy reasoning method, in M. Gupta, R. Ragade, and R. Yager, eds., *Advances in Fuzzy Set Theory and Applications*, pp. 137–149, (Amsterdam: Elsevier).
Yager, R. R. (1983). On the implication operator in fuzzy logic. *Inf. Sci.*, **31**, 141–164.
Yager, R. R. (1985). Strong truth and rules of inference in fuzzy logic and approximate reasoning. *Cybern. Syst.*, **16**, 23–63.
Zadeh, L. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges. *J. Cybern.*, **2**, 4–34.
Zadeh, L. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man, Cybern.*, **3**, 28–44.
Zadeh, L. (1975a). The concept of a linguistic variable and its application to approximate reasoning—I. *Inf. Sci.*, **8**, 199–249.
Zadeh, L. (1975b). The concept of a linguistic variable and its application to approximate reasoning—II. *Inf. Sci.*, **8**, 301–357.
Zadeh, L. (1976). The concept of a linguistic variable and its application to approximate reasoning—Part 3. *Inf. Sci.*, **9**, 43–80.
Zadeh, L. (1979). A theory of approximate reasoning, in J. Hayes, D. Michie, and L. Mikulich, eds., *Machine Intelligence*, pp. 149–194 (New York: Halstead Press).

Problems

5.1 Under what conditions of P and Q is the implication $P \rightarrow Q$ a tautology?

5.2 The “exclusive or” is given by the expression $P\,XOR\,Q = (\overline{P} \wedge Q) \vee (P \wedge \overline{Q})$.

Show that the “logical or” given by $P \vee Q$, gives a different result from the “exclusive or,” and comment on this difference using an example in your own field.

5.3 For a proposition R of the form $P \rightarrow Q$, show the following:

a. R and its contrapositive are equivalent, that is, prove that $(P \rightarrow Q) \leftrightarrow (\overline{Q} \rightarrow \overline{P})$.
b. The converse of R and the inverse of R are equivalent, that is, prove that $(Q \rightarrow P) \leftrightarrow (\overline{P} \rightarrow \overline{Q})$.

5.4 Show that the dual of the equivalence $((P \vee Q) \vee (\overline{P} \wedge \overline{Q})) \leftrightarrow X$ is also true.

5.5 Show that De Morgan’s principles are duals.

5.6 Show that the compound proposition $((P \rightarrow Q) \wedge (R \rightarrow \overline{S}) \wedge (Q \rightarrow R)) \rightarrow (P \rightarrow \overline{S})$ is a tautology.

5.7 Show that the following propositions from Lewis Carroll are tautologies (Gill, 1976):

a. No ducks waltz; no officers ever decline to waltz; all my poultry are ducks; therefore, none of my poultry are officers.
b. Babies are illogical; despised persons cannot manage crocodiles; illogical persons are despised; therefore, babies cannot manage crocodiles.
c. Promise-breakers are untrustworthy; wine-drinkers are communicative; a man who keeps his promise is honest; all pawnbrokers are wine-drinkers; we can always trust a communicative person; therefore, all pawnbrokers are honest. (This problem requires $2^6 = 64$ lines of a truth table; perhaps it could be solved with a computer).

5.8 Prove the following statements by contradiction.

a. $((\mathrm{P} \to \mathrm{Q}) \wedge \mathrm{P}) \to \mathrm{Q}$
b. $\left((\mathrm{P} \to \overline{Q}) \wedge (\mathrm{Q} \vee \overline{R}) \wedge (\mathrm{R} \wedge \overline{S})\right) \to \overline{P}$

5.9 Prove that $\left((\mathrm{P} \to \overline{Q}) \wedge (\mathrm{R} \to \overline{Q}) \wedge (\mathrm{P} \vee \mathrm{R})\right) \to \mathrm{R}$ is not a tautology (i.e., a fallacy) by developing a counterexample.

5.10 Prove that the following statements are tautologies.

a. $((\mathrm{P} \to \mathrm{Q}) \wedge \mathrm{P}) \to \mathrm{Q}$
b. $\mathrm{P} \to (\mathrm{P} \vee \mathrm{Q})$
c. $(\mathrm{P} \wedge \mathrm{Q}) \to \mathrm{P}$
d. $\left((\mathrm{P} \vee \mathrm{Q}) \wedge \overline{P}\right) \to \mathrm{Q}$.

5.11 For this inference rule, $[(\mathrm{A} \to \mathrm{B}) \wedge (\mathrm{B} \to \mathrm{C})] \to (\mathrm{A} \to \mathrm{C})$, prove that the rule is a tautology.

5.12 Consider the following two discrete fuzzy sets, which are defined on a universe $\mathrm{X} = [0, 9]$

a. $\underset{\sim}{A} = \left\{\frac{0.4}{0} + \frac{0.7}{1} + \frac{0.2}{2} + \frac{0.1}{3} + \frac{0.3}{4}\right\}$ $\underset{\sim}{B} = \left\{\frac{0}{5} + \frac{0.5}{6} + \frac{0.6}{7} + \frac{0.3}{8} + \frac{0.7}{9}\right\}$
b. Construct the relation for the rule. If $\underset{\sim}{A}$, THEN $\underset{\sim}{B}$ using Mamdani implication, that is, using $\mu_{\underset{\sim}{R}}(\mathrm{x}, \mathrm{y}) = \min\left[\mu_{\underset{\sim}{A}}(\mathrm{x}), \mu_{\underset{\sim}{B}}(\mathrm{y})\right]$.
c. If we introduce a new antecedent $\underset{\sim}{A}' = \left\{\frac{0.3}{0} + \frac{0.7}{1} + \frac{0.6}{2} + \frac{0.3}{3} + \frac{0.5}{4}\right\}$ find the new consequent B', using max-min composition; that is find B' = A'oR for R from part(a)

5.13 Given the fuzzy sets A and B on X and Y, respectively,

$$\underset{\sim}{A} = \int\left\{\frac{1-0.1x}{x}\right\}, \textit{ for } x \in [0, 4]$$

$$\underset{\sim}{B} = \int\left\{\frac{0.25y}{y}\right\}, \textit{ for } y \in [1, 4]$$

a. Construct a fuzzy relation $\underset{\sim}{R}$ for the implication $\underset{\sim}{A} \to \underset{\sim}{B}$ using the classical implication operation construct, that is, $\underset{\sim}{R} = (\underset{\sim}{A} \times \underset{\sim}{B}) \cup (\underset{\sim}{A} \times Y)$.

b. Use max–min composition to find $\underset{\sim}{B}'$, given $\underset{\sim}{A}' = \left\{ \frac{0.6}{2} \right\}$, a singleton at point 2 that does not have full membership.

5.14 Rail track is monitored to measure the thermal expansion. Data is recorded for one whole day and made a group of sets for different temperature. It is known that expansion or contraction of rail track is directly related to temperature. Two fuzzy sets for temperature $\underset{\sim}{T}$ (°F) and Elongation, $\underset{\sim}{\Delta L}$ (in) are given for the recording made from monitoring till night.

$$\underset{\sim}{T} = \left\{ \frac{0.2}{65} + \frac{0.3}{75} + \frac{0.5}{85} + \frac{0.6}{95} + \frac{0.8}{105} \right\}$$

$$\underset{\sim}{\Delta L} = \left\{ \frac{0.1}{0.1} + \frac{0.3}{0.2} + \frac{0.5}{0.3} + \frac{0.8}{0.4} + \frac{0.7}{0.5} \right\}$$

a. Using the proposition, If x is $\underset{\sim}{T}$, then Y is $\underset{\sim}{\Delta L}$ find this relation using the following forms of the implication $\underset{\sim}{T} \to \underset{\sim}{\Delta L}$
   i. Classical $\mu_{\underset{\sim}{R}} = \max[\min(\mu_{\underset{\sim}{T}}, \mu_{\underset{\sim}{\Delta L}}), (1-\mu_{\underset{\sim}{T}})]$
   ii. Mamdani $\mu_{\underset{\sim}{R}} = \min(\mu_{\underset{\sim}{T}}, \mu_{\underset{\sim}{\Delta L}})$
   iii. Product $\mu_{\underset{\sim}{R}} = \mu_{\underset{\sim}{T}} \cdot \mu_{\underset{\sim}{\Delta L}}$

b. Now define another antecedent, say $\underset{\sim}{\Delta L}'$ = "contraction"
   i. $\underset{\sim}{\Delta L}' = \left\{ \frac{0.5}{0.1} + \frac{0.8}{0.2} + \frac{0.7}{0.3} + \frac{0.2}{0.4} + \frac{0.0}{0.5} \right\}$
   ii. Find the temperature change using from higher to lower as a result of contraction using
   iii. Classical implication from (a)

c. Max-min Composition
   i. Max-product Composition

5.15 An outlet in a dam is regulated to maintain the level of water. Gates are opened once water reaches more than head level. Let X be the universe of excess level of water in feet X = {10, 20, 30, 40, 50} and Y be a universe of discharge in cubic feet $Y = \{5 \times 10^4, 10 \times 10^4, 15 \times 10^4, 20 \times 10^4, 25 \times 10^4\}$. The fuzzy sets for both parameters are given as follows:

$$\mu_{\underset{\sim}{H}}(x) = \left\{ \frac{0.7}{10} + \frac{0.6}{20} + \frac{0.3}{30} + \frac{0.4}{40} + \frac{0.2}{50} \right\}$$

$$\mu_{\underset{\sim}{D}}(y) = \left\{ \frac{0.8}{5x10^4,} + \frac{0.5}{10x10^4,} + \frac{0.6}{15x10^4,} + \frac{0.4}{20x10^4,} + \frac{0.1}{25x10^4,} \right\}$$

a. Use classical implication to find the relation IF high water level, THEN more discharge

b. Suppose we are given a new water level as $\mu'_{\underset{\sim}{H}}(x) = \left\{\frac{0.7}{10} + \frac{0.3}{20} + \frac{0.5}{30} + \frac{0.6}{40} + \frac{0.4}{50}\right\}$

Using max-min composition, find the discharge associated with this new water level.

5.16 We want to consider the engineering of amplifiers. Here, the amplifier is a simple voltage-measuring input and current output, as shown in Figure P5.16. We define two fuzzy linguistic variables for a fuzzy relation: $\underset{\sim}{V}_{in}$, the input voltage, and $\underset{\sim}{I}_{out}$, the output current:

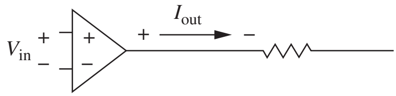

Figure P5.16

$$\underset{\sim}{V}_{in} = \text{“small”} = \left\{\frac{1.0}{0.1} + \frac{0.9}{0.2} + \frac{0.8}{0.3} + \frac{0.4}{0.4} + \frac{0.3}{0.5}\right\} \text{ (volts)}$$

$$\underset{\sim}{I}_{out} = \text{“big”} = \left\{\frac{0.4}{0.7} + \frac{0.6}{1.0} + \frac{0.8}{1.3} + \frac{1.0}{1.6}\right\} \text{ (amperes)}$$

a. Find the relation, IF $\underset{\sim}{V}_{in}$, THEN $\underset{\sim}{I}_{out}$, using classical implication.

b. Another fuzzy linguistic variable in this problem is input impedance, $\underset{\sim}{Z}$. The higher the impedance, generally the better the amplifier. For the following impedance defined on a universe of resistances,

$$\underset{\sim}{Z} = \text{“high impedance”} = \left\{\frac{0.0}{10^3} + \frac{0.2}{10^4} + \frac{0.5}{10^5} + \frac{0.8}{10^6} + \frac{1.0}{10^7}\right\} \text{ (ohms)}$$

Find the relation, IF $\underset{\sim}{Z}$, THEN $\underset{\sim}{I}_{out}$, using Mamdani implication.

5.17 For sets

$$\underset{\sim}{A} = \left\{\frac{0.7}{1} + \frac{0.8}{2} + \frac{0.5}{3} + \frac{0.3}{4} + \frac{0.2}{5}\right\} \quad \underset{\sim}{B} = \left\{\frac{0.3}{10} + \frac{0.5}{20} + \frac{0.7}{30} + \frac{0.9}{40}\right\}$$

Calculate the fuzzy relation $\underset{\sim}{R}$ using Equations [5.19 to 5.23]

$$\mu_{\underset{\sim}{R}}(x, y) = \max\left[\mu_{\underset{\sim}{B}}(y), 1 - \mu_{\underset{\sim}{A}}(x)\right] \tag{5.19}$$

$$\mu_{\underset{\sim}{R}}(x, y) = \min\left[\mu_{\underset{\sim}{A}}(x), \mu_{\underset{\sim}{B}}(y)\right] \tag{5.20}$$

$$\mu_{\underset{\sim}{R}}(x, y) = \min\left\{1, \left[1 - \mu_{\underset{\sim}{A}}(x) + \mu_{\underset{\sim}{B}}(y)\right]\right\} \tag{5.21}$$

$$\mu_{\underset{\sim}{R}}(x, y) = \mu_{\underset{\sim}{A}}(x) \cdot \mu_{\underset{\sim}{B}}(y) \tag{5.22}$$

$$\mu_{\underset{\sim}{R}}(x, y) = \begin{cases} 1 & \text{for } \mu_{\underset{\sim}{A}}(x) \le \mu_{\underset{\sim}{B}}(y) \\ \mu_{\underset{\sim}{B}}(y) & \text{Otherwise} \end{cases} \tag{5.23}$$

5.18 Fill in the following table using the five equations from problem 5.17, one equation for each row of the table, to determine the values of the implication $\underset{\sim}{A} \to \underset{\sim}{B}$. Comment on the similarities and dissimilarities of the various implication methods with respect to the various values for $\underset{\sim}{A}$ and $\underset{\sim}{B}$.

$\underset{\sim}{A}$	$\underset{\sim}{B}$	$\underset{\sim}{A} \to \underset{\sim}{B}$
0		1
1		0
0.5		0.5
0.2		0.7
0.8		0.4

5.19 The weather department collected data of rainfall during the summer and the monsoon season. Because summer experiences less rainfall than the monsoon, the summer is categorized as 'Low," and monsoon is categorized as "High." Two parameters are made for rainfall, "precipitation" (inches) and "intensity" (inches/hour). We characterize each parameter in fuzzy linguistic terms as follows

$$\text{"Low } \underset{\sim}{\text{precipitation}}\text{"} = \left\{ \frac{0.8}{20} + \frac{0.6}{30} + \frac{0.5}{40} + \frac{0.2}{50} + \frac{0.1}{60} \right\}$$

$$\text{"High } \underset{\sim}{\text{precipitation}}\text{"} = \left\{ \frac{0.2}{20} + \frac{0.3}{30} + \frac{0.6}{40} + \frac{0.8}{50} + \frac{1.0}{60} \right\}$$

$$\text{"Low } \underset{\sim}{\text{Intensity}}\text{"} = \left\{ \frac{0.9}{2.5} + \frac{0.7}{5.0} + \frac{0.4}{7.5} + \frac{0.2}{10.0} + \frac{0.1}{12.5} \right\}$$

$$\text{"High } \underset{\sim}{\text{Intensity}}\text{"} = \left\{ \frac{0.3}{2.5} + \frac{0.2}{5.0} + \frac{0.5}{7.5} + \frac{0.8}{10.0} + \frac{0.9}{12.5} \right\}$$

Find the following membership functions using linguistic hedges

i. Very low precipitation
ii. Very high Intensity
iii. Not very high precipitation
iv. Not very high and not very low Intensity
v. Slightly high precipitation or slightly low precipitation

5.20 Testing of foundations under two different soils for the same axial loading was done to find the settlement conditions. Experiments gave "less settlement" for a type 1 soil and "Large settlement" for a type 2 soil. X is the universe for settlement in inches. Let us characterize settlement linguistically with two terms “less settlement” and “large settlement.”

$$\text{“Less } \underset{\sim}{\text{settlement}}\text{”} = \left\{ \frac{0.8}{0} + \frac{0.7}{0.5} + \frac{0.6}{1.0} + \frac{0.3}{1.5} + \frac{0.2}{2.0} \right\}$$

$$\text{“Large } \underset{\sim}{\text{settlement}}\text{”} = \left\{ \frac{0.2}{0} + \frac{0.3}{0.5} + \frac{0.5}{1.0} + \frac{0.8}{1.5} + \frac{0.9}{2.0} \right\}$$

Find the membership function for the following linguistic expressions:

a. Very less settlement
b. Slightly large settlement
c. Very, very large settlement
d. Not slightly less and very, very large settlement

5.21 For two pressure heads in a manometer readings show 30 cm and 120 cm. The atmospheric pressure correlated to these readings was found to be 80 kPa. Write a simple rule that relates these parameters linguistically.

5.22 Two compounds, Ca^{2+} and $So^{2-}{}_4$, react to give a salt, $CaSO_4$. Write a simple rule that relates these parameters linguistically.

5.23 In Example 5.16, recalculate the response function shown in Figure 5.15b using the following membership function shapes:

a. Two triangles for the input X and two triangles for the output Y;
b. Two trapezoids for the input X and two trapezoids for the output Y.

5.24 In Example 5.17, recalculate the response function shown in Figure 5.17d using the following membership function shapes for the inputs:

a. triangles for small, medium, large
b. trapezoids for small, medium, large.

5.25 From thermodynamics it is known that for an ideal gas in an adiabatic reversible process

$$\frac{T_2}{T_1} = \left(\frac{P_2}{P_1}^{\frac{\gamma-1}{\gamma}} \right)$$

where T_1 and T_2 are temperatures in kelvin (K) and P_1 and P_2 are pressures in bars, for an ideal gas.

To pursue a Sugeno solution, we will use the following functions for the consequents of the three rules:

Rule 1: $T_1 = 320$ K and $\gamma = 1.5$
Rule 2: $T_1 = 300$ K and $\gamma = 1.4$
Rule 3: $T_1 = 300$ K and $\gamma = 1.3$

For the Sugeno solution, T_1 will be fixed at 300 K and the fuzzy model will predict P_2 for the given input variables P_1 and T_2. In other words, we are interested in finding the final pressure, P_2, of the system if the temperature of the system is changed to T_2 from an original pressure equal to P_1. A real application could use a similar model built from experimental data to do a prediction on non-ideal gases.

For a Tsukamoto solution the rules to be used are:

Rule 1: IF P_1 = atmP AND T_2 = lowT THEN P_2 = lowP.
Rule 2: IF P_1 = atmP AND T_2 = midT THEN P_2 = lowP.
Rule 3: IF P_1 = lowP AND T_2 = lowT THEN P_2 = very highP.

Given the rule-base, the membership functions shown in Figure P5.25, and the following pair of input values, $P_1 = 1.6$ bar and $T_2 = 415$ K, conduct a simulation to determine P_2 for the two inference methods of Sugeno and Tsukamoto. For another solution approach the Sugeno consequents could use the ideal gas formula constant $\gamma = 1 + \dfrac{R}{Cv} = 1.4$.

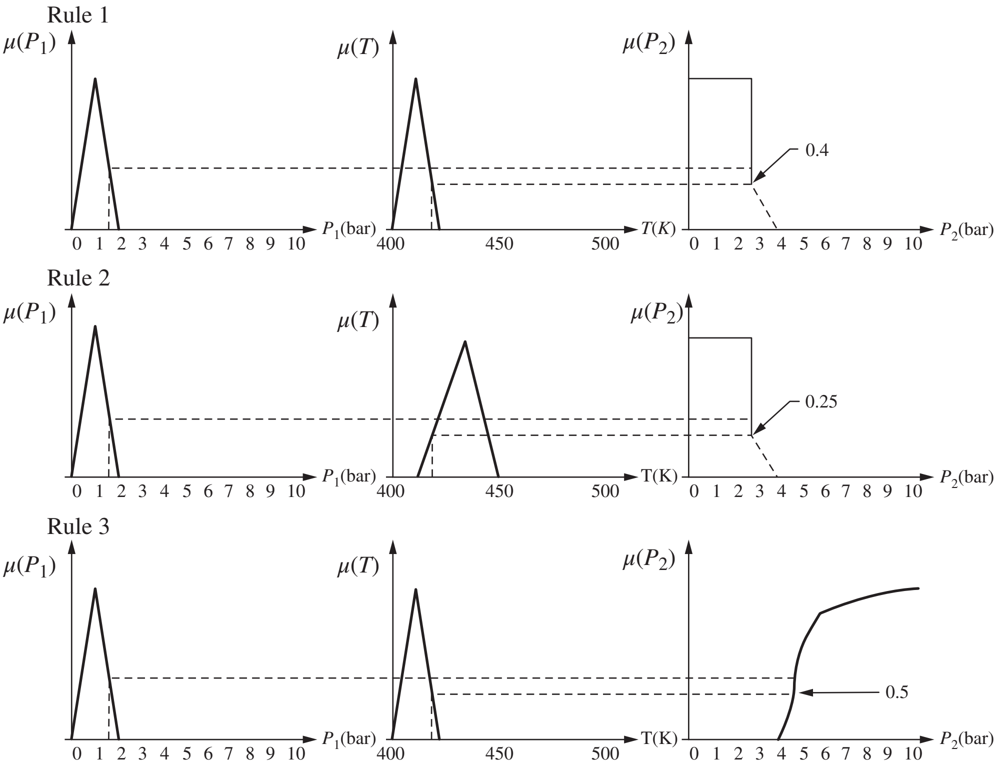

Figure P5.25

5.26 In finding the Nusselt number (a dimensionless number for determining heat transfer) for a hexagonal cylinder in cross flow, there are two correlations (which are to be used as the consequent terms in a Sugeno inference method):

$Nu1 = 0.16\ Re^{0.638}\ Pr^{1/3}$	$5000 < Re < 19650,$
$Nu2 = 0.0385\ Re^{0.638}\ Pr^{1/3}$	$Re > 19650,$

Re is the Reynolds number and Pr is the Prandtl number. In the equations, we seek to know whether Nu is low ($Nu1$) or is Nu medium ($Nu2$)?

The Nusselt number is a function of the convective heat transfer (h), diameter of the hexagonal cylinder (D) over which cooling fluid travels, and the conductivity of the material (K):

$$Nu = \frac{hD}{K}$$

Both Re and Pr can be fuzzy because of uncertainty in the variables in velocity. It would be convenient to find Nu (output) based on Re and Pr (inputs) without having to do all the calculations. More specifically, there is uncertainty in calculating the Reynolds number because velocity is not known exactly:

$$\mathrm{Re} = \frac{\rho VD}{\mu},$$

where ρ is the density, V is the velocity, D is the characteristic length (or pipe diameter), and μ is the dynamic viscosity. And there is also uncertainty in the value for the Prandtl number due to its constituents

$$\mathrm{Pr} = \frac{\nu}{\alpha}$$

where ν is the kinematic viscosity and α is the specific gravity.

Calculation of Nu is involved and the incorporation of a rule-base can be used to bypass these calculations; we have the following rules to govern this process:

If Re is high and Pr is low $\rightarrow$ Then Nu is low.
If Re is low and Pr is low $\rightarrow$ Then Nu is low.
If Re is high and Pr is high $\rightarrow$ Then Nu is medium.
If Re is low and Pr is high $\rightarrow$ Then Nu is medium.

For this problem, conduct a Mamdani and a Sugeno inference, based on the membership functions given in Figure P5.26 (a–c), and use the following inputs:

$$Re = 1.965 \times 10^4.$$
$$\mathrm{Pr} = 275.$$

Comment on the differences in the results.

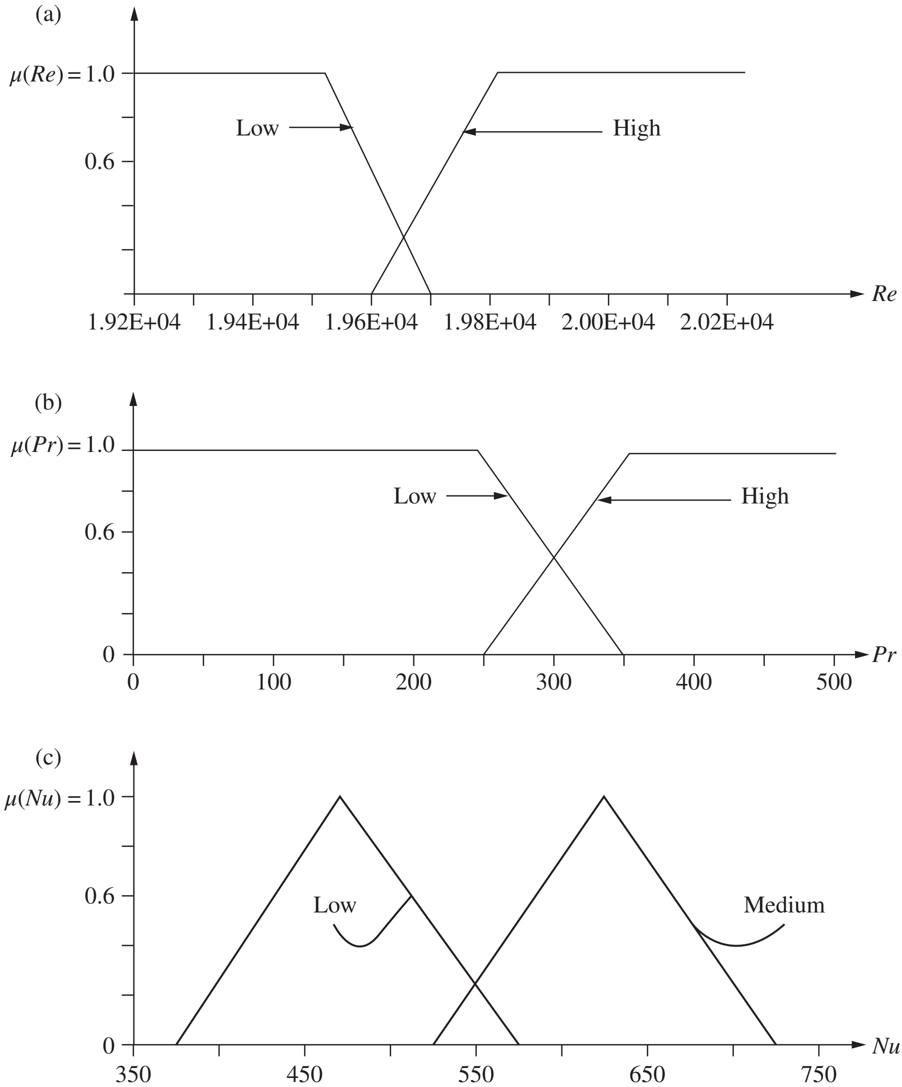

Figure P5.26

5.27 Prove, using a symbolic approach, a truth table, or a graphical approach similar to the diagrams in Figure 5.3, that Equation 5.5 is equivalent to Equation 5.6.

6

Historical Methods of Developing Membership Functions

So far as the laws of mathematics refer to reality, they are not certain. And so far as they are certain, they do not refer to reality.

Albert Einstein,
theoretical physicist and Nobel Laureate "Geometrie und Erfahrung,"
Lecture to Prussian Academy, 1922

Uncertainty is the refuge of hope.

Henri Frederic Amiel,
nineteenth-century Swiss philosopher.

The quest for certainty blocks the search for meaning. Uncertainty is the very condition to impel man to unfold his powers.

Erich Fromm,
twentieth-century psychologist

Not to be absolutely certain is, I think, one of the essential things in rationality.

Bertrand Russell,
twentieth-century British philosopher

The statements from Albert Einstein to Bertrand Russell attest to the fact that few things in real life are certain or can be conveniently reduced to the axioms of mathematical theories and models. A metaphorical expression that represents this idea is known as the *Law of Probable Dispersal*; to wit, "Whatever it is that hits the fan will not be evenly distributed." As this enlightened law implies, most things in nature cannot be characterized by simple or convenient shapes or distributions. Membership functions characterize the fuzziness in a fuzzy set—whether the elements in the set are discrete or continuous—in a graphical form for eventual use in the mathematical formalisms of fuzzy set theory. But the shapes used to describe the fuzziness have few restrictions indeed; some of these have been described in Chapters 1 and 4. Just as there are an infinite number of ways to characterize fuzziness, there are an infinite number of ways to graphically depict the membership functions that describe this fuzziness. This chapter describes a few procedures to develop these membership functions based on deductive intuition or numerical data; Chapter 7 develops this idea further with an explanation of additional procedures that build membership functions and deductive rules from measured observations of systems.

Because the membership function essentially embodies all fuzziness for a particular fuzzy set, its description is the essence of a fuzzy property or operation. Because of the importance of the "shape" of the membership function, a great deal of attention has been focused on development of these functions. This chapter describes, then illustrates, six procedures that have been used to build membership functions. There are many more; references at the end of this chapter can be consulted on this topic.

Membership Value Assignments

There are possibly more ways to assign membership values or functions to fuzzy variables than there are to assign probability density functions to random variables (see Dubois and Prade, 1980). This assignment process can be intuitive, or it can be based on some algorithmic or logical operations. The following is a list of six straightforward methods described in the literature to assign membership values or functions to fuzzy variables. Each of these methods will be illustrated in simple examples in this chapter. The literature on this topic is rich with references, and a short list of those consulted is provided in the summary of this chapter.

1. Intuition
2. Inference
3. Rank ordering
4. Neural networks
5. Genetic algorithms
6. Inductive reasoning.

Intuition

This method needs little or no introduction. It is simply derived from the capacity of humans to develop membership functions through their own innate intelligence and understanding. Intuition involves contextual and semantic knowledge about an issue; it can also involve linguistic truth values about this knowledge (see Zadeh, 1972). As an example, consider the membership functions for the fuzzy variable temperature. Figure 6.1 shows various shapes on the universe of temperature as measured in units of degrees Celsius. Each curve is a membership function corresponding to various fuzzy variables, such as cold, cool, warm, and hot. Of course, these curves are a function of context and the analyst developing them. For example, if the temperatures are referred to the range of human comfort we get one set of curves, and if they are referred to the range of safe operating temperatures for a steam turbine we get another set. However, the important character of these curves for purposes of use in fuzzy operations is the fact that they overlap. In numerous examples throughout the rest of this text we shall see that the precise shapes of these curves are not so important in their utility. Rather, it is the approximate

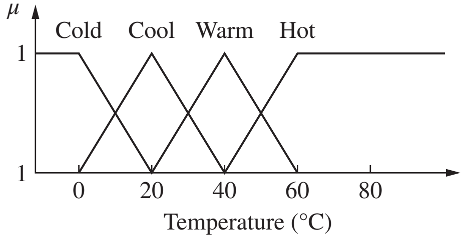

Figure 6.1 Membership functions for the fuzzy variable "temperature."

placement of the curves on the universe of discourse, the number of curves (partitions) used, and the overlapping character that are the most important ideas.

Inference

In the inference method we use knowledge to perform deductive reasoning. That is, we wish to deduce or infer a conclusion, given a body of facts and knowledge. There are many forms of this method documented in the literature, but the one we illustrate here relates to our formal knowledge of geometry and geometric shapes, similar to ideas posed in Chapter 1.

In the identification of a triangle, let A, B, and C be the inner angles of a triangle, in the order $A \geq B \geq C \geq 0$, and let U be the universe of triangles, that is,

$$U = \{(A, B, C) | A \geq B \geq C > 0; A + B + C = 180°\}. \tag{6.1}$$

We define a number of geometric shapes that we wish to be able to identify for any collection of angles fulfilling the constraints given in Equation (6.1). For this purpose, we will define the following five types of triangles:

$\underset{\sim}{\mathrm{I}}$ Approximate isosceles triangle
$\underset{\sim}{\mathrm{R}}$ Approximate right triangle
$\underset{\sim}{\mathrm{IR}}$ Approximate isosceles *and* right triangle
$\underset{\sim}{\mathrm{E}}$ Approximate equilateral triangle
$\underset{\sim}{\mathrm{T}}$ Other triangles.

We can infer membership values for all of these triangle types through the method of inference because we possess knowledge about geometry that helps us to make the membership assignments. So we shall list this knowledge here to develop an algorithm to assist us in making these membership assignments for any collection of angles meeting the constraints of Equation (6.1).

For the approximate isosceles triangle, we have the following algorithm for the membership, again for the situation of $A \geq B \geq C > 0$ and $A + B + C = 180°$:

$$\mu_{\underset{\sim}{\mathrm{I}}}(A, B, C) = 1 - \frac{1}{60°} \min(A - B, B - C). \tag{6.2}$$

So, for example, if $A = B$ or $B = C$, the membership value in the approximate isosceles triangle is $\mu_{\underset{\sim}{\mathrm{I}}} = 1$; if $A = 120°$, $B = 60°$, and $C = 0°$, then $\mu_{\underset{\sim}{\mathrm{I}}} = 0$. For a fuzzy right triangle, we have

$$\mu_{\underset{\sim}{\mathrm{R}}}(A, B, C) = 1 - \frac{1}{90^\circ}|A - 90^\circ|. \tag{6.3}$$

For instance, when $A = 90^\circ$, the membership value in the fuzzy right triangle, $\mu_{\underset{\sim}{\mathrm{R}}} = 1$, or when $A = 180^\circ$, this membership vanishes, that is, $\mu_{\underset{\sim}{\mathrm{R}}} = 0$. For the case of an approximate isosceles *and* right triangle (there is only one of these in the crisp domain), we can find this membership function by taking the logical intersection (*and* operator) of the isosceles and right-triangle membership functions, or

$$\underset{\sim}{\mathrm{IR}} = \underset{\sim}{\mathrm{I}} \cap \underset{\sim}{\mathrm{R}},$$

which results in

$$\begin{aligned}\mu_{\underset{\sim}{\mathrm{IR}}}(A,B,C) &= \min\left[\mu_{\underset{\sim}{\mathrm{I}}}(A,B,C), \mu_{\underset{\sim}{\mathrm{R}}}(A,B,C)\right] \\ &= 1 - \max\left[\frac{1}{60^\circ}\min(A-B, B-C), \frac{1}{90^\circ}|A-90^\circ|\right].\end{aligned} \tag{6.4}$$

For the case of a fuzzy equilateral triangle, the membership function is given by

$$\mu_{\underset{\sim}{\mathrm{E}}}(A,B,C) = 1 - \frac{1}{180^\circ}(A - C). \tag{6.5}$$

For example, when $A = B = C$, the membership value is $\mu_{\underset{\sim}{\mathrm{E}}}(A,B,C) = 1$; when $A = 180^\circ$, the membership value vanishes, or $\mu_{\underset{\sim}{\mathrm{E}}} = 0$. Finally, for the set of "all other triangles" (all triangular shapes other than I, R, and E) we simply invoke the complement of the logical union of the three previous cases (or, from De Morgan's principles, Equation [2.13], the intersection of the complements of the triangular shapes),

$$\underset{\sim}{\mathrm{T}} = (\overline{\underset{\sim}{\mathrm{I}} \cup \underset{\sim}{\mathrm{R}} \cup \underset{\sim}{\mathrm{E}}}) = \bar{\mathrm{I}} \cap \bar{\mathrm{R}} \cap \bar{\mathrm{E}},$$

which results in

$$\begin{aligned}\mu_{\underset{\sim}{\mathrm{T}}}(A,B,C) &= \min\left\{1-\mu_{\underset{\sim}{\mathrm{I}}}(A,B,C), 1-\mu_{\underset{\sim}{\mathrm{E}}}(A,B,C), 1-\mu_{\underset{\sim}{\mathrm{R}}}(A,B,C)\right\} \\ &= \frac{1}{180^\circ}\min\{3(A-B), 3(B-C), 2|A-90^\circ|, A-C\}.\end{aligned} \tag{6.6}$$

Example 6.1 (Ross, 1995).
Define a specific triangle, as shown in Figure 6.2, with these three ordered angles:

$$\{\mathrm{X}: A = 85^\circ \geq B = 50^\circ \geq C = 45^\circ, \text{where } A + B + C = 180^\circ\}.$$

The membership values for the fuzzy triangle shown in Figure 6.2 for each of the fuzzy triangles types are determined from Equations (6.2) to (6.6), as listed here:

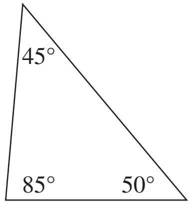

Figure 6.2 A specific triangle.

$$\mu_{\underset{\sim}{R}}(x) = 0.94,$$
$$\mu_{\underset{\sim}{I}}(x) = 0.916,$$
$$\mu_{\underset{\sim}{IR}}(x) = 0.916,$$
$$\mu_{\underset{\sim}{E}}(x) = 0.7,$$
$$\mu_{\underset{\sim}{T}}(x) = 0.05.$$

Hence, it appears that the triangle given in Figure 6.2 has the highest membership in the set of fuzzy right triangles, that is, in $\underset{\sim}{R}$. Notice, however, that the triangle in Figure 6.2 also has high membership in the isosceles triangle fuzzy set, and reasonably high membership in the equilateral fuzzy triangle set.

Rank Ordering

Assessing preferences by a single individual, a committee, a poll, and other opinion methods can be used to assign membership values to a fuzzy variable. Preference is determined by pairwise comparisons, and these determine the ordering of the membership. More is provided on the area of rank ordering in Chapter 9. This method is similar to a relative preferences method developed by Saaty (1974).

Example 6.2 (Ross, 1995).
Suppose 1000 people respond to a questionnaire about their pairwise preferences among five colors, X = {red, orange, yellow, green, blue}. Define a fuzzy set as $\underset{\sim}{A}$ on the universe of colors "best color." Table 6.1 is a summary of the opinion survey. In this table, for example, out of 1000 people 517 preferred the color red to the color orange, 841 preferred the color orange to the color yellow, and so on. Note that the color columns in the table represent an "antisymmetric" matrix. Such a matrix will be seen to relate to a *reciprocal relation*, which is introduced in Chapter 9. The total number of responses is 10 000 (10 comparisons). If the sum of the preferences of each color (row sum) is normalized to the total number of responses, a rank ordering can be determined as shown in the last two columns of the table.

If the percentage preference (the percentage column of Table 6.1) of these colors is plotted to a normalized scale on the universe of colors in an ascending order on the color universe, the membership function for "best color" shown in Figure 6.3 would result. Alternatively, the membership function could be formed based on the rank order developed (last column of Table 6.1).

Table 6.1 Example in rank ordering.

	Number who preferred							
	Red	Orange	Yellow	Green	Blue	Total	Percentage	Rank order
Red	–	517	525	545	661	2 248	22.5	2
Orange	483	–	841	477	576	2 377	23.8	1
Yellow	475	159	–	534	614	1 782	17.8	4
Green	455	523	466	–	643	2 087	20.9	3
Blue	339	424	386	357	–	1 506	15	5
Total						10 000		

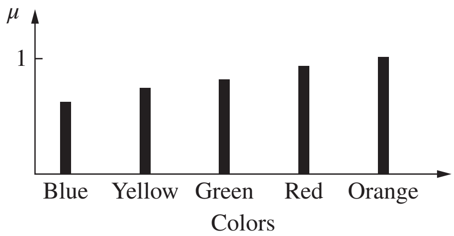

Figure 6.3 Membership function for best color.

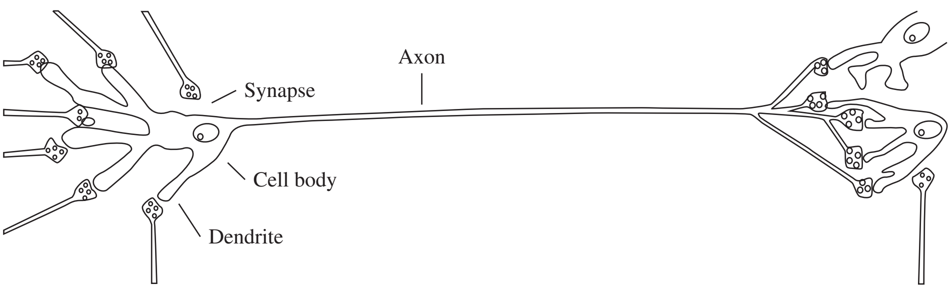

Figure 6.4 A simple schematic of a human neuron.

Neural Networks

In this section we explain how a neural network can be used to determine membership functions. We first present a brief introduction to neural networks and then show how they can be used to determine membership functions.

A neural network is a technique that seeks to build an intelligent program (to implement intelligence) using models that simulate the working network of the neurons in the human brain (Yamakawa, 1992; Hopfield, 1982; Hopfield and Tank, 1986). A neuron, Figure 6.4, is made up of several protrusions called *dendrites* and a long branch called the *axon*. A neuron is joined to other neurons through the dendrites. The dendrites of different neurons meet to form synapses, the areas where messages pass. The neurons receive the impulses via the synapses. If the total of the impulses received exceeds a certain threshold value, then the neuron sends

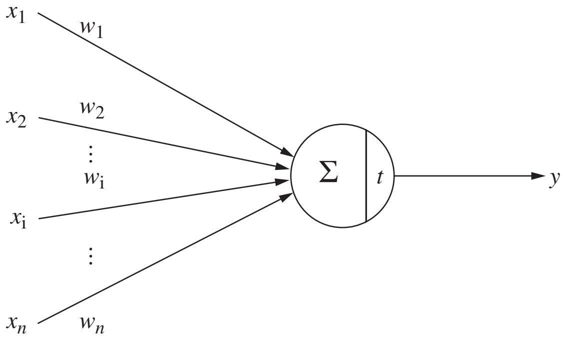

Figure 6.5 A threshold element as an analog to a neuron.

an impulse down the axon where the axon is connected to other neurons through more synapses. The synapses may be excitatory or inhibitory in nature. An excitatory synapse adds to the total of the impulses reaching the neuron, whereas an inhibitory neuron reduces the total of the impulses reaching the neuron. In a global sense, a neuron receives a set of input pulses and sends out another pulse that is a function of the input pulses.

This concept of how neurons work in the human brain is used in performing computations on computers. Researchers have long felt that the neurons are responsible for the human capacity to learn, and it is in this sense that the physical structure is being emulated by a neural network to accomplish machine learning. Each computational unit computes some function of its inputs and passes the result to connected units in the network. The knowledge of the system comes out of the entire network of the neurons.

Figure 6.5 shows the analog of a neuron as a threshold element. The variables $x_1, x_2, \ldots, x_i, \ldots, x_n$ are the n inputs to the threshold element. These are analogous to impulses arriving from several different neurons to one neuron. The variables w_1, $w_2, \ldots$, $w_i, \ldots$, w_n are the weights associated with the impulses/inputs, signifying the relative importance that is associated with the path from which the input is coming. When w_i is positive, input x_i acts as an excitatory signal for the element. When w_i is negative, input x_i acts as an inhibitory signal for the element. The threshold element sums the product of these inputs and their associated weights ($\Sigma w_i x_i$), compares it to a prescribed threshold value, and, if the summation is greater than the threshold value, computes an output using a nonlinear function (F). The signal output y (Figure 6.5) is a nonlinear function (F) of the difference between the preceding computed summation and the threshold value and is expressed as

$$y = F\left(\sum w_i x_i - t\right), \tag{6.7}$$

where

x_i signal input ($i = 1, 2, \ldots, n$)
w_i weight associated with the signal input x_i
t threshold level prescribed by user

$F(s)$ is a nonlinear function, for example, a sigmoid function $F(s) = \dfrac{1}{1 + e^{-s}}$.

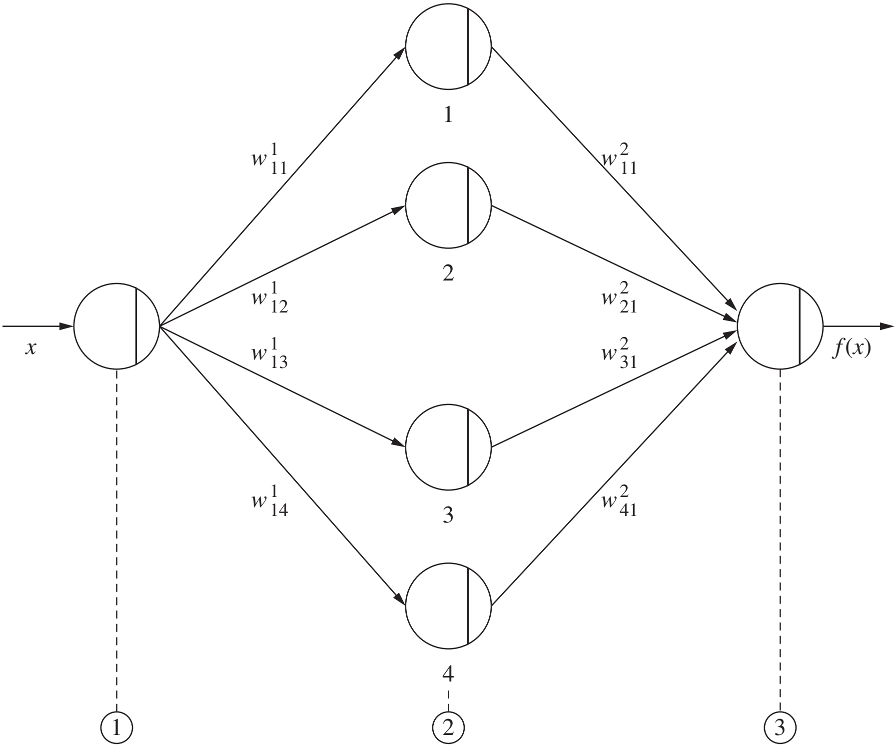

Figure 6.6 A simple $1 \times 4 \times 1$ neural network, where w^j_{jk} represents the weight associated with the path connecting the j th element of the ith layer to the kth element of the $(i + 1)$th layer.

The nonlinear function, F, is a modeling choice and is a function of the type of output signal desired in the neural network model. Popular choices for this function are a sigmoid function, a step function, and a ramp function on the unit interval.

Figure 6.6 shows a simple neural network for a system with single-input signal x and a corresponding single-output signal $f(x)$. The first layer has only one element that has a single input, but the element sends its output to four other elements in the second layer. Elements shown in the second layer are all single-input, single-output elements. The third layer has only one element that has four inputs, and it computes the output for the system. This neural network is termed a $(1 \times 4 \times 1)$ neural network. The numbers represent the number of elements in each layer of the network. The layers other than the first (input layer) and the last (output layer) layers constitute the set of hidden layers. (Systems can have more than three layers, in which case we would have more than one hidden layer.)

Neural systems solve problems by adapting to the nature of the data (signals) they receive. One of the ways to accomplish this is to use a training data set and a checking data set of input and output data/signals (x, y); for a multiple-input, multiple-output system using a neural network, we may use input–output sets comprising vectors $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$. We start with a random assignment of weights w^i_{jk} to the paths joining the elements in the different layers (Figure 6.6). Then an input x from the training data set is passed through the neural network.

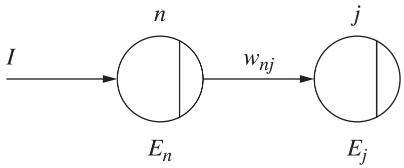

Figure 6.7 Distribution of error to different elements.

The neural network computes a value ($f(x)_{\text{output}}$), which is compared with the actual value ($f(x)_{\text{actual}} = y$). The error measure E is computed from these two output values as

$$E = f(x)_{\text{actual}} - f(x)_{\text{output}}. \tag{6.8}$$

This is the error measure associated with the last layer of the neural network (for Figure 6.6); in this case the error measure E would be associated with the third layer in the neural network. Next we try to distribute this error to the elements in the hidden layers using a technique called *back-propagation*.

The error measure associated with the different elements in the hidden layers is computed as follows. Let E_j be the error associated with the jth element (Figure 6.7). Let w_{nj} be the weight associated with the line from element n to element j and let I be the input to unit n. The error for element n is computed as

$$E_n = F'(I) w_{nj} E_j, \tag{6.9}$$

where, for $F(I) = 1/(1 + e^{-I})$, the sigmoid function, we have

$$F'(I) = F(I)(1 - F(\mathrm{I})). \tag{6.10}$$

Next the different weights w^i_{jk} connecting different elements in the network are corrected so that they can approximate the final output more closely. For updating the weights, the error measure on the elements is used to update the weights on the lines joining the elements.

For an element with an error E associated with it, as shown in Figure 6.8, the associated weights may be updated as

$$w_i(\text{new}) = w_i(\text{old}) + \alpha E x_i \tag{6.11}$$

where

α = learning constant
E = associated error measure
x_i = input to the element

The input value x_i is passed through the neural network (now having the updated weights) again, and the errors, if any, are computed again. This technique is iterated until the error value of the final output is within some user-prescribed limits.

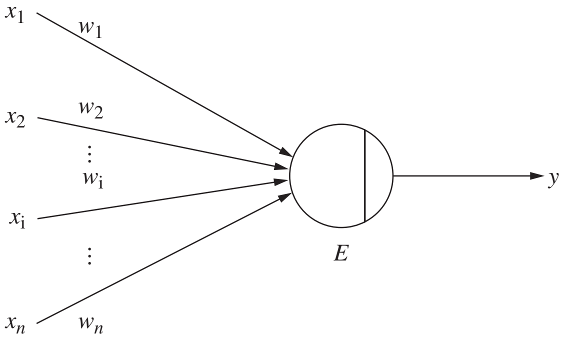

Figure 6.8 A threshold element with an error E associated with it.

The neural network then uses the next set of input–output data. This method is continued for all data in the training data set. This technique makes the neural network simulate the nonlinear relation between the input–output data sets. Finally, a checking data set is used to verify how well the neural network can simulate the nonlinear relationship.

For systems where we may have data sets of inputs and corresponding outputs, and where the relationship between the input and output may be highly nonlinear or not known at all, we may want to use fuzzy logic to classify the input and the output data sets broadly into different fuzzy classes. Furthermore, for systems that are dynamic in nature (the system parameters may change in a nondeterministic fashion) the fuzzy membership functions would have to be repeatedly updated. For these types of systems, it is advantageous to use a neural network because the network can modify itself (by changing the weight assignments in the neural network) to accommodate the changes. Unlike symbolic learning algorithms, for example, conventional expert systems (Luger and Stubblefield, 1989), neural networks do not learn by adding new rules to their knowledge base; they learn by modifying their overall structure. The lack of intuitive knowledge in the learning process is one of the major drawbacks of neural networks for use in cognitive learning.

Generation of Membership Functions Using a Neural Network

We consider here a method by which fuzzy membership functions may be created for fuzzy classes of an input data set (Takagi and Hayashi, 1991). We select a number of input data values and divide them into a training data set and a checking data set. The training data set is used to train the neural network. Let us consider an input training data set as shown in Figure 6.9a. Table 6.2 shows the coordinate values of the different data points considered (e.g., crosses in Figure 6.9a). The data points are expressed with two coordinates each because the data shown in Figure 6.9a represent a two-dimensional problem. The data points are first divided into different classes (Figure 6.9a) by conventional clustering techniques (these are explained in Chapter 10).

As shown in Figure 6.9a the data points have been divided into three regions, or classes, R^1, R^2, and R^3. Let us consider data point 1, which has input coordinate values of $x_1 = 0.7$ and $x_2 = 0.8$ (Figure 6.9d). As this is in region R_2, we assign to it a complete membership

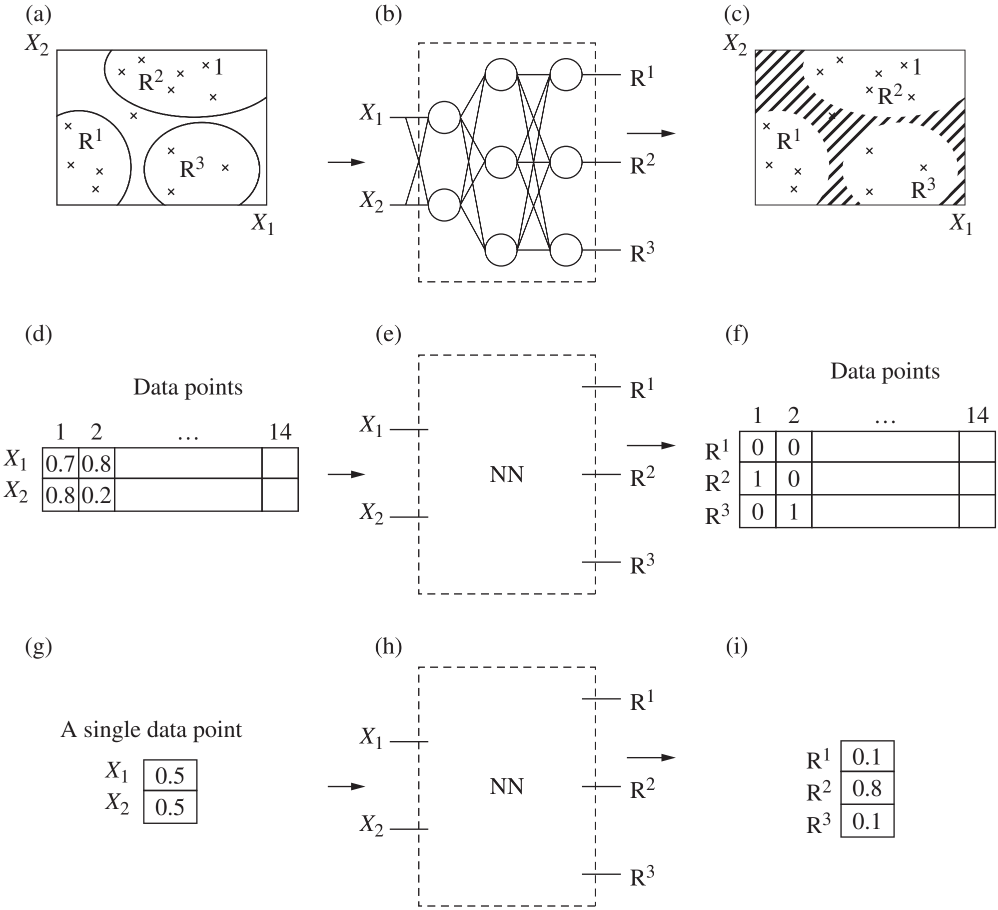

Figure 6.9 Using a neural network to determine membership functions. Adapted from Takagi and Hayashi, 1991.

Table 6.2 Variables describing the data points to be used as a training data set.

Data point	1	2	3	4	5	6	7	8	9	10
x_1	0.05	0.09	0.12	0.15	0.20	0.75	0.80	0.82	0.90	0.95
x_2	0.02	0.11	0.20	0.22	0.25	0.75	0.83	0.80	0.89	0.89

of one in class R_2 and zero membership in classes R_1 and R_3 (Figure 6.9f). Similarly, the other data points are assigned membership values of unity for the classes they belong to initially. A neural network is created (Figures 6.9b, e, and h) that uses the data point marked 1 and the corresponding membership values in different classes for training itself to simulate the relationship between coordinate locations and the membership values. Figure 6.9c represents the output of the neural network, which classifies data points into one of the three regions. The neural network then uses the next set of data values (e.g., point 2) and membership values

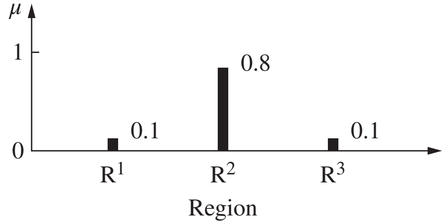

Figure 6.10 Membership function for data point (X_1, $X_2 = (0.5, 0.5)$).

Table 6.3 Variables describing the data points to be used as a checking data set.

Data point	11	12	13	14	15	16	17	18	19	20
x_1	0.09	0.10	0.14	0.18	0.22	0.77	0.79	0.84	0.94	0.98
x_2	0.04	0.10	0.21	0.24	0.28	0.78	0.81	0.82	0.93	0.99

to train itself further as seen in Figure 6.9d. This repetitive process is continued until the neural network can simulate the entire set of input–output (coordinate location–membership value) values. The performance of the neural network is then checked using the checking data set. Once the neural network is ready, its final version (Figure 6.9h) can be used to determine the membership values (function) of any input data (Figure 6.9g) in the different regions (Figure 6.9i).

Notice that the points shown in the table in Figure 6.9i are actually the membership values in each region for the data point shown in Figure 6.9g. These could be plotted as a membership function, as shown in Figure 6.10. A complete mapping of the membership of different data points in the different fuzzy classes can be derived to determine the overlap of the different classes (the hatched portion in Figure 6.9c shows the overlap of the three fuzzy classes). These steps will become clearer as we go through the computations in the following example.

Example 6.3

Let us consider a system that has 20 data points described in two-dimensional format (two variables) as shown in Tables 6.2 and 6.3. We have placed these data points in two fuzzy classes, R_1 and R_2, using a clustering technique (see Chapter 10). We would like to form a neural network that can determine the membership values of any data point in the two classes. We would use the data points in Table 6.2 to train the neural network and the data points in Table 6.3 to check its performance. The membership values in Table 6.4 are to be used to train and check the performance of the neural network. The data points that are to be used for training and checking the performance of the neural network have been assigned membership values of unity for the classes into which they have been originally assigned, as seen in Table 6.4.

We select a $2 \times 3 \times 3 \times 2$ neural network to simulate the relationship between the data points and their membership in the two fuzzy sets, R_1 and R_2 (Figure 6.11). The coordinates x_1 and x_2 for each data point are used as the input values, and the corresponding membership values in the two fuzzy classes for each data point are the output values for the neural network.

Table 6.5 shows the initial quasi-random values that have been assigned to the different weights connecting the paths between the elements in the layers in the network shown in

Table 6.4 Membership values (output) of the data points in the training and checking data sets to be used for training and checking the performance of the neural network.

Data points	1 & 11	2 & 12	3 & 13	4 & 14	5 & 15	6 & 16
R_1	1.0	1.0	1.0	1.0	1.0	0.0
R_2	0.0	0.0	0.0	0.0	0.0	1.0
	7 & 17	8 & 18	9 & 19	10 & 20		
	0.0	0.0	0.0	0.0		
	1.0	1.0	1.0	1.0		

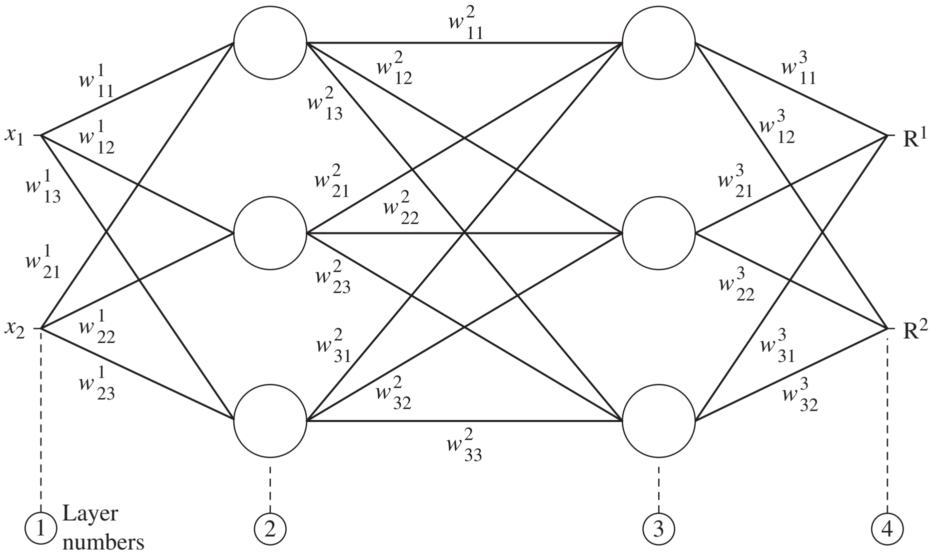

Figure 6.11 The $[2 \times 3 \times 3 \times 2]$ neural network to be trained for the data set of Example 6.3.

Table 6.5 The initial quasi-random values that have been assigned to the different weights connecting the paths between the elements in the layers in the network of Figure 6.11.

$w^1_{11} = 0.5$	$w^2_{11} = 0.10$	$w^3_{11} = 0.30$
$w^1_{12} = 0.4$	$w^2_{12} = 0.55$	$w^3_{12} = 0.35$
$w^1_{13} = 0.1$	$w^2_{13} = 0.35$	$w^3_{21} = 0.35$
$w^1_{21} = 0.2$	$w^2_{21} = 0.20$	$w^3_{22} = 0.25$
$w^1_{22} = 0.6$	$w^2_{22} = 0.45$	$w^3_{31} = 0.45$
$w^1_{23} = 0.2$	$w^2_{23} = 0.35$	$w^3_{32} = 0.30$
	$w^2_{31} = 0.25$	
	$w^2_{32} = 0.15$	
	$w^2_{33} = 0.60$	

Figure 6.11. We take the first data point ($x_1 = 0.05$, $x_2 = 0.02$) as the input to the neural network. We will use Equation (6.7) in the form

$$O = \frac{1}{1 + \exp\left[-\left(\sum x_i w_i - t\right)\right]} \tag{6.12}$$

where

O = output of the threshold element computed using the sigmoidal function
x_i = inputs to the threshold element ($i = 1, 2, \ldots, n$)
w_i = weights attached to the inputs
t = threshold for the element.

First iteration: We start off with the first iteration in training the neural network using Equation (6.12) to determine the outputs of the different elements by calculating the outputs for each of the neural network layers. We select a threshold value of $t = 0$.

Outputs for the second layer:

$$O_1^2 = \frac{1}{1 + \exp\{-[(0.05 \times 0.50) + (0.02 \times 0.20) - 0.0]\}} = 0.507249,$$

$$O_2^2 = \frac{1}{1 + \exp\{-[(0.05 \times 0.40) + (0.02 \times 0.60) - 0.0]\}} = 0.507999,$$

$$O_3^2 = \frac{1}{1 + \exp\{-[(0.05 \times 0.10) + (0.02 \times 0.20) - 0.0]\}} = 0.502250.$$

Outputs for the third layer:

$$O_1^2 = \frac{1}{1 + \exp\{-[(0.507249 \times 0.10) + (0.507999 \times 0.20) + (0.502250 \times 0.20) - 0.0]\}}$$

$$= 0.569028,$$

$$O_2^3 = \frac{1}{1 + \exp\{-[(0.507249 \times 0.55) + (0.507999 \times 0.45) + (0.502250 \times 0.15) - 0.0]\}}$$

$$= 0.641740,$$

$$O_3^3 = \frac{1}{1 + \exp\{-[(0.507249 \times 0.35) + (0.507999 \times 0.35) + (0.502250 \times 0.60) - 0.0]\}}$$

$$= 0.658516.$$

Outputs for the fourth layer:

$$O_1^4 = \frac{1}{1+\exp\{-[(0.509028\times 0.30)+(0.641740\times 0.35)+(0.658516\times 0.45)-0.0]\}}$$
$$=0.666334,$$
$$O_2^4 = \frac{1}{1+\exp\{-[(0.569028\times 0.35)+(0.641740\times 0.25)+(0.658516\times 0.30)-0.0]\}}$$
$$=0.635793.$$

Determining errors:

$$\mathrm{R}_1 : E_1^4 = O_{1\,\mathrm{actual}}^4 - O_1^4 = 1.0-0.666334 = 0.333666,$$
$$\mathrm{R}_2 : E_2^4 = O_{2\,\mathrm{actual}}^4 - O_2^4 = 0.0-0.635793 = -0.635793.$$

Now that we know the final errors for the neural network for the first iteration, we distribute this error to the other nodes (elements) in the network using Equations (6.9) and (6.10) in the form

$$E_n = O_n(1-O_n)\sum_j w_{nj}E_j. \tag{6.13}$$

Assigning errors: First, we assign errors to the elements in the third layer,

$$E_1^3 = 0.569028(1.0-0.569028)[(0.30\times 0.333666)+(0.35\times(-0.635793))] = -0.030024,$$
$$E_2^3 = 0.641740(1.0-0.641740)[(0.35\times 0.333666)+(0.25\times(-0.635793))] = -0.009694,$$
$$E_3^3 = 0.658516(1.0-0.658516)[(0.45\times 0.333666)+(0.30\times(-0.635793))] = -0.009127,$$

and then assign errors to the elements in the second layer,

$$E_1^2 = 0.507249(1.0-0.507249)[(0.10\times(-0.030024))+(0.55\times(-0.009694))$$
$$+(0.35\times(-0.009127))] = -0.002882,$$
$$E_2^2 = 0.507999(1.0-0.507999)[(0.20\times(-0.030024))+(0.45\times(-0.009694))$$
$$+(0.35\times(-0.009127))] = -0.003390,$$
$$E_3^2 = 0.502250(1.0-0.502250)[(0.25\times(-0.030024))+(0.15\times(-0.009694))$$
$$+(0.60\times(-0.009127))] = -0.003609.$$

Now that we know the errors associated with each element in the network we can update the weights associated with these elements so that the network approximates the output more closely. To update the weights, we use Equation (6.11) in the form

$$w_{jk}^j(\mathrm{new}) = w_{jk}^j(\mathrm{old}) + \alpha E_k^{i+1} x_{jk}, \tag{6.14}$$

where

w_{jk}^{j} = represents the weight associated with the path connecting the jth element of the ith layer to the kth element of the $(i+1)$th layer
α = learning constant, which we will take as 0.3 for this example
E_k^{i+1} = error associated with the kth element of the $(i+1)$th layer
x_{jk} = input from the jth element in the ith layer to the kth element in the $(i+1)$th layer $\left(O_j^i\right)$.

Updating weights: We will update the weights connecting elements in the third and the fourth layers,

$$w_{11}^3 = 0.30 + 0.3 \times 0.333666 \times 0.569028 = 0.356960,$$
$$w_{21}^3 = 0.35 + 0.3 \times 0.333666 \times 0.641740 = 0.414238,$$
$$w_{31}^3 = 0.45 + 0.3 \times 0.333666 \times 0.658516 = 0.515917,$$
$$w_{12}^3 = 0.35 + 0.3 \times (-0.635793) \times 0.569028 = 0.241465,$$
$$w_{22}^3 = 0.25 + 0.3 \times (-0.635793) \times 0.641740 = 0.127596,$$
$$w_{22}^3 = 0.30 + 0.3 \times (-0.635793) \times 0.658516 = 0.174396,$$

then update weights connecting elements in the second and the third layers,

$$w_{11}^2 = 0.10 + 0.3 \times (-0.030024) \times 0.507249 = 0.095431,$$
$$w_{21}^3 = 0.20 + 0.3 \times (-0.030024) \times 0.507999 = 0.195424,$$
$$w_{31}^2 = 0.25 + 0.3 \times (-0.030024) \times 0.502250 = 0.245476,$$
$$w_{12}^2 = 0.55 + 0.3 \times (-0.009694) \times 0.507249 = 0.548525,$$
$$w_{22}^2 = 0.45 + 0.3 \times (-0.009694) \times 0.507999 = 0.448523,$$
$$w_{32}^3 = 0.15 + 0.3 \times (-0.009694) \times 0.502250 = 0.148540,$$
$$w_{13}^2 = 0.35 + 0.3 \times (-0.009127) \times 0.507249 = 0.348611,$$
$$w_{23}^2 = 0.35 + 0.3 \times (-0.009127) \times 0.507999 = 0.348609,$$
$$w_{33}^2 = 0.60 + 0.3 \times (-0.009127) \times 0.502250 = 0.598625,$$

and then, finally, update weights connecting elements in the first and the second layers,

$$w_{11}^1 = 0.50 + 0.3 \times (-0.002882) \times 0.05 = 0.499957,$$
$$w_{12}^1 = 0.40 + 0.3 \times (-0.003390) \times 0.05 = 0.399949,$$
$$w_{13}^1 = 0.10 + 0.3 \times (-0.003609) \times 0.05 = 0.099946,$$
$$w_{21}^1 = 0.20 + 0.3 \times (-0.003609) \times 0.02 = 0.199983,$$
$$w_{22}^1 = 0.60 + 0.3 \times (-0.003390) \times 0.02 = 0.5899980,$$
$$w_{23}^1 = 0.20 + 0.3 \times (-0.003609) \times 0.02 = 0.199978.$$

Now that all the weights in the neural network have been updated, the input data point ($x_1 = 0.05$, $x_2 = 0.02$) is again passed through the neural network. The errors in approximating the output are computed again and redistributed as before. This process is continued until the errors are within acceptable limits. Next, the second data point ($x_1 = 0.09$, $x_2 = 0.11$, Table 6.2) and the corresponding membership values ($R^1 = 1$, $R^2 = 0$, Table 6.4) are used to train the network. This process is continued until all the data points in the *training* data set (Table 6.2) are used. The performance of the neural network (how closely it can predict the value of the membership of the data point) is then checked using the data points in the *checking* data set (Table 6.3).

Once the neural network is trained and verified to be performing satisfactorily, it can be used to find the membership of any other data points in the two fuzzy classes. A complete mapping of the membership of different data points in the different fuzzy classes can be derived to determine the overlap of the different classes (R_1 and R_2).

Genetic Algorithms

As in the previous section, we will first provide a brief introduction to genetic algorithms and then show how these can be used to determine membership functions. In the previous section we introduced the concept of a neural network. In implementing a neural network algorithm, we try to recreate the working of neurons in the human brain. In this section we introduce another class of algorithms, which use the concept of Darwin's theory of evolution. Darwin's theory basically stressed the fact that the existence of all living things is based on the rule of "survival of the fittest." Darwin also postulated that new breeds or classes of living things come into existence through the processes of reproduction, crossover, and mutation among existing organisms (Forrest, 1993).

These concepts in the theory of evolution have been translated into algorithms to search for solutions to problems in a more "natural" way. First, different possible solutions to a problem are created. These solutions are then tested for their performance (i.e., how good a solution they provide). Among all possible solutions, a fraction of the good solutions is selected, and the others are eliminated (survival of the fittest). The selected solutions undergo the processes of reproduction, crossover, and mutation to create a new generation of possible solutions (which are expected to perform better than the previous generation). This process of production of a new generation and its evaluation is repeated until there is convergence within a generation. The benefit of this technique is that it searches for a solution from a broad spectrum of possible solutions, rather than restrict the search to a narrow domain where the results would be normally expected. Genetic algorithms try to perform an intelligent search for a solution from a nearly infinite number of possible solutions.

We show how the concepts of genetics are translated into a search algorithm (Goldberg, 1989). In a genetic algorithm, the parameter set of the problem is coded as a finite string of bits. For example, given a set of two-dimensional data ((x, y) data points), we want to fit a linear curve (straight line) through the data. To get a linear fit, we encode the parameter set for a line ($y = C_1x + C_2$) by creating independent bit strings for the two unknown constants C_1 and C_2 (parameter set describing the line) and then join them (concatenate the strings). The bit strings are combinations of zeros and ones, which represent the value of a number in binary form.

An n-bit string can accommodate all integers up to the value $2^n - 1$. For example, the number 7 requires a 3-bit string, that is, $2^3 - 1 = 7$, and the bit string would look like "111," where the first unit digit is in the 2^2 place (=4), the second unit digit is in the 2^1 place (=2), and the last unit digit is in the 2^0 place (=1); hence, $4 + 2 + 1 = 7$. The number 10 would look like "1010," that is, $2^3 + 2^1 = 10$, from a 4-bit string. This bit string may be mapped to the value of a parameter, say C_i, $i = 1, 2$, by the mapping

$$C_i = C_{\min} + \frac{b}{2^L - 1}(C_{\max_i} - C_{\min_i}), \tag{6.15}$$

where b is the number in decimal form that is being represented in binary form (e.g., 152 may be represented in binary form as 10011000), L is the length of the bit string (i.e., the number of bits in each string), and $C_{\max}$ and $C_{\min}$ are user-defined constants between which C_1 and C_2 vary linearly. The parameters C_1 and C_2 depend on the problem. The length of the bit strings is based on the handling capacity of the computer being used, that is, on how long a string (strings of each parameter are concatenated to make one long string representing the whole parameter set) the computer can manipulate at an optimum speed.

All genetic algorithms contain three basic operators: reproduction, crossover, and mutation, where all three are analogous to their namesakes in genetics. Let us consider the overall process of a genetic algorithm before trying to understand the basic processes.

First, an initial population of n strings (for n parameters) of length L is created. The strings are created in a random fashion, that is, the values of the parameters that are coded in the strings are random values (created by randomly placing the zeros and ones in the strings). Each of the strings is decoded into a set of parameters that it represents. This set of parameters is passed through a numerical model of the problem space. The numerical model gives out a solution based on the input set of parameters. On the basis of the quality of this solution, the string is assigned a fitness value. The fitness values are determined for each string in the entire population of strings. With these fitness values, the three genetic operators are used to create a new generation of strings, which is expected to perform better than the previous generations (better fitness values). The new set of strings is again decoded and evaluated, and a new generation is created using the three basic genetic operators. This process is continued until convergence is achieved within a population.

Among the three genetic operators, reproduction is the process by which strings with better fitness values receive correspondingly *better copies* in the new generation, that is, we try to ensure that better solutions persist and contribute to better offspring (new strings) during successive generations. This is a way of ensuring the "survival of the fittest" strings. Because the total number of strings in each generation is kept a constant (for computational economy and efficiency), strings with lower fitness values are eliminated.

The second operator, crossover, is the process in which the strings are able to mix and match their desirable qualities in a random fashion. After reproduction, crossover proceeds in three simple steps. First, two new strings are selected at random (Figure 6.12a). Second, a random location in both strings is selected (Figure 6.12b). Third, the portions of the strings to the right of the randomly selected location in the two strings are exchanged (Figure 6.12c). In this way information is exchanged between strings, and portions of high-quality solutions are exchanged and combined.

Reproduction and crossover together give genetic algorithms most of their searching power. The third genetic operator, mutation, helps to increase the searching power. To understand the

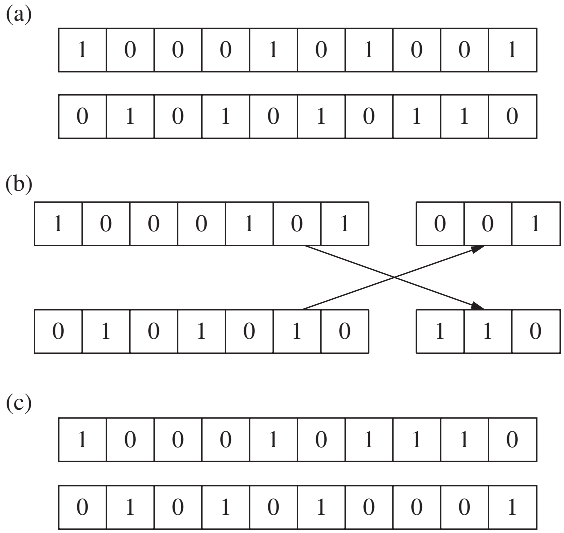

Figure 6.12 Crossover in strings. (a) Two strings are selected at random to be mated; (b) a random location in the strings is located (here the location is before the last three bit locations); and (c) the string portions following the selected location are exchanged.

need for mutation, let us consider the case where reproduction or crossover may not be able to find an optimum solution to a problem. During the creation of a generation it is possible that the entire population of strings is missing a vital bit of information (e.g., none of the strings has a one at the fourth location) that is important for determining the correct or the most nearly optimum solution. Future generations that would be created using reproduction and crossover would not be able to alleviate this problem. Here mutation becomes important. Occasionally, the value at a certain string location is changed, that is, if there is a one originally at a location in the bit string, it is changed to a zero, or vice versa. Mutation thus ensures that the vital bit of information is introduced into the generation. Mutation, as it does in nature, takes place rarely, on the order of once in a thousand bit string locations (a suggested mutation rate is 0.005/bit/generation (Forrest, 1993)).

Let us now consider an example that shows how a line may be fit through a given data set using a genetic algorithm.

Example 6.4

Let us consider the data set in Table 6.6. For performing a line ($y = C_1x + C_2$) fit, as mentioned previously, we first encode the parameter set (C_1, C_2) in the form of bit strings. Bit strings are created with random assignment of ones and zeros at different bit locations. We start with an initial population of four strings (Table 6.7a, column 2). The strings are 12 bits in length. The first 6 bits encode the parameter C_1, and the next 6 bits encode the parameter C_2. Table 6.7a, columns 3 and 5, shows the decimal equivalent of their binary coding. These binary values for C_1 and C_2 are then mapped into values relevant to the problem using Equation (6.15). We assume that the minimum value to which we would expect C_1 or C_2 to go would be −2 and the maximum would be 5 (these are arbitrary values; any other values could just as easily have

Table 6.6 Data set through which a line fit is required.

Data number	x	y'
1	1.0	1.0
2	2.0	2.0
3	4.0	4.0
4	6.0	6.0

been chosen). Therefore, for Equation (6.15), $C_{\min_i} = -2$ and $C_{\max_i} = 5$. Using these values, we compute C_1 and C_2 (Table 6.7a, columns 4 and 6). The values shown in Table 6.7a, columns 7, 8, 9, and 10, are the values computed using the equation $y = C_1x + C_2$, using the values of C_1 and C_2 from columns 4 and 6, respectively, for different values of x as given in Table 6.6. These computed values for the y are compared with the correct values (Table 6.6), and the square of the errors in estimating the y is calculated for each string. This summation is subtracted from a large number (400 in this problem) (Table 6.7a, column 11) to convert the problem into a maximization problem. The values in Table 6.7a, column 11, are the fitness values for the four strings. These fitness values are added. Their average is also computed. The fitness value of each string is divided by the average fitness value of the whole population of strings to give an estimate of the relative fitness of each string (Table 6.7a, column 12). This measure also acts as a guide as to which strings are eliminated from consideration for the next generation and which string "gets reproduced" in the next generation. In this problem a cutoff value of 0.80 (relative fitness) has been used for the acceptability of a string succeeding into the next generation. Table 6.7a, column 13, shows the number of copies of each of the four strings that would be used to create the next generation of strings.

Table 6.7b is a continuation of Table 6.7a. The first column in Table 6.7b shows the four strings selected from the previous generation aligned for crossover at the locations shown in the strings in the column. After crossover, the new strings generated are shown in Table 6.7b, column 2. These strings undergo the same process of decoding and evaluation as the previous generation. This process is shown in Table 6.7b, columns 3–13. We notice that the average fitness of the second generation is greater than that of the first generation of strings.

The process of generation of strings and their evaluation is continued until we get a convergence to the solution (i.e., final values for C_1 and C_2) within a generation.

Computing Membership Functions Using Genetic Algorithms

Genetic algorithms as just described can be used to compute membership functions (Karr and Gentry, 1993). Given some functional mapping for a system, some membership functions and their shapes are assumed for the various fuzzy variables defined for a problem. These membership functions are then coded as bit strings that are then concatenated. An evaluation (fitness) function is used to evaluate the fitness of each set of membership functions (parameters that define the functional mapping). This procedure is illustrated for a simple problem in the next example.

Table 6.7a First iteration using a genetic algorithm, Example 6.4.

(1) String number	(2) String	(3) C_1	(4) C_1 (binary)	(5) C_2	(6) C_2 (binary)	(7) y_1	(8) y_2	(9) y_3	(10) y_4	(11) $f(x)=400-\sum(y_i-y_i')^2$	(12) Expected count $= f/f_{av}$	(13) Actual count
1	000111 010100	7	−1.22	20	0.22	−1.00	−2.22	−4.66	−7.11	147.49	0.48	0
2	010010 001100	18	0.00	12	−0.67	−0.67	−0.67	−0.67	−0.67	332.22	1.08	1
3	010101 101010	21	0.33	42	2.67	3.00	3.33	5.00	4.67	391.44	1.27	2
4	100100 001001	36	2.00	9	−1.00	1.00	3.00	3.67	11.00	358.00	1.17	1
									Sum	1229.15		
									Average	307.29		
									Maximum	391.44		

Table 6.7b Second iteration using a genetic algorithm. Example 6.4.

(1) Selected strings	(2) New strings	(3) C_1	(4) C_1 (binary)	(5) C_2	(6) C_2 (binary)	(7) y_1	(8) y_2	(9) y_3	(10) y_4	(11) $f(x)=400-\sum(y_i-y_i')^2$	(12) Expected count $= f/f_{av}$	(13) Actual count
0101|01 101010	010110 001100	22	0.44	12	−0.67	−0.22	0.22	1.11	2.00	375.78	1.15	1
0100|10 001100	010001 101010	17	−0.11	42	2.67	2.56	2.44	2.22	2.00	380.78	1.17	2
010101 101|010	010101 101001	21	0.33	41	2.56	2.89	3.22	3.89	4.56	292.06	0.90	1
100100 001|001	100100 001010	36	2.0	10	−0.89	1.11	3.11	7.11	11.11	255.73	0.78	0
									Sum	1304.35		
									Average	326.09		
									Maximum	380.78		

Example 6.5

Let us consider that we have a single-input (x), single-output (y) system with input–output values as shown in Table 6.8. Table 6.9 shows a functional mapping for this system between the input (x) and the output (y).

In Table 6.9 we see that each of the variables x and y makes use of two fuzzy classes (x uses S (small) and L (large); y uses S (small) and VL (very large)). The functional mapping tells us that a *small x* maps to a *small y*, and a *large x* maps to a *very large y*. We assume that the range of the variable x is [0, 5] and that of y is [0, 25]. We assume that each membership function has the shape of a right triangle, as shown in Figure 6.13.

The membership function on the right side of Figure 6.13 is constrained to have the right-angle wedge at the upper limit of the range of the fuzzy variable. The membership function on the left side is constrained to have the right-angle wedge on the lower limit of the range of the fuzzy variable. It is intuitively obvious that under the foregoing constraints the only thing needed to describe the shape and position of the membership function fully is the length of the base of the right-triangle membership functions. We use this fact in encoding the membership functions as bit strings.

The unknown variables in this problem are the lengths of the bases of the four membership functions (x(S, L) and y(S, VL)). We use 6-bit binary strings to define the base of each of the membership functions. (The binary values are later mapped to decimal values using Equation (6.15).) These strings are then concatenated to give us a 24-bit (6×4) string. As shown in Table 6.10a, column 1, we start with an initial population of four strings. These are decoded to the binary values of the variables as shown in Table 6.10a, columns 2, 3, 4,

Table 6.8 Data for a single-input, single-output system.

x	1	2	3	4	5
y	1	4	9	16	25

Table 6.9 Functional mapping for the system.

x	S	L
y	S	VL

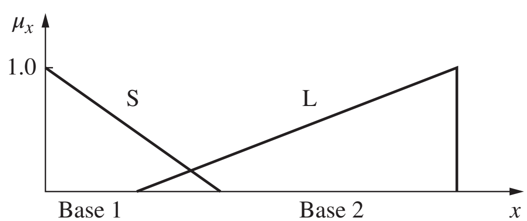

Figure 6.13 Membership functions for the input (and output) variables are assumed to be right triangles.

Table 6.10a First iteration using a genetic algorithm for determining optimal membership functions.

String number	(1) String	(2) Base 1 (binary)	(3) Base 2 (binary)	(4) Base 3 (binary)	(5) Base 4 (binary)	(6) Base 1	(7) Base 2	(8) Base 3	(9) Base 4	(10) y' $(x=1)$	(11) y' $(x=2)$	(12) y' $(x=3)$	(13) y' $(x=4)$	(14) y' $(x=5)$	(15) $1000-\sum(y_i-y_i')^2$	(16) Expected count $= f/f_{av}$
1	000111 010100 010110 110011	7	20	22	51	0.56	1.59	8.73	20.24	0	0	0	12.25	25	887.94	1.24
2	010010 001100 101100 100110	18	12	44	38	1.43	0.95	17.46	15.08	12.22	0	0	0	25	521.11	0.73
3	010101 101010 001101 101000	21	42	13	40	1.67	3.33	5.16	15.87	3.1	10.72	15.48	20.24	25	890.46	1.25
4	100100 001001 101100 100011	36	9	44	35	2.86	0.71	17.46	13.89	6.98	12.22	0	0	25	559.67	0.78
														Sum	2859.18	
														Average	714.80	
														Maximum	890.46	

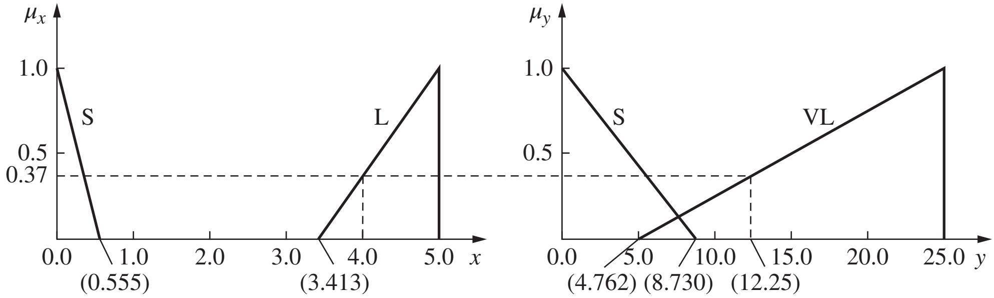

Figure 6.14 Physical representation of the first string in Table 6.12 and the graphical determination of y for a given x.

and 5. The binary values are mapped to decimal values for the fuzzy variables using Equation (6.15) (Table 6.10a, columns 6, 7, 8, and 9). For the fuzzy variable x (range $x = 0, 5$) we use $C_{\min} = 0$ and $C_{\max} = 5$ for both the membership functions S (Small) and L (Large). For the fuzzy variable y (range $y = 0, 25$) we use $C_{\min} = 0$ and $C_{\max} = 25$.

The physical representation of the first string is shown in Figure 6.14. In this figure the base values are obtained from Table 6.10a, columns 6, 7, 8, and 9. So, for example, the base values for the x variable for string number 1 are 0.56 and $5 - 1.59 = 3.41$, and the base values for the y variable are 8.73 and $25 - 20.24 = 4.76$. To determine the fitness of the combination of membership functions in each of the strings, we want a measure of the square of the errors that are produced in estimating the value of the outputs y, given the inputs x from Table 6.8. Figure 6.14 shows how the value of the output y can be computed graphically from the membership functions for string number 1 in Table 6.10a. For example, for $x = 4$ we see that the membership of x in the fuzzy class Large is 0.37. Referring to the rules in Table 6.9, we see that if x is Large then y is Very Large. Therefore, we look for the value in the fuzzy class Very Large (VL) of fuzzy variable y that has a membership of 0.37. We determine this to be equal to 12.25. The corresponding actual value for y is 16 (Table 6.8). Therefore, the squared error is $(16 - 12.25)^2 = 14.06$. Columns 10, 11, 12, 13, and 14 of Table 6.10a show the values computed for y using the respective membership functions. Table 6.10a, column 15, shows the sum of the squared errors subtracted from 1000 (this is done to convert the fitness function from a minimization problem to a maximization problem). Table 6.10a, column 15, thus shows the fitness values for the four strings. We find the sum of all the fitness values in the generation and the average fitness of the generation. The average fitness of the generation is used to determine the relative fitness of the strings in the generation, as seen in Table 6.10a, column 16. These relative fitness values are used to determine which strings are to be eliminated and which string gets how many copies to make the next generation of strings. In this problem a cutoff value of 0.75 (relative fitness) has been used for the acceptability of a string propagating into the next generation. Table 6.10a, column 17, shows the number of copies of each of the four strings that would be used to create the next generation of strings.

Table 6.10b is a continuation of Table 6.10a. The first column in Table 6.10b shows the four strings selected from the previous generation aligned for crossover at the locations shown in the strings in the column. After crossover, the new strings generated are shown in Table 6.10b, column 2. These strings undergo the same process of decoding and evaluation as the previous able 6.10b Second iteration using a genetic algorithm for determining optimal membership functions.

) elected rings	(2) New Strings	(3) Base 1 (binary)	(4) Base 2 (binary)	(5) Base 3 (binary)	(6) Base 4 (binary)	(7) Base 1	(8) Base 2	(9) Base 3	(10) Base 4	(11) y' (x = 1))	(12) y' (x = 2)	(13) y' (x = 3)	(14) y' (x = 4)	(15) y' (x = 5)	(16) 1000 − $\sum(y_i - y_i')^2$	(17) Expected count = f/f_{av}
00111	000111	7	22	13	40	0.56	1.75	5.16	15.87	0	0	0	15.93	25	902.00	1.10
101|00	010110															
10110	001101															
10011	101000															
10101	010101	21	40	22	51	1.67	3.17	8.73	20.24	5.24	5.85	12.23	18.62	25	961.30	1.18
010|10	101000															
01101	010110															
01000	110011															
10101	010101	21	42	13	35	1.67	3.33	5.16	13.89	3.1	12.51	16.68	20.84	25	840.78	1.03
01010	101010															
01101	001101															
01000	10|0011															
00100	100100	36	9	44	40	2.86	0.71	17.46	15.87	6.11	12.22	0	0	25	569.32	0.70
01001	001001															
01100	101100															
00011	10|1000															
														Sum	3 273.40	
														Average	818.35	
														Maximum	961.30	

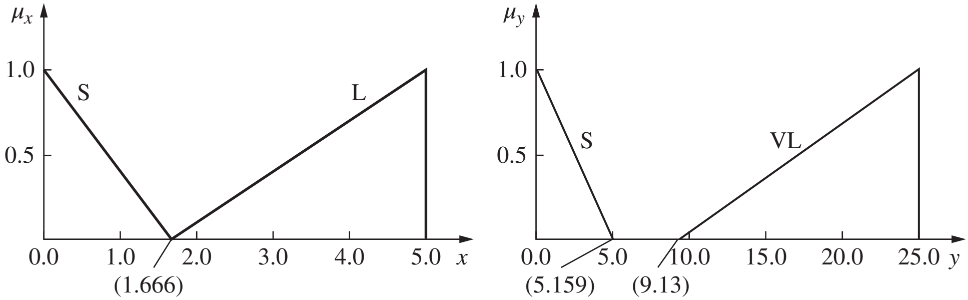

Figure 6.15 Physical mapping of the best string in the first generation of strings in the genetic algorithm.

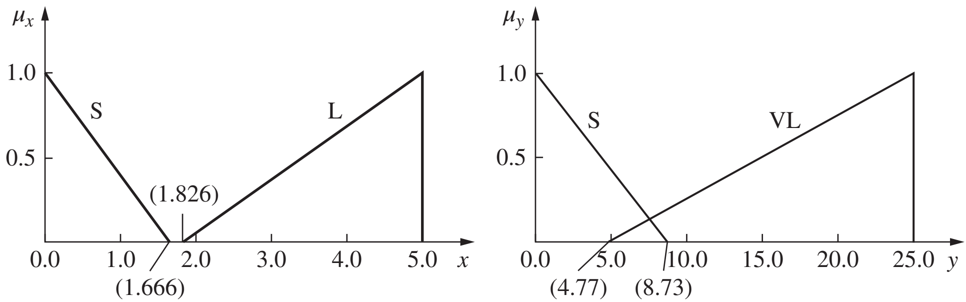

Figure 6.16 Physical mapping of the best string in the second generation of strings in the genetic algorithm.

generation. This process is shown in Table 6.10b, columns 3–18. Also, the fitness of the best string in the second generation is greater than the fitness of the best string in the first generation. Figure 6.15 shows the physical mapping of the best string in the first generation. Figure 6.16 shows the physical mapping of the best string in the second generation; notice that the membership values for the *y* variable in Figure 6.16 show overlap, which is a desirable property of membership functions.

The process of generating and evaluating strings is continued until we get a convergence to the solution within a generation, that is, we get the membership functions with the best fitness value.

Inductive Reasoning

An automatic generation of membership functions can also be accommodated by using the essential characteristic of *inductive reasoning*, which derives a general consensus from the particular (derives the generic from the specific). The induction is performed by the entropy minimization principle, which clusters most optimally the parameters corresponding to the output classes (De Luca and Termini, 1972).

This method is based on an ideal scheme that describes the input and output relationships for a well-established database, that is, the method generates membership functions based solely on the data provided. The method can be quite useful for complex systems where the data are abundant and static. In situations where the data are dynamic, the method may not be useful because the membership functions will continually change with time (see the chapter summary for a discussion on the merits of this method).

The intent of induction is to discover a law having objective validity and universal application. Beginning with the particular, induction concludes with the general. The essential principles of induction have been known for centuries. Three laws of induction are summarized here (Christensen, 1980):

1. Given a set of irreducible outcomes of an experiment, the induced probabilities are those probabilities consistent with all available information that maximize the entropy of the set.
2. The induced probability of a set of independent observations is proportional to the probability density of the induced probability of a single observation.
3. The induced rule is that rule consistent with all available information of which the entropy is minimum.

Among the three laws, the third one is appropriate for classification (or, for our purposes, membership function development) and the second one for calculating the mean probability of each step of separation (or partitioning). In classification, the probability aspects of the problem are completely disregarded because the issue is simply a binary one: a data point is either in a class or not.

A key goal of entropy minimization analysis is to determine the quantity of information in a given data set. The entropy of a probability distribution is a measure of the uncertainty of the distribution (Yager and Filev, 1994). This information measure compares the contents of data to a prior probability for the same data. The higher the prior estimate of the probability for an outcome to occur, the lower will be the information gained by observing it to occur. The entropy on a set of possible outcomes of a trial where one and only one outcome is true is defined by the summation of probability and the logarithm of the probability for all outcomes. In other words, the entropy is the expected value of information.

For a simple one-dimensional (one uncertain variable) case, let us assume that the probability of the ith sample w_i to be true is $\{p(w_i)\}$. If we actually observe the sample w_i in the future and discover that it is true, then we gain the following information, $I(w_i)$:

$$I(w_i) = -k \ln p(w_i), \tag{6.16}$$

where k is a normalizing parameter. If we discover that it is false, we still gain this information:

$$I(\overline{w}_i) = -k \ln [1 - p(w_i)]. \tag{6.17}$$

Then the entropy of the inner product of all the samples (N) is

$$S = -k \sum_{i=1}^{N} [p_i \ln p_i + (1 - p_i) \ln (1 - p_i)], \tag{6.18}$$

where $p_i = p(w_i)$. The minus sign before parameter k in Equation (6.18) ensures that $S \geq 0$, because $\ln x \leq 0$ for $0 \leq x \leq 1$.

The third law of induction, which is typical in pattern classification, says that the entropy of a rule should be minimized. Minimum entropy (S) is associated with all the p_i being as close to ones or zeros as possible, which in turn implies that they have a high probability of either happening or not happening, respectively. Note in Equation (6.18) that if $p_i = 1$ then $S = 0$. This result makes sense since p_i is the probability measure of whether a value belongs to a partition or not.

Membership Function Generation

To subdivide our data set into membership functions we need some procedure to establish fuzzy thresholds between classes of data. We can determine a threshold line with an entropy minimization screening method, then start the segmentation process, first into two classes. By partitioning first into two classes one more time, we can have three different classes. Therefore, a repeated partitioning with threshold value calculations will allow us to partition the data set into a number of classes, or fuzzy sets, depending on the shape used to describe membership in each set.

Membership function generation is based on a partitioning or analog screening concept, which draws a threshold line between two classes of sample data. The main idea behind drawing the threshold line is to classify the samples while minimizing the entropy for an optimum partitioning. The following is a brief review of the threshold value calculation using the induction principle for a two-class problem. First, we assume that we are seeking a threshold value for a sample in the range between x_1 and x_2. Considering this sample alone, we write an entropy equation for the regions $[x_1, x]$ and $[x, x_2]$. We denote the first region p and the second region q, as is shown in Figure 6.17. By moving an imaginary threshold value x between x_1 and x_2, we calculate entropy for each value of x.

An entropy with each value of x in the region x_1 and x_2 is expressed by Christensen (1980) as

$$S(x) = p(x)S_p(x) + q(x)S_q(x), \tag{6.19}$$

where

$$S_p(x) = -[p_1(x)\ln p_1(x) + p_2(x)\ln p_2(x)], \tag{6.20}$$

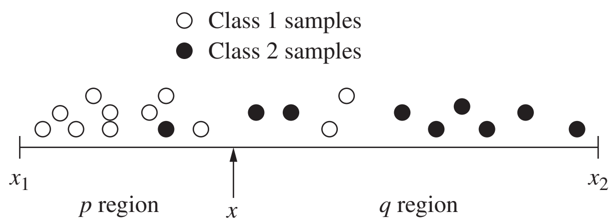

Figure 6.17 Illustration of threshold value idea.

$$S_p(x) = -[q_1(x)\ln q_1(x) + q_2(x)\ln q_2(x)], \tag{6.21}$$

where

$$\begin{aligned}
p_k(x) \text{ and } q_k(x) &= \text{conditional probabilities that the class } k \text{ sample is in the} \\
&\quad \text{region } [x_1, x_1 + x] \text{ and } [x_1 + x, x_2], \text{ respectively} \\
p(x) \text{ and } q(x) &= \text{probabilities that all samples are in the region } [x_1, x_1 + x] \\
&\quad \text{and } [x_1 + x, x_2], \text{ respectively} \\
p(x) + q(x) &= 1.
\end{aligned}$$

A value of x that gives the minimum entropy is the optimum threshold value. We calculate entropy estimates of $p_k(x)$, $q_k(x)$, $p(x)$, and $q(x)$, as follows (Christensen, 1980):

$$p_k(x) = \frac{n_k(x) + 1}{n(x) + 1}, \tag{6.22}$$

$$q_k(x) = \frac{N_k(x) + 1}{N(x) + 1}, \tag{6.23}$$

$$p(x) = \frac{n(x)}{n}, \tag{6.24}$$

$$q(x) = 1 - p(x), \tag{6.25}$$

where

$$\begin{aligned}
n_k(x) &= \text{number of class } k \text{ samples located in } [x_l, x_l + x] \\
n(x) &= \text{the total number of samples located in } [x_l, x_l + x] \\
N_k(x) &= \text{number of class } k \text{ samples located in } [x_l + x, x_2] \\
N(x) &= \text{the total number of samples located in } [x_l + x, x_2] \\
n &= \text{total number of samples in } [x_1, x_2] \\
l &= \text{a general length along the interval } [x_1, x_2].
\end{aligned}$$

While moving x in the region $[x_1, x_2]$ we calculate the values of entropy for each position of x. The value of x that holds the minimum entropy we will call the primary threshold (PRI) value. With this PRI value, we divide the region $[x_1, x_2]$ in two. We may say that the left side of the primary threshold is the *negative* side and the right the *positive* side; these labels are purely arbitrary but should hold some contextual meaning for the particular problem. With this first PRI value we can choose a shape for the two membership functions; one such shape uses two trapezoids, as seen in Figure 6.18a. But the particular choice of shape is arbitrary; we could just as well have chosen to make the threshold crisp and use two rectangles as membership functions. However, we do want to employ some amount of overlap

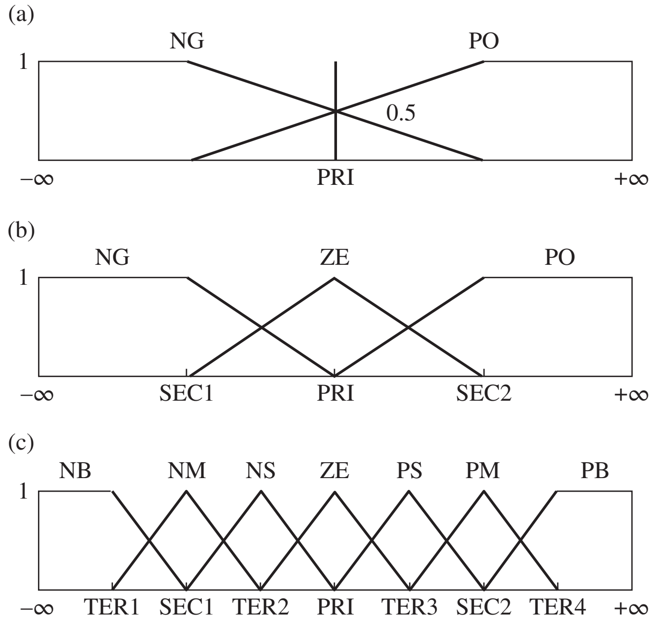

Figure 6.18 Repeated partitions and corresponding fuzzy set labels: (a) the first partition, (b) the second partition, and (c) the third partition.

since this develops the power of a membership function. As we get more and more subdivisions of the region $[x_1, x_2]$, the choice of shape for the membership function becomes less and less important as long as there is overlap between sets. Therefore, selection of simple shapes like triangles and trapezoids, which exhibit some degree of overlap, is judicious. In the next sequence we conduct the segmentation again, on each of the regions shown in Figure 6.18a; this process will determine *secondary* threshold values. The same procedure is applied to calculate these secondary threshold values. If we denote a secondary threshold in the negative area as SEC1 and the other secondary threshold in the positive area SEC2, we now have three threshold lines in the sample space. The thresholds SEC1 and SEC2 are the minimum entropy points that divide the respective areas into two classes. Then we can use three labels of PO (positive), ZE (zero), and NG (negative) for each of the classes, and the three threshold values (PRI, SEC1, SEC2) are used as the *toes* of the three separate membership shapes shown in Figure 6.18b. In fuzzy logic applications we often use an odd number of membership functions to partition a region, say five labels or seven. To develop seven partitions we would need *tertiary* threshold values in each of the three classes of Figure 6.18b. Each threshold level, in turn, gradually separates the region into more and more classes. We have four tertiary threshold values: TER1, TER2, TER3, and TER4. Two of the tertiary thresholds lie between primary and secondary thresholds, and the other two lie between secondary thresholds and the ends of the sample space; this arrangement is shown in Figure 6.18c. In this figure we use labels such as negative-big (NB), negative-medium (NM), negative-small (NS), ZE, positive-small (PS), positive-medium (PM), and positive-big (PB).

Example 6.6

The shape of an ellipse may be characterized by the ratio of the length of two chords a and b, as shown in Figure 6.19 (a similar problem was originally posed in Chapter 1; see Figure 6.3).

Let $x = a/b$; then as the ratio $a/b \rightarrow \infty$, the shape of the ellipse tends to a horizontal line, whereas as $a/b \rightarrow 0$, the shape tends to a vertical line. For $a/b = 1$ the shape is a circle. Given a set of a/b values that have been classified into two classes (class division is not necessarily based on the value of x alone; other properties like line thickness, shading of the ellipse, etc., may also be criteria), divide the variable $x = a/b$ into fuzzy partitions, as illustrated in Table 6.11.

First we determine the entropy for different values of x. The value of x is selected as approximately the mid-value between any two adjacent values. Equations (6.19) to (6.25) are then used to compute p_1, p_2, q_1, q_2, $p(x)$, $q(x)$, $S_p(x)$, $S_q(x)$, and S ; and the results are displayed in Table 6.12. The value of x that gives the minimum value of the entropy (S) is selected as the first threshold partition point, PRI. From Table 6.12 (see checkmark at $S = 0.4$) we see that the first partition point is selected at $x = 1.5$, and its location for determining membership function selection is shown in Figure 6.20.

The same process as displayed in Table 6.12 is repeated for the negative and positive partitions for different values of x. For example, in determining the threshold value to partition the negative (NG) side of Figure 6.20, Table 6.13 displays the appropriate calculations.

Table 6.14 illustrates the calculations to determine the threshold value to partition the positive side of Figure 6.20.

The partitions are selected based on the minimum entropy principle; the S values with a checkmark in Tables 6.13 and 6.14 are those selected. The resulting fuzzy partitions are as shown in Figure 6.21. If required, these partitions can be further subdivided into more fuzzy subpartitions of the variable x.

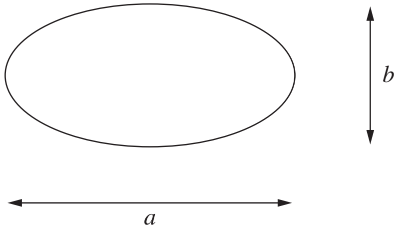

Figure 6.19 Geometry of an ellipse.

Table 6.11 Segmentation of x into two arbitrary classes (from raw data).

$x = a/b$	0	0.1	0.15	0.2	0.2	0.5	0.9	1.1	1.9	5	50	100
Class	1	1	1	1	1	2	1	1	2	2	2	2

Table 6.12 Calculations for selection of partition point PRI.

x	0.7	1.0	1.5	3.45
p_1	$\frac{5+1}{6+1}=\frac{6}{7}$	$\frac{6+1}{7+1}=\frac{7}{8}$	$\frac{7+1}{8+1}=\frac{8}{9}$	$\frac{7+1}{9+1}=\frac{8}{10}$
p_2	$\frac{1+1}{6+1}=\frac{2}{7}$	$\frac{1+1}{7+1}=\frac{2}{8}$	$\frac{1+1}{8+1}=\frac{2}{9}$	$\frac{2+1}{9+1}=\frac{3}{10}$
q_1	$\frac{2+1}{6+1}=\frac{3}{7}$	$\frac{1+1}{5+1}=\frac{2}{6}$	$\frac{0+1}{4+1}=\frac{1}{5}$	$\frac{0+1}{3+1}=\frac{1}{4}$
q_2	$\frac{4+1}{6+1}=\frac{5}{7}$	$\frac{4+1}{5+1}=\frac{5}{6}$	$\frac{4+1}{4+1}=1.0$	$\frac{3+1}{3+1}=1.0$
$P(x)$	$\frac{6}{12}$	$\frac{7}{12}$	$\frac{8}{12}$	$\frac{9}{12}$
$q(x)$	$\frac{6}{12}$	$\frac{5}{12}$	$\frac{4}{12}$	$\frac{3}{12}$
$S_P(x)$	0.49	0.463	0.439	0.54
$S_q(x)$	0.603	0.518	0.32	0.347
S	0.547	0.486	0.4√	0.49

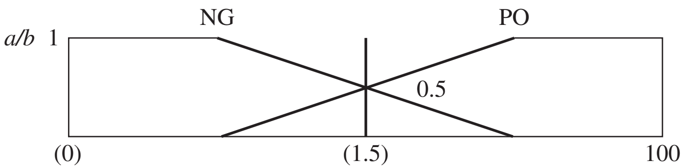

Figure 6.20 Partitioning of the variable $x = a/b$ into positive (PO) and negative (NG) partitions.

Table 6.13 Calculations to determine secondary threshold value: NG side.

x	0.175	0.35	0.7
p_1	$\frac{3+1}{3+1}=1.0$	$\frac{5+1}{5+1}=1.0$	$\frac{5+1}{6+1}=\frac{6}{7}$
p_2	$\frac{0+1}{3+1}=\frac{1}{4}$	$\frac{0+1}{5+1}=\frac{1}{6}$	$\frac{1+1}{6+1}=\frac{2}{7}$
q_1	$\frac{4+1}{5+1}=\frac{5}{6}$	$\frac{2+1}{3+1}=\frac{3}{4}$	$\frac{2+1}{2+1}=1.0$
q_2	$\frac{1+1}{5+1}=\frac{2}{6}$	$\frac{1+1}{3+1}=\frac{2}{4}$	$\frac{0+1}{2+1}=\frac{1}{3}$
$P(x)$	$\frac{3}{8}$	$\frac{5}{8}$	$\frac{6}{8}$
$q(x)$	$\frac{5}{8}$	$\frac{3}{8}$	$\frac{2}{8}$
$S_P(x)$	0.347	0.299	0.49
$S_q(x)$	0.518	0.562	0.366
S	0.454	0.398√	0.459

Table 6.14 Calculations to determine secondary threshold value: PO side.

x	27.5
p_1	$\frac{0+1}{2+1}=\frac{1}{3}$
p_2	$\frac{2+1}{2+1}=1.0$
q_1	$\frac{0+1}{2+1}=\frac{1}{3}$
q_2	$\frac{2+1}{2+1}=1.0$
$P(x)$	$\frac{2}{4}$
$q(x)$	$\frac{2}{4}$
$S_P(x)$	0.366
$S_q(x)$	0.366
S	0.366√

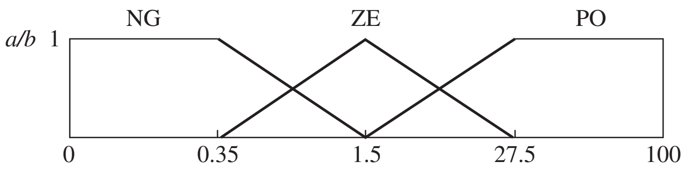

Figure 6.21 Secondary partitioning for Example 6.6.

Summary

This chapter attempts to summarize several methods—classical and modern—that have been and are being used to develop membership functions. This field is rapidly developing, and this chapter is simply an introduction. Many methods for developing membership functions have not been discussed. Ideas such as deformable prototypes (Bremermann, 1976), implicit analytical definition (Kochen and Badre, 1976), relative preferences (Saaty, 1974), and various uses of statistics (Dubois and Prade, 1980) are just a few of the many omitted here for brevity.

This chapter has dealt at length with only six of the methods currently used in developing membership functions. There are a growing number of papers in the area of cognitive systems, where learning methods like neural networks and reasoning systems like fuzzy systems are being combined to form powerful problem solvers. In these cases, the membership functions are generally tuned in a cyclic fashion and are inextricably tied to their associated rule structure (e.g., see Hayashi, Nomura, Yamasaki, and Wakami, 1992).

In the case of genetic algorithms, a number of works have appeared (see Karr and Gentry, 1993; Lee and Takagi, 1993). Vast improvements have been made in the ability of genetic algorithms to find optimum solutions: for example, the *best* shape for a membership function. One of these improvements makes use of gray codes in solving a traditional binary coding problem, where sometimes all the bits used to map a decimal number had to be changed to increase that number by 1 (Forrest, 1993). This problem had made it difficult for some algorithms to find an optimum solution from a point in the solution space that was already close to the optimum. Both neural network and genetic algorithm approaches to determining membership functions generally make use of associated rules in the knowledge base.

In inductive reasoning, as long as the database is not dynamic the method will produce good results; when the database changes, the partitioning must be reaccomplished. Compared to neural networks and genetic algorithms, inductive reasoning has an advantage in the fact that the method may not require a convergence analysis, which in the case of genetic algorithms and neural networks is computationally expensive. On the other hand, the inductive reasoning method uses the entire database to formulate rules and membership functions and, if the database is large, this method can also be computationally expensive. The choice of which of the three methods to use depends entirely on the problem size and problem type.

An illustration of the use of inductive reasoning on a modern research problem is given in Azarbayejani, Reda Taha, and Ross (2008). In this research the authors showed the use of inductive reasoning to develop the membership functions to describe damage states in a multistory frame structure. Information entropy was used to identify the fuzzy damage sets, and a fuzzy pattern recognition method (such as those described in Chapter 10) was used to specify the level of damage (e.g., light, severe) in the structure.

References

Azarbayejani, M., Reda Taha, M. M., and Ross, T. J. (2008). An inductive fuzzy damage classification approach for structural health monitoring. *Int. J. Mater. Struct. Integrity*, **2** (3), 193–206.

Bremermann, H. (1976). Pattern recognition, in *Systems Theory in the Social Sciences*, in H. Bossel, S. Klaszko, and N. Müler, eds., pp. 116–159 (Basel: Birkhaeuser).

Christensen, R. (1980). *Entropy Minimax Sourcebook*: *Fundamentals of Inductive Teasoning*, Vols. 1–4, (Lincoln, MA: Entropy Ltd.).

De Luca, A., and Termini, S. (1972). A definition of a non-probabilistic entropy in the setting of fuzzy sets theory. *Inf. Control*, **20**, 301–312.

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications* (New York: Academic Press).

Einstein, A. (1922). Geometry and experience, in *Sidelights of Relativity* (London: Methuen).

Forrest, S. (1993). Genetic algorithms: Principles of natural selection applied to computation. *Science*, **261**, 872–878.

Goldberg, D. (1989). *Genetic Algorithms* (New York: Addison-Wesley).

Hayashi, I., Nomura, H., Yamasaki, H., and Wakami, N. (1992). Construction of fuzzy inference rules by NDF and NDFL. *Int. J. Approx. Reasoning*, **6**, 241–266.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, **79**, 2554–2558.

Hopfield, J., and Tank, D. (1986). Computing with neural circuits: a model. *Science*, **233**, 625–633.

Karr, C. L., and Gentry, E. J. (1993). Fuzzy control of pH using genetic algorithms. *IEEE Trans. Fuzzy Syst.*, **1** (1), 46–53.

Kochen, M., and Badre, A. (1976). On the precision of adjectives which denote fuzzy sets. *J. Cybern.*, **4** (1), 49–59.

Lee, M., and Takagi, H. (1993). Integrating design stages of fuzzy systems using genetic algorithms. *IEEE Trans.*, Paper 0-7803-0614-7/93.
Luger, G., and Stubblefield, W. (1989). *Artificial Intelligence and the Design of Expert Systems* (Redwood City, CA: Benjamin-Cummings).
Ross, T. (1995). *Fuzzy Logic with Engineering Applications* (New York: McGraw-Hill).
Saaty, T. (1974). Measuring the fuzziness of sets. *J. Cybern.*, **4** (4), 53–61.
Takagi, H., and Hayashi, I. (1991). NN-driven fuzzy reasoning. *Int. J. Approx. Reasoning*, **5**, 191–212.
Yager, R., and Filev, D. (1994). Template-based fuzzy systems modeling. *Intell. Fuzzy Syst.*, **2** (1), 39–54.
Yamakawa, T. (1992). A fuzzy logic controller. *J. Biotechnol.*, **24**, 1–32.
Zadeh, L. (1972). A rationale for fuzzy control. *J. Dyn. Syst. Meas. Control Trans. ASME*, **94**, 3–4.

Problems

6.1 Using your intuition, develop fuzzy membership functions on the real line for the fuzzy number 8, using the following function shapes
   a. Symmetric triangle
   b. Trapezoid
   c. Gaussian function

6.2 Using your own intuition develop fuzzy membership functions on the real line for the fuzzy member "approximately 4 or approximately 6" using the following function shapes:
   a. Symmetric triangle
   b. Trapezoid
   c. Gaussian function

6.3 Using your own intuition and your own definition of the universe discourse, plot fuzzy membership functions for the following variables:
   a. Difficulty level
      i. Beginners
      ii. Intermediate
      iii. Professionals

   b. Size of column
      i. Fairly large
      ii. Large
      iii. Very large
      iv. Not very large
      v. More or less large

6.4 Using the inference approach outlined in this chapter, find the membership values for each of the triangular shapes for each of the following triangles (numbers are in degrees):
   a. 90, 50, 40
   b. 120, 40, 20
   c. 40, 65, 75

6.5 Develop a membership function for rectangles that is similar to the algorithm on triangles in this chapter. This function should have two independent variables; hence, it can be plotted.

6.6 For the data shown in the accompanying table, show the first iteration in trying to compute the membership values for the input variables x_1, x_2, and x_3 in the output regions R^1

and R^2. Assume a random set of weights for your neural network. Use a $3 \times 3 \times 2$ neural network.

X_1	X_2	X_3	R^1	R^2
0.3	2.4	2	0	1

6.7 For the data shown in the accompanying table, show the first iteration in trying to compute the membership values for the input variables x_1, x_2, and x_3 in the output regions R^1 and R^2. Assume a random set of weights for your neural network.

X_1	X_2	X_3	R^1	R^2
1.2	0.8	2	1	0

a. Use a $3 \times 3 \times 1$ neural network.
b. Use a $3 \times 3 \times 2$ neural network.
c. Explain the difference in results when using networks in parts a and b.

6.8 For the data shown in the accompanying Table A, show the first two iterations using a genetic algorithm in trying to find the optimum membership functions (use right-triangle functions) for the input variable x and output variable y in the rule table, Table B.

Table A
Data

X	0	45	90
Y	0	0.71	1

Table B
Rules

X	SM	MD
Y	SM	LG

For the rule table, the labels SM, MD and LG mean small, medium and large, respectively.

6.9 For the data shown in the following Table A, show the first two iterations using a genetic algorithm in trying to find the optimum membership functions (use right-triangle functions) for the input variable x and output variable y in the rule table, Table B. For the rule table, Table B, the symbols ZE, S, and LG mean zero, small, and large, respectively.

Table A
Data.

X	0	0.3	0.6	1.0	100
Y	1	0.74	0.55	0.37	0

Table B
Rules.

X	LG	S
Y	ZE	S

6.10 The results of a price survey for 30 automobiles is presented here:

Class	Automobile prices (in units of \$1000)
Economy	5.5, 5.8, 7.5, 7.9, 8.2, 8.5, 9.2, 10.4, 11.2, 13.5
Midsize	11.9, 12.5, 13.2, 14.9, 15.6, 17.8, 18.2, 19.5, 20.5, 24.0
Luxury	22.0, 23.5, 25.0, 26.0, 27.5, 29.0, 32.0, 37.0, 43.0, 47.5

Consider the automobile prices as a variable and the classes as economy, midsize, and luxury automobiles. Develop three membership function envelopes for car prices using the method of inductive reasoning.

6.11 The ductility of a long slender steel rod under tension may be characterized by a decrease in diameter, d, together with a significant increase in elongation, L. Let $x = d/L$, where x is dimensionless. As x approaches zero the elongation is maximized and the subsequent failure of the member in tension will occur. However, before the member will fail, it will first go from an elastic state, to a plastic state.

Failure can be defined as exceeding the elastic limit, or in breaking apart in the plastic state. We define two failure states as a function of x: state (1) is where the member exceeds the elastic limit but can still support further tension; state (2) where the member exceeds the onset of plastic deformation and will break if further tension is applied. The following table defines tests on 12 slender rods, whose final state is shown. Divide x into fuzzy partitions and develop the associated membership functions using inductive reasoning.

State	Deformations
$x = d/L$	0.11, 0.10, 0.08, 0.06, 0.04, 0.03, 0.01, 0.009, 0.007, 0.005, 0.003, 0
state	1, 1, 1, 1, 1, 2, 1, 1, 2, 2, 2, 2

6.12 The following raw data were determined in a pairwise comparison of new premium car preferences in a poll of 100 people. When it was compared with a Porsche (P), 79 of those polled preferred a BMW (B), 85 preferred a Mercedes (M), 59 preferred a Lexus (L), and 67 preferred an Infinity (I). When a BMW was compared, the preferences were 21 – P, 23 – M, 37 – L, and 45 – I. When a Mercedes was compared, the preferences were 15 – P, 77 – B, 35 – L, and 48 – I. When a Lexus was compared, the preferences were 41 – P, 63 – B, 65 – M, and 51 – I. Finally, when an Infinity was compared, the preferences were 33 – P, 55 – B, 52 – M, and 49 – L. Using rank ordering, plot the membership function for "most preferred car."

7

Automated Methods for Fuzzy Systems

Measure what is measurable, and make measurable what is not so.
Galileo Galilei, circa 1630

For my part I know nothing with any certainty, but the sight of the stars makes me dream.
Vincent van Gogh, twentieth-century painter

It is often difficult or impossible to accurately model complicated natural processes or engineered systems using a conventional nonlinear mathematical approach with limited prior knowledge. Ideally, the analyst uses the information and knowledge gained from prior experiments or trials with the system to develop a model and predict the outcome, but for new systems where little is known or where experimental analyses are too costly to perform (e.g., astronomy), prior knowledge and information is often unavailable. This lack of data on, or extensive knowledge of, the system makes developing a model using conventional methods extremely difficult and often impossible. Furthermore, forming a linguistic rule-base of the system may be impractical without conducting additional observations. Fortunately, for situations such as these, fuzzy modeling is practical and can be used to develop a model for the system using the "limited" available information. *Batch least squares (BLS), recursive least squares (RLS), gradient method (GM), learning from example (LFE), modified learning from example (MLFE),* and *clustering method (CM)* are some of the algorithms available for developing a fuzzy model (Passino and Yurkovich, 1998). The choice of which method to implement depends on factors such as the amount of prior knowledge of the system to be modeled. These methods, which are referred to as *automated methods*, are provided as additional procedures to develop membership functions, like those in Chapter 6, and to provide rules as well.

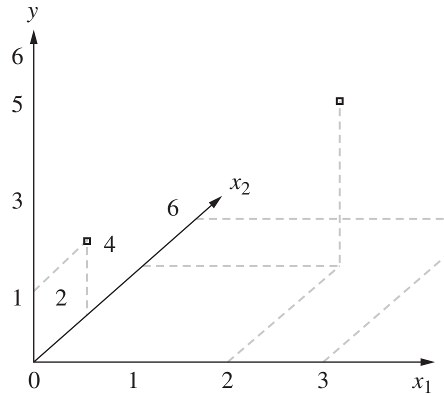

Figure 7.1 Example of two-input, single-output system for three data points. Reproduced by permission of Kevin M. Passino and Stephen Yurkovich, from: Kevin M. Passino and Stephen Yurkovich, *Fuzzy Control*, Addison Wesley Longman, Menlo Park, CA, 1998.

Definitions

The description of these methods provided by Passino and Yurkovich (1998) is expanded in this chapter with a detailed example for developing a fuzzy model using each of the algorithms mentioned previously. The explanation is given in such a manner that allows the reader to easily prepare a MATLAB code for other applications (see the preface for instructions on accessing this software). Only two-input, single-output systems are illustrated here, but the algorithms can be extended to multiple-input, single-output systems and even multiple-input, multiple-output systems. An example of a two-input, single-output system is illustrated in Figure 7.1, where the information is provided by three points and where the inputs are x_1 and x_2 and the output is y. Most of the algorithms used in the examples of this chapter incorporate Gaussian membership functions for the inputs $\mu(x)$,

$$\mu(x) = \exp\left[-\frac{1}{2}\left(\frac{x_i - c_i}{\sigma_i}\right)^2\right], \tag{7.1}$$

where x_i is the ith input variable, c_i is the ith center of the membership function (i.e., where the membership function achieves a maximum value), and σ_i is a constant related to the spread of the ith membership function. Figure 7.2 illustrates a typical Gaussian membership function and these parameters.

For demonstrative purposes triangular membership functions are used in the example given for the LFE algorithm; Equation (7.2) shows the formula used for the triangular membership function, whereas Figure 7.3 illustrates the membership function. In fact, any type of membership function may be used for the input and output functions but only the Gaussian and triangular membership functions are illustrated here. In most of the examples provided in this chapter, the output membership function is a delta function, which is an impulse function of zero width with only one value with full membership located at b_i and all other values have zero

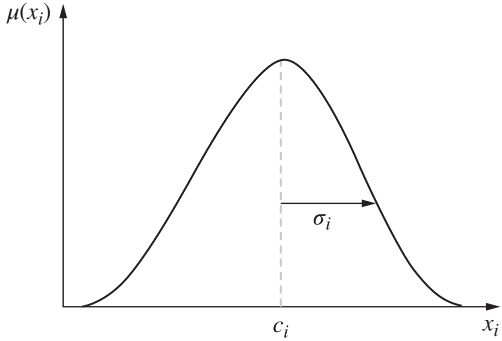

Figure 7.2 Typical Gaussian membership function.

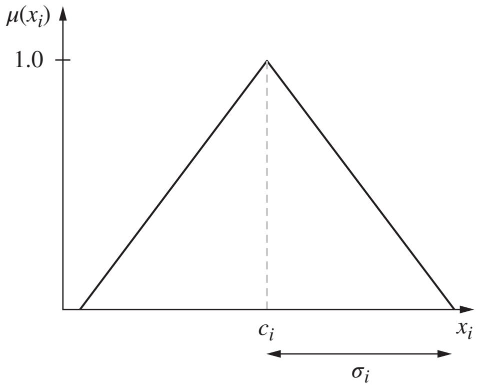

Figure 7.3 Typical triangular membership function.

(Figure 7.4). However, the algorithms provided here accommodate any type of output membership function.

$$\mu(x) = \begin{cases} \max\left\{0, 1 + \dfrac{x_i - c_i}{0.5\sigma_i}\right\}, & \text{if } x_i \le c_i \\ \max\left\{0, 1 + \dfrac{c_i - x_i}{0.5\sigma_i}\right\}, & \text{otherwise.} \end{cases} \tag{7.2}$$

The six automated methods presented here either develop a rule-base or use a predetermined rule-base (such as the LFE method, batch, and RLS algorithms) to model the system and predict outputs given the inputs; in any case the rules comprise a premise clause and a consequence. A typical example of a rule for a multiple-input, single-output system is as follows:

IF *premise*$_1$ and *premise*$_2$ THEN *consequence*.

These rules are developed by the algorithms to predict and/or govern an output for the system with given inputs. Most importantly, the algorithms incorporate the use of fuzzy values rather

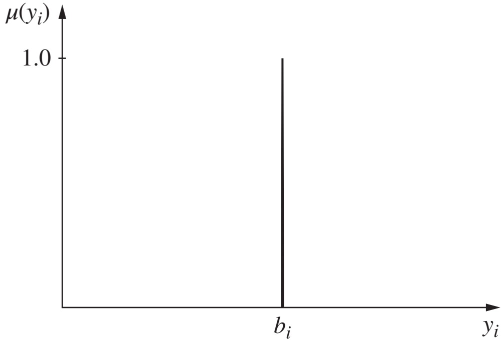

Figure 7.4 Delta membership function.

than fuzzy linguistic terms in these rules (hence the membership functions). In other words, the premise and consequence are fuzzy values. In the BLS, RLS, and GM algorithms, this rule-base must be specified by the user of the algorithm from other automated procedures (e.g., MLFE); however, the GM has the capability to update the parameters of the rule-base (i.e., the parameters of the membership functions). The CM and MLFE form a rule-base from the input–output, which is then used to model the system. The LFE algorithm relies on complete specification of the membership functions and only constructs the rules of the rule-base. Because of this, some algorithms can be used together to develop a refined model for the system. For instance, the MLFE can be used in conjunction with the RLS to develop a more effective model. Once the parameters of the membership functions of the rule-base have been specified, they are used by the algorithms to predict an output given the inputs. A detailed description of this process is provided later in this chapter.

The examples to follow all employ a center-average defuzzification, a product t-norm for the premise and a product implication, as given in Equation (7.3). A Takagi–Sugeno or other inference mechanism may be used instead but their application and respective discussion are not included in this chapter (Passino and Yurkovich, 1998). As mentioned, most of our examples use Gaussian membership functions for the premise and delta functions for the output, resulting in the following equation to predict the output given an input data-tuple x_j:

$$f(x|\theta) = \frac{\sum_{i=1}^{R} b_i \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]}{\sum_{i=1}^{R} \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]}, \tag{7.3}$$

where R is the number of rules in the rule-base and n is the number of inputs per data-tuple. For instance, the system of Figure 7.1 has two inputs (x_1 and x_2); thus, $n=2$ and if there were two rules in the rule-base, $R=2$. The parameter R is not known a priori for some methods, but is determined by the algorithms. The symbol θ is a vector that includes the membership function parameters for the rule-base, c_i, σ_i, and b_i.

Table 7.1 Training data set, $Z = \{([x_1, x_2], y)\}$.

x_1	x_2	y
0	2	1
2	4	5
3	6	6

The data-tuples we shall use for our examples are the same as those used in Passino and Yurkovich (1998). Table 7.1 and Figure 7.1 contain these data, which are presumably a representative portion of a larger nonlinear data set, $Z = \{([x_1, x_2], y)\}$. The data of Table 7.1 are used to train the fuzzy system to model the output y given the two inputs x_1 and x_2. The BLS and RLS methods are presented first followed by the GM, CM, and LFE, and finally the MLFE methods. In consideration for space, only the training of the fuzzy set is demonstrated in each example. However, at the end of this chapter the result of a more thorough application of fuzzy modeling is presented for a system described by numerous input–output data from an experiment on a new, crushable foam material called *syntactic foam*; the strength of the foam will be verified by triaxial compression tests. Unfortunately, because of the costs involved in preparing the foam only a few cylindrical specimens (2.800 inches in length and 1.400 inches in diameter) are tested under various triaxial compressive loads to determine the longitudinal deformations associated with these loads (Figure 7.11). The collected input–output data consist of the longitudinal stress (major principal stress) and the lateral stress (minor principal stress), and their respective longitudinal deformation is the output. These input–output data are then used by various fuzzy algorithms to develop a model to predict the longitudinal deformation given the lateral and longitudinal stress.

Batch Least Squares Algorithm

The following example demonstrates the development of a nonlinear fuzzy model for the data in Table 7.1 using the BLS algorithm. The algorithm constructs a fuzzy model from numerical data, which can then be used to predict outputs given any input. Thus, the data set Z can be thought of as a training set used to model the system. When using the BLS algorithm to develop a fuzzy model it is helpful to have knowledge about the behavior of the data set to form a rule-base. In the cases where this knowledge is not available another algorithm with rule-forming capabilities (such as MLFE or CM) may be used to develop a rule-base.

To begin with, let us denote the number of input data-tuples, $m = 3$, where there are two inputs for each data-tuple, $n = 2$ (i.e., x_1 and x_2). As required by the algorithm we must designate the number of rules (two rules, $R = 2$) and the rule parameters. The consequence in each rule is denoted by the output membership function centers b_1 and b_2. Recall that there are no other parameters needed for the consequence. The two premises of each rule are defined by the input membership function centers (c_i) and their respective spread (σ_i):

$$\text{IF } premise_1 \text{ and } premise_2 \text{ THEN } consequence.$$

Say we designate values for the premise and consequence of the rule-base, which are close to the first two data-tuples of Z (presume that Z is a good representation of the data contained in a larger data set). This way the premise and consequence capture as much of the data set as possible thus improving the predictability of the model. We have the following for the input membership functions centers c_j^i, where i is the rule number and j denotes input number:

$$\begin{array}{ll} c_1^1 = 1.5 & c_1^2 = 3 \\ c_2^1 = 3 & c_2^2 = 5 \end{array}.$$

This places the peaks of the membership functions between the input portions of the training data pairs. We could make a conjecture as to the whereabouts of the respective output centers for these input centers as well, but for demonstrative purposes we use the output from the first two data sets for now. Thus, we have the following two rules in our rule-base:

Rule 1: If x_1 is ‘about 1.5’ and x_2 is ‘about 3’ then b_1 is 1.
Rule 2: If x_1 is ‘about 3’ and x_2 is ‘about 5’ then b_1 is 5.

Next we pick the spreads, σ_j^i, for the input membership functions we selected. As a good start we select $\sigma_j^i = 2$, for $i = 1, 2$ and $j = 1, 2$, to provide reasonable overlap between membership functions. We may have to increase or decrease the overlap among the input membership functions in the rule-base to improve the output of the fuzzy model. The input membership functions for Rules 1 and 2 are Gaussian membership functions and are displayed in Figures 7.5 and 7.6. The output membership functions for the rules are delta functions that are displayed in Figure 7.7.

The objective is to determine the predicted output using Equation (7.3) when given an input data-tuple. Up to now, we have only defined the rule-base but have not developed an output mapping function; we do this next using the training data set.

We calculate the membership value that each input data-tuple has in the specified rules of the rule-base and multiply these two values by one another, resulting in the membership value that the input data-tuple has in a particular rule. This is accomplished by

$$\mu_i(x) = \prod_{j=1}^{n} \exp\left(-\frac{1}{2}\left(\frac{x - c_j^i}{\sigma_j^i} \right)^2 \right), \tag{7.4}$$

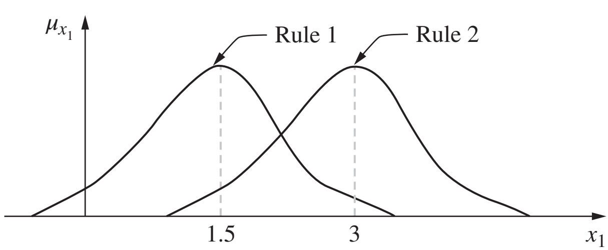

Figure 7.5 Input membership functions for x_1.

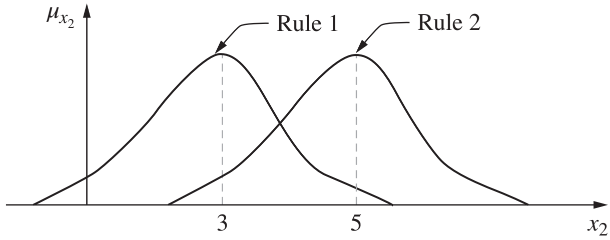

Figure 7.6 Input membership functions for x_2.

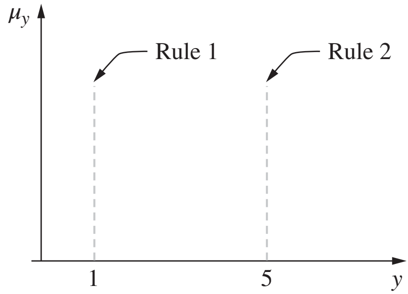

Figure 7.7 Output membership functions for y.

where $n = 2$, x are the input data-tuples, and c_j^i and σ_j^i are the rule-base input parameters. This reduces Equation (7.3) to the following form:

$$f(x|\theta) = \frac{\sum_{i=1}^{R} b_i \mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)}. \tag{7.5}$$

Passino and Yurkovich (1998) point out that this is also equal to

$$f(x|\theta) = \frac{\sum_{i=1}^{R} b_i \mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)} = \frac{b_1 \mu_1(x)}{\sum_{i=1}^{R} \mu_i(x)} + \frac{b_2 \mu_2(x)}{\sum_{i=1}^{R} \mu_i(x)} + \cdots + \frac{b_R \mu_R(x)}{\sum_{i=1}^{R} \mu_i(x)},$$

and if we define the regression vector ξ as

$$\xi_i(x) = \frac{\mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)} = \frac{\prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]}{\sum_{i=1}^{R} \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j - c_j^i}{\sigma_j^i}\right)^2\right]}, \quad \text{for } i = 1, 2 \tag{7.6}$$

we can use Equation (7.7) to calculate the output. However, because we are using the BLS approach, the output is calculated a little differently (for theory, see Passino and Yurkovich, 1998). The resulting mapping, for the BLS approach, is

$$f\left(x|\hat{\theta}\right) = \hat{\theta}^{\mathrm{T}} \xi(x), \tag{7.7}$$

where $\hat{\theta}$ is the least squares estimate vector from the training set, and $\hat{\theta}^{\mathrm{T}}$ is the transpose. The calculation of the least squares estimate and ξ_i will be explained.

For each input data-tuple, we have two ($i = 2$) values for ξ, one for Rule 1 and another for Rule 2, resulting in a total of six values:

$$\begin{array}{cc} \xi_1(x^1) & \xi_2(x^1) \\ \xi_1(x^2) & \xi_2(x^2). \\ \xi_1(x^3) & \xi_2(x^3) \end{array}$$

Using Equation (7.6), we get

$$\begin{aligned}
\xi_1(x^1) &= \frac{\exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^1}{\sigma_1^1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^1}{\sigma_2^1}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^1}{\sigma_1^1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^1}{\sigma_2^1}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^2}{\sigma_1^2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^2}{\sigma_2^2}\right)^2\right]}, \\
&= \frac{\exp\left[-\frac{1}{2}\left(\frac{0 - 1.5}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 3}{2}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{0 - 1.5}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 3}{2}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{0 - 3}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 5}{2}\right)^2\right]}, \\
&= \frac{0.66614}{0.77154} = 0.8634
\end{aligned}$$

and

$$\begin{aligned}
\xi_2(x^1) &= \frac{\exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^2}{\sigma_1^2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^2}{\sigma_2^2}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^1}{\sigma_1^1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^1}{\sigma_2^1}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{x_1 - c_1^2}{\sigma_1^2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{x_2 - c_2^2}{\sigma_2^2}\right)^2\right]}, \\
&= \frac{\exp\left[-\frac{1}{2}\left(\frac{0 - 3}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 5}{2}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{0 - 1.5}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 3}{2}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{0 - 3}{2}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2 - 5}{2}\right)^2\right]}, \\
&= 0.13661.
\end{aligned}$$

For x^2 and x^3 of data set Z we obtain the following values of $\xi_i(x)$:

$$\xi_1(x^2) = 0.5234 \quad \xi_2(x^2) = 0.4766$$
$$\xi_1(x^3) = 0.2173 \quad \xi_2(x^3) = 0.7827.$$

With $\xi_i(x)$ completely specified, the transpose of $\xi_i(x)$ is determined and placed into a matrix, $\mathbf{\Phi}$:

$$\mathbf{\Phi} = \begin{bmatrix} \xi^{\mathrm{T}}(x^1) \\ \xi^{\mathrm{T}}(x^2) \\ \xi^{\mathrm{T}}(x^3) \end{bmatrix} = \begin{bmatrix} 0.8634 & 0.1366 \\ 0.5234 & 0.4766 \\ 0.2173 & 0.7827 \end{bmatrix}.$$

And from Z we have the following outputs placed in vector Y:

$$\mathbf{Y} = \begin{bmatrix} y^1 & y^2 & y^3 \end{bmatrix}^{\mathrm{T}} = [1 \;\; 5 \;\; 6]^{\mathrm{T}}.$$

Using Y and $\mathbf{\Phi}$, we determine $\hat{\theta}$,

$$\hat{\theta} = \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{Y} \tag{7.8}$$

thus producing

$$\hat{\theta} = \left\{ \begin{bmatrix} 0.8634 & 0.5234 & 0.2173 \\ 0.1366 & 0.4766 & 0.7827 \end{bmatrix} * \begin{bmatrix} 0.8634 & 0.1366 \\ 0.5234 & 0.4766 \\ 0.2173 & 0.7827 \end{bmatrix} \right\}^{-1} * \begin{bmatrix} 0.8634 & 0.5234 & 0.2173 \\ 0.1366 & 0.4766 & 0.7827 \end{bmatrix} * \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix},$$

$$\hat{\theta} = \begin{bmatrix} 0.3647 \\ 8.1775 \end{bmatrix}.$$

Using Equation (7.7), we calculate the output for the training data set:

$$f\left(x|\hat{\theta}\right) = \hat{\theta}^{\mathrm{T}}\xi(x),$$

$$f\left(x^1|\hat{\theta}\right) = [0.3647 \quad 8.1775] * \begin{bmatrix} 0.8634 \\ 0.1366 \end{bmatrix},$$

$$= 1.4319,$$

and similarly, $f\left(x^2|\hat{\theta}\right) = 4.0883$,

$$f\left(x^3|\hat{\theta}\right) = 6.4798.$$

As seen, the fuzzy system maps the training data set reasonably accurately and if we use additional points not in the training set, as a test set, to see how the system interpolates, we find, for example,

$$f\left([1,2]^{\mathrm{T}}|\hat{\theta}\right) = 1.8267; f\left([2.5,5]^{\mathrm{T}}|\hat{\theta}\right) = 5.3981; f\left([4,7]^{\mathrm{T}}|\hat{\theta}\right) = 7.3673.$$

The accuracy of the fuzzy model developed using BLS primarily depends on the rules specified in the rule-base and the data set used to train the fuzzy model.

Recursive Least Squares Algorithm

The RLS algorithm is similar to the BLS algorithm; however, the RLS algorithm makes updating $\hat{\theta}$ much easier. The algorithm is a recursive version of the BLS method (the theory behind this algorithm is available; see Passino and Yurkovich, 1998). It operates without using all the training data and most importantly without having to compute the inverse of $\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}$ each time the $\hat{\theta}$ is updated. RLS calculates $\hat{\theta}(k)$ at each time step k from the past estimate $\hat{\theta}(k-1)$ and the latest data pair that is received, x, y^k. The following example demonstrates the training of a fuzzy model using the RLS algorithm given data set Z (see Table 7.1).

As before, we use Gaussian membership functions for the input and a delta function for the output in the rule-base. Recall that b_i is the point in the output space at which the output membership function for the ith rule is a delta function, and c_j^i is the point in the jth input universe of discourse where the membership function for the ith rule achieves a maximum. The relative width, σ_j^i, of the jth input membership function for the ith rule is always greater than zero.

The RLS algorithm requires that the rule-base be specified (i.e., number of rules, input membership function centers, input membership function relative widths, and the output centers). The training data set should include a good representative subset of the data set. If the analyst does not have enough knowledge of the system to specify the parameters needed to define the rule-base, he or she can do so by using another algorithm first, such as the MLFE. In this example, we are able to specify these parameters for the rule-base. We decide on using two rules to model the system and make an educated guess as to where to set the input membership function centers based on some type of regular spacing so as to lie in the middle of the training data, just as we did in the BLS example.

Like the BLS, we can vary the spread for each premise of the rules and thus achieve greater or lesser overlap among the input membership functions $\mu_{x_j^i}$. This is useful when dealing with inputs of different ranges where we would like the spreads of the inputs to reflect this variability. Again, we select $\sigma_j^i = 2$, for $i = 1, 2$ and $j = 1, 2$, which should provide sufficient overlap between membership functions for the data in Table 7.1. We tune $f(x|\theta)$ to interpolate the data set Z by selecting two rules ($R = 2$). If we choose the same values for c_j^i that we used in the BLS example we have,

$$\begin{array}{ll} c_1^1 = 1.5 & c_1^2 = 3 \\ c_2^1 = 3 & c_2^2 = 5 \end{array}.$$

We have two inputs for each data-tuple, $n = 2$, and three input data-tuples in our training set, $m = 3$. We assume that the training set may be increased by one each time step, so we let the time index $k=m$. In the RLS, we can cycle through the training data a number of times to develop the fuzzy model but in this example we elect to cycle through the data only once for demonstrative purposes.

Now we calculate the regression vector based on the training data set and obtain the same regression vector ξ, using Equation (7.6) as we did for the BLS example. Recall that in the least squares algorithm the training data x^i are mapped into $\xi(x^i)$ which is then used to develop an output $f(x^i)$ for the model. We get the identical results for $\xi(x^i)$ as for the BLS approach, that is,

$$\begin{array}{ll} \xi_1(x^1)=0.8634 & \xi_2(x^1)=0.1366 \\ \xi_1(x^2)=0.5234 & \xi_2(x^2)=0.4766. \\ \xi_1(x^3)=0.2173 & \xi_2(x^3)=0.7827 \end{array}$$

If we decide to use a *weighted recursive least squares* (WRLS) algorithm because the parameters of the physical system θ vary slowly, we employ a "forgetting factor," λ, which gives the more recent data more weight. The forgetting factor varies from 0 to 1, where $\lambda = 1$ results in a standard RLS. For our example, we choose to use $\lambda = 1$ to weight all the training data equally. Before proceeding to find an estimate of the output, we need to decide in what order to have the RLS process the data pairs (x^i, y^i). There are many possibilities, but in this example we choose to cycle through the data just once beginning with the first input pair and ending with the last (note that all the data-tuples are weighted equally). As mentioned, we could repeat this a number of times which may improve our results; however, for illustrative purposes we decide to cycle through just once.

For the RLS algorithm, we use a covariance matrix to determine $\hat{\theta}$, which is calculated using the regression vector and a previous covariant (Equation (7.11)). To do this, we must first calculate an initial covariance matrix P_0 using a parameter α and the identity matrix, I (Equation (7.9)). P_0 is the covariance matrix at time step 0 ($k=0$) and is used to update the covariance matrix, P, in the next time step. A recursive relation is established to calculate values of the P matrix for each time step (Equation (7.10)). The value of the parameter α should be greater than 0. Here, we use a value of $\alpha = 2000$; I, the identity matrix, is an $R \times R$ array. Next we set our initial conditions for $\hat{\theta}$, at time step 0 ($k=0$); a good starting point for this would be to use the results from our BLS example, thus

$$\hat{\theta}(0) = \begin{bmatrix} 0.3647 \\ 8.1775 \end{bmatrix}.$$

If these values are not readily available, another set of values may be used but more cycles may be needed to arrive at good values. As mentioned previously, this example only demonstrates the training of the fuzzy model using one cycle.

$$\begin{aligned} &\mathrm{P}_0 = \alpha \mathrm{I}, \\ &\mathrm{P}_0 = \mathrm{P}(0) = 2000 * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2000 & 0 \\ 0 & 2000 \end{bmatrix}. \end{aligned} \tag{7.9}$$

Once P_0 is determined, we use it along with $\xi_i(x^{k=1})$ to calculate the next P and $\hat{\theta}$ for the next step, $P(k=1)$ and $\hat{\theta}(k=1)$. This is accomplished using Equations (7.10) and (7.11):

$$P(k) = \frac{1}{\lambda}\left\{I - P(k-1)\xi(x^k)\left[\lambda I + \left(\xi(x^k)\right)^T P(k-1)\xi(x^k)\right]^{-1}\left(\xi(x^k)\right)^T\right\}P(k-1). \tag{7.10}$$

$$\hat{\theta}(k) = \hat{\theta}(k-1) + P(k)\xi(x^k)\left[y^k - \left(\xi(x^k)\right)^T\hat{\theta}(k-1)\right]. \tag{7.11}$$

For $k=1$ and $\xi_1(x^1) = 0.8634$, $\xi_2(x^1) = 0.1366$,

$$\begin{aligned}
P(1) &= \frac{1}{1}\left\{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \left(\begin{bmatrix} 2000 & 0 \\ 0 & 2000 \end{bmatrix} * \begin{bmatrix} 0.8634 \\ 0.1366 \end{bmatrix}\right) * \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.8634 & 0.1366 \end{bmatrix}\right.\right. \\
&\quad \left.\left. * \begin{bmatrix} 2000 & 0 \\ 0 & 2000 \end{bmatrix} * \begin{bmatrix} 0.8364 \\ 0.1366 \end{bmatrix}\right)^{-1} * \begin{bmatrix} 0.8634 & 0.1366 \end{bmatrix}\right\} * \begin{bmatrix} 2000 & 0 \\ 0 & 2000 \end{bmatrix}, \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.9749 & 0.1543 \\ 0.1543 & 0.0244 \end{bmatrix}\right) * \begin{bmatrix} 2000 & 0 \\ 0 & 2000 \end{bmatrix}, \\
&= \begin{bmatrix} 50.12 & -308.5 \\ -308.5 & 1951 \end{bmatrix}.
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}(1) &= \begin{bmatrix} 0.3647 \\ 8.1775 \end{bmatrix} + \begin{bmatrix} 50.12 & -308.5 \\ -308.5 & 1951 \end{bmatrix} * \begin{bmatrix} 0.8634 \\ 0.1366 \end{bmatrix} * \left(1 - \begin{bmatrix} 0.8634 & 0.1366 \end{bmatrix} * \begin{bmatrix} 0.3647 \\ 8.1775 \end{bmatrix}\right), \\
&= \begin{bmatrix} -0.1232 \\ 8.1007 \end{bmatrix}.
\end{aligned}$$

The next time step, with $k=2$ and $\xi_1(x^2) = 0.5234$, $\xi_2(x^2) = 0.4766$, results in the following:

$$P(2) = \begin{bmatrix} 2.1193 & -3.1614 \\ -3.1614 & 8.7762 \end{bmatrix}.$$

$$\hat{\theta}(2) = \begin{bmatrix} -0.6016 \\ 11.1438 \end{bmatrix}.$$

Finally for the third time step of the cycle, $k=3$ and $\xi_1(x^3) = 0.2173$, $\xi_2(x^3) = 0.7827$, results are

$$P(3) = \begin{bmatrix} 1.3684 & -0.8564 \\ -0.8564 & 1.7003 \end{bmatrix}.$$

$$\hat{\theta}(3) = \begin{bmatrix} 0.3646 \\ 8.1779 \end{bmatrix}.$$

We have now calculated the vector parameters $\hat{\theta}$ based on the three inputs needed to model the system. For this example, performing another cycle with the training data set changes $\hat{\theta}$ very little; this is left as an exercise at the end of the chapter. Now that $\hat{\theta}$ has been determined it is used in conjunction with ξ in Equation (7.7) to calculate the resulting output values for the training data-tuples:

$$f\left(x^1|\hat{\theta}\right) = [0.3646 \quad 8.1779] * \begin{bmatrix} 0.8634 \\ 0.1366 \end{bmatrix} = 1.432,$$

$$f\left(x^2|\hat{\theta}\right) = [0.3646 \quad 8.1779] * \begin{bmatrix} 0.5234 \\ 0.4766 \end{bmatrix} = 4.088,$$

$$f\left(x^3|\hat{\theta}\right) = [0.3646 \quad 8.1779] * \begin{bmatrix} 0.2173 \\ 0.7827 \end{bmatrix} = 6.480,$$

This compares well with the original output (Table 7.1). Modifying the input membership function parameters may improve the predicted output; this is left as an exercise at the end of this chapter.

Gradient Method

In the RLS method, we noticed that the predicted output for the training data set could have been improved and recommended modifying the input parameters. The GM does just that and provides a means for tuning the parameters of the fuzzy model (i.e., the parameters of the rule-base). Recall that for the input membership function, we have the membership function centers and the spread of the membership functions. In addition to the input parameters, the GM provides a method to tune the output membership function.

Using the training data set Z of Table 7.1, we illustrate the development of a fuzzy model using the GM. Like the least squares algorithms, we must specify the rules; however, the GM has the ability to tune the parameters associated with the rules based on the training set. Thus, the data used in the training set are of utmost importance in achieving a good approximation. We shall illustrate the method with two rules ($R = 2$). The GM's goal is to minimize the error between the predicted output value, $f(x^m|\theta)$, and the actual output value y^m through the quadratic function e_m, which we call the "error surface." The equation for this error surface is

$$e_m = \frac{1}{2}[f(x^m|\theta) - y^m]^2. \tag{7.12}$$

Here, m denotes the input–output data-tuple from the training data set. We want to find the minimum value on this error surface that may be used to determine when the model has achieved desired predictability. In this example, we demonstrate how cycling through the training data updates the rule-base parameters thus reducing the difference between the predicted output and the actual output as provided here,

$$\varepsilon_m = f(x^m|\theta) - y^m. \tag{7.13}$$

We can keep cycling through the training data set each time step (k) modifying the rule parameters, thus decreasing ε_m and obtaining an improved fuzzy system. In this example, we update the rule-base using the first data-tuple of Table 7.1 which is then used in the first time step. The second and third time steps, for the remaining data-tuples in Table 7.1, are reserved for an exercise at the end of the chapter.

The GM requires that a step size λ be specified for each of the three parameters being tuned (b_i, c_j^i, and σ_j^i), which are used by the algorithm to determine the updated rule parameters and decrease the error value. Selecting a large step size will converge faster but may risk overstepping the minimum value of e_m, and selecting a small step size means the parameter converges very slowly (Passino and Yurkovich, 1998). Here, we designate the step size for the error surface of the output membership centers, input membership centers, and input membership spreads equal to 1, λ_1, λ_2, and $\lambda_3 = 1$, respectively. In this example, the step size values were selected primarily to simplify the calculations.

The algorithm requires that initial values for the rules be designated, but these rules are updated through the iterations with each time step (i.e., the next data-tuple). Thus, to initiate the algorithm for the first rule, we choose x^1, y^1 for the input and output membership function centers and select the input spreads to be equal to 1. For the second rule, we choose x^2, y^2 as the input and output membership function centers and select the input spreads to be equal to 1. It is important to note that these values initiate the algorithm and are updated by the process to obtain a better model to predict the output in the first time step (i.e., $k=1$). These initiating values correspond to the zero time step ($k=0$):

$$\begin{bmatrix} c_1^1(0) \\ c_2^1(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \begin{bmatrix} \sigma_1^1(0) \\ \sigma_2^1(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b_1(0) = 1$$
$$\begin{bmatrix} c_1^2(0) \\ c_2^2(0) \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \begin{bmatrix} \sigma_1^2(0) \\ \sigma_2^2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b_2(0) = 5.$$

Let us calculate the predicted outputs for the current fuzzy model. First we need to calculate the membership values for data-tuples of Table 7.1, using

$$\mu_i(x^m, k=0) = \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j^m - c_j^i(k=0)}{\sigma_j^i(k=0)}\right)^2\right]. \tag{7.14}$$

$$\mu_1(x^1, 0) = \exp\left[-\frac{1}{2}\left(\frac{0-0}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2-2}{1}\right)^2\right] = 1.$$

$$\mu_1(x^2, 0) = \exp\left[-\frac{1}{2}\left(\frac{2-0}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{4-2}{1}\right)^2\right] = 0.0183156.$$

$$\mu_1(x^3, 0) = \exp\left[-\frac{1}{2}\left(\frac{3-0}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{6-2}{1}\right)^2\right] = 3.72665 \times 10^{-6}.$$

$$\mu_2(x^1, 0) = \exp\left[-\frac{1}{2}\left(\frac{0-2}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{2-4}{1}\right)^2\right] = 0.0183156.$$

$$\mu_2(x^2, 0) = \exp\left[-\frac{1}{2}\left(\frac{2-2}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{4-4}{1}\right)^2\right] = 1.$$

$$\mu_2(x^3, 0) = \exp\left[-\frac{1}{2}\left(\frac{3-2}{1}\right)^2\right] * \exp\left[-\frac{1}{2}\left(\frac{6-4}{1}\right)^2\right] = 0.082085.$$

From the membership values the training data set has in the current rule-base, we obtain the fuzzy output from Equation (7.3) as follows:

$$f(x^m|\theta(k=0)) = \frac{\sum_{i=1}^{R} b_i(0) \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j^m - c_j^i(k=0)}{\sigma_j^i(k=0)}\right)^2\right]}{\sum_{i=1}^{R} \prod_{j=1}^{n} \exp\left[-\frac{1}{2}\left(\frac{x_j^m - c_j^i(k=0)}{\sigma_j^i(k=0)}\right)^2\right]}.$$

$$f(x^2|\theta(0)) = \frac{1*\exp\left[-\frac{1}{2}\left(\frac{2-0}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{4-2}{1}\right)^2\right] + 5*\exp\left[-\frac{1}{2}\left(\frac{2-2}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{4-4}{1}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{2-0}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{4-2}{1}\right)^2\right] + \exp\left[-\frac{1}{2}\left(\frac{2-2}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{4-4}{1}\right)^2\right]},$$

$$= \frac{1*\mu_1(x^2,0) + 5*\mu_2(x^2,0)}{\mu_1(x^2,0)+\mu_2(x^2,0)},$$

$$= \frac{1*0.0183156 + 5*1}{1.0183156} = 4.92805.$$

$$f(x^3|\theta(0)) = \frac{1*\mu_1(x^3,0) + 5*\mu_2(x^3,0)}{\mu_1(x^3,0)+\mu_2(x^3,0)},$$

$$= \frac{1*0.00000372665 + 5*0.082085}{0.00000372665 + 0.082085} = 4.999818.$$

$$f(x^1|\theta(0)) = \frac{1*\mu_1(x^1,0) + 5*\mu_2(x^1,0)}{\mu_1(x^1,0)+\mu_2(x^1,0)},$$

$$= \frac{1*1 + 5*0.0183156}{1+0.0183156} = 1.0719447.$$

To compute the approximate error between the predicted output values and the actual output values, we use Equation (7.12):

$$e_m = \frac{1}{2}[f(x^m|\theta(k=0)) - y^m]^2.$$

$$e_1 = \frac{1}{2}[1.0719447-1]^2 = 2.58802 \times 10^{-3}.$$

$$e_2 = \frac{1}{2}[4.928055-5]^2 = 2.58802 \times 10^{-3}.$$

$$e_3 = \frac{1}{2}[4.999818-6]^2 = 0.500182.$$

From these results it can be seen that the algorithm maps the first two data points much better than the third. The predicted output is improved by cycling through the model with the training data set. The rule-base parameters are modified and improved after each time step; through this process the algorithm will learn to map the third data pair but does not forget how to map the first two data pairs.

Now we demonstrate how the GM updates the rule-base parameters b_i, c_j^i, and σ_j^i using the first time step, $k=1$. Note that the first time step uses the first data-tuple of the training set, the second time step uses the second data-tuple, and the third time step uses the third data-tuple. We could repeatedly cycle through the training data set to improve the predictability of the algorithm, which is what the GM does. We start by calculating the difference between the predicted fuzzy output and the actual output for the first data-tuple of the training data set, using Equation (7.15). We then use this value to update the parameters of our rule-base using Equations (7.16) to (7.18).

$$\begin{aligned}\varepsilon_m(k=0) &= f(x^m|\theta(k=0)) - y^m.\\ \varepsilon_1(0) &= 1.0719447 - 1 = 0.0719447.\end{aligned} \tag{7.15}$$

We begin with the output membership function centers using

$$b_i(k) = b_i(k-1) - \lambda_1(\varepsilon_k(k-1))\frac{\mu_i(x^k, k-1)}{\sum_{i=1}^{R}\mu_i(x^k, k-1)}. \tag{7.16}$$

$$\begin{aligned}b_1(1) &= b_1(0) - \lambda_1 * (\varepsilon_1(0))\frac{\mu_1(x^1,0)}{\mu_1(x^1,0)+\mu_2(x^1,0)},\\ &= 1 - 1*(0.0719447)\left(\frac{1}{1+0.0183156}\right) = 0.9293493.\\ b_2(1) &= b_2(0) - \lambda_2 * (\varepsilon_1(0))\frac{\mu_2(x^1,0)}{\mu_1(x^1,0)+\mu_2(x^1,0)},\\ &= 5 - 1*(0.0719447)\left(\frac{0.0183156}{1+.0183156}\right) = 4.998706.\end{aligned}$$

Then the input membership function centers for the rule-base are updated based on the first time step $k=1$:

$$\begin{aligned}c_j^i(k) = c_j^i(k-1) - \lambda_2\varepsilon_k(k-1)&\left(\frac{b_i(k-1) - f(x^k|\theta(k-1))}{\sum_{i=1}^{R}\mu_i(x^k, k-1)}\right)\\ &*\mu_i(x^k, k-1)\left(\frac{x_j^k - c_j^i(k-1)}{\left(\sigma_j^i(k-1)\right)^2}\right).\end{aligned} \tag{7.17}$$

$$c_1^1(1) = c_1^1(0) - 1 * \varepsilon_1(0)\left(\frac{b_1(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_1(x^1,0)\left(\frac{x_1^1 - c_1^1(0)}{\left(\sigma_1^1(0)\right)^2}\right),$$

$$= 0 - 1 * (0.0719447) * \left(\frac{1 - 1.0719447}{1 + 0.0183156}\right) * 1 * \left(\frac{0 - 0}{(1)^2}\right) = 0.$$

$$c_2^1(1) = c_2^1(0) - 1 * \varepsilon_1(0)\left(\frac{b_1(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_2(x^1,0)\left(\frac{x_2^1 - c_2^1(0)}{\left(\sigma_2^1(0)\right)^2}\right),$$

$$= 2 - 1 * (0.0719447) * \left(\frac{1 - 1.0719447}{1 + 0.0183156}\right) * 0.0183156 * \left(\frac{2 - 2}{(1)^2}\right) = 2.$$

Because the input membership functions for the first rule are the first data-tuples, the updated centers do not change. The time step will affect the second rule because the rule's parameters are based on the second data-tuple of the training data set:

$$c_1^2(1) = c_1^2(0) - 1 * \varepsilon_1(0)\left(\frac{b_2(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_2(x^1,0)\left(\frac{x_1^1 - c_1^2(0)}{\left(\sigma_1^2(0)\right)^2}\right),$$

$$= 2 - 1 * (0.0719447) * \left(\frac{5 - 1.0719447}{1 + 0.0183156}\right) * 0.0183156 * \left(\frac{0 - 2}{(1)^2}\right) = 2.010166.$$

$$c_2^2(1) = c_2^2(0) - 1 * \varepsilon_1(0)\left(\frac{b_2(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_2(x^1,0)\left(\frac{x_2^1 - c_2^2(0)}{\left(\sigma_2^2(0)\right)^2}\right),$$

$$= 4 - 1 * (0.0719447) * \left(\frac{5 - 1.0719447}{1 + 0.0183156}\right) * 0.0183156 * \left(\frac{2 - 4}{(1)^2}\right) = 4.010166.$$

As expected, the first time step would have an effect on the input membership functions for the second rule. This is an iterative process and may take several iterations (time steps) to obtain a desired fuzzy system model.

Finally, we update the input membership function spreads, using the following equation:

$$\sigma_j^i(k) = \sigma_j^i(k-1) - \lambda_3 * \varepsilon_k(k-1) * \left\{\frac{b_i(k-1) - f\left(x^k|\theta(k-1)\right)}{\sum_{i=1}^{R} \mu_i\left(x^k, k-1\right)}\right\}$$

$$* \mu_i\left(x^k, k-1\right) * \left(\frac{\left(x_j^k - c_j^i(k-1)\right)^2}{\left(\sigma_j^i(k-1)\right)^3}\right). \tag{7.18}$$

$$\sigma_1^1(1) = \sigma_1^1(0) - 1 * \varepsilon_1(0)\left(\frac{b_1(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_1(x^1,0) * \left(\frac{\left(x_1^1 - c_1^1(0)\right)^2}{\left(\sigma_1^1(0)\right)^3}\right),$$

$$= 1 - 1 * (0.0719447)\left(\frac{1 - 1.0719447}{1.0183156}\right) * 1 * \left(\frac{(0-0)^2}{(1)^3}\right) = 1.$$

$$\sigma_2^1(1) = \sigma_2^1(0) - 1 * (0.0719447)\left(\frac{b_1(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_1(x^1,0) * \left(\frac{\left(x_2^1 - c_2^1(0)\right)^2}{\left(\sigma_2^1(0)\right)^3}\right),$$

$$= 1 - 1 * (0.0719447)\left(\frac{1 - 1.0719447}{1.0183156}\right) * 1 * \left(\frac{(2-2)^2}{(1)^3}\right) = 1.$$

$$\sigma_1^2(1) = \sigma_1^2(0) - 1 * (0.0719447)\left(\frac{b_2(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_2(x^1,0) * \left(\frac{\left(x_1^1 - c_1^2(0)\right)^2}{\left(\sigma_1^2(0)\right)^3}\right),$$

$$= 1 - 1 * (0.0719447)\left(\frac{5 - 1.0719447}{1.0183156}\right) * 0.0183156 * \left(\frac{(0-2)^2}{(1)^3}\right) = 0.979668.$$

$$\sigma_2^2(1) = \sigma_2^2(0) - 1 * (0.0719447)\left(\frac{b_2(0) - f(x^1|\theta(0))}{\mu_1(x^1,0) + \mu_2(x^1,0)}\right) * \mu_2(x^1,0) * \left(\frac{\left(x_2^1 - c_2^2(0)\right)^2}{\left(\sigma_2^2(0)\right)^3}\right),$$

$$= 1 - 1 * (0.0719447)\left(\frac{5 - 1.0719447}{1.0183156}\right) * 0.0183156 * \left(\frac{(2-4)^2}{(1)^3}\right) = 0.979668.$$

We now have completed one iteration (using one time step), which updated the parameters of the rule-base. Further iterations with the training set will improve the predictive power of the rule-base.

Clustering Method

Fuzzy clustering is the partitioning of a collection of data into fuzzy subsets or clusters based on similarities between the data (Passino and Yurkovich, 1998). The CM, like the other methods described previously, develops a fuzzy estimation model to predict the output given the input. The algorithm forms rules (or clusters) with training data using a nearest neighbor approach for the fuzzy system. This is demonstrated in the following example where the same training data set used in the previous examples is again used here (see Table 7.1).

Recall that these data consist of two inputs ($n = 2$) and one output for each data-tuple. Again we employ Gaussian membership functions for the input fuzzy sets, and delta functions for the output functions. In addition, we make use of center-average defuzzification and product premise for developing our fuzzy model, which is given by $f(x|\theta)$ in Equation (7.3); however, for the CM we employ slightly different variables as shown:

$$f(x|\theta) = \frac{\sum_{i=1}^{R} A_i \prod_{j=1}^{n} \exp\left[-\left(\frac{x_j - v_j^i}{2\sigma}\right)^2\right]}{\sum_{i=1}^{R} B_i \prod_{j=1}^{n} \exp\left[-\left(\frac{x_j - v_j^i}{2\sigma}\right)^2\right]}. \tag{7.19}$$

In the preceding equation, R is the total number of rules, v_j^i are the input membership function centers, x_j is the input, and σ is spread for the input membership functions.

In this example we initiate the parameters A_i and B_i which are then updated or optimized by the algorithm during training of the fuzzy model to predict the output. This is clarified later on in this section. Passino and Yurkovich (1998) make the following recommendations on σ:

- A small σ provides narrow membership functions that may yield a less smooth fuzzy system, which may cause the fuzzy system mapping not to generalize well for the data points in the training set.
- Increasing the parameter σ will result in a smoother fuzzy system mapping.

The use of only one value for the spread may pose a problem when developing a model for the inputs of different ranges. For instance, suppose it is desired to use five input membership function centers for each input variable. Presume that the first input has a range of 0–10, that the second input has a range of 20–200, and that the resulting width of the membership functions for the first input would have to be much smaller than that of the second. The use of only one spread to develop a model for systems such as this would not work well. This could be remedied by increasing the number of input membership function centers for the second input.

These parameters make up the vector θ, shown below, which is developed during the training of the fuzzy model $f(x|\theta)$. The dimensions of θ are determined by the number of inputs n and the number of rules, R, in the rule-base:

$$\theta = \left[A_1, \ldots, A_R, B_1, \ldots, B_R, v_1^1, \ldots, v_n^1, \ldots, v_1^R, \ldots, v_n^R, \sigma\right]^{\mathrm{T}}.$$

The CM develops its rules by first forming cluster centers $v^j = [v_1^j, v_2^j, \ldots, v_n^j]^{\mathrm{T}}$ for the input data. These cluster centers are formed by measuring the distance between the existing R cluster centers and the input training data; if the distance is greater than a specified maximum ε_f we add another cluster center (i.e., rule), otherwise the available cluster center will suffice and we update the parameters A and B for that rule.

Let us begin the training of the fuzzy model by specifying the parameters using the first data-tuple in Table 7.1 to initiate the process. For this we use $A_1 = y^1 = 1$, $B_1 = 1$, $v_1^1 = x_1^1 = 0$, $v_2^1 = x_2^1 = 2$, and $\sigma = 0.3$. We also specify the maximum distance between our cluster centers and the input data as $\varepsilon_f = 3.0$. Our fuzzy model $f(x|\theta)$ now has one rule ($R = 1$) and the cluster center for the input of this first rule is

$$\underline{v}^1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Now using the second data-tuple

$$(x^2, y^2) = \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, 5 \right),$$

we check if the clusters are applicable for this data-tuple or if another rule is needed. We accomplish this by measuring the distance ε_{ij} for the ith data-tuple and the jth rule center and compare this value to the specified ε_f:

$$\begin{aligned}
\varepsilon_{ij} &= |x_j^i - \underline{v}^1|. \\
\varepsilon_{21} &= |x_1^2 - v_1^1|, \\
&= |2-0| = 2 < \varepsilon_f = 3. \\
\varepsilon_{22} &= |x_2^2 - v_2^1|, \\
&= |4-2| = 2 < \varepsilon_f = 3.
\end{aligned} \tag{7.20}$$

Thus there is no need to incorporate an additional cluster, as both ε_{11} and ε_{12} are less than ε_f. However, we must update the existing parameters A_1 and B_1 to account for the output of the current input data-tuple. Updating A_1 modifies the numerator to better predict the output value for the current input data-tuple while the updated B_1 normalizes this predicted output value. This is accomplished using the following equations:

$$A_i = A_{i-1}^{\text{old}} + y^i. \tag{7.21}$$

$$B_i = B_{i-1}^{\text{old}} + 1. \tag{7.22}$$

$$\begin{aligned}
A_2 &= A_1^{\text{old}} + y^2 = 1 + 5 = 6. \\
B_2 &= B_1^{\text{old}} + 1 = 1 + 1 = 2.
\end{aligned}$$

This results in the following parameters for our fuzzy model:

$$\theta = \left[A_2 = 6, B_2 = 2, v_1^1 = 0, v_2^1 = 2, \sigma = 0.3\right]^{\mathrm{T}}.$$

Let us continue training our fuzzy model using the third data-tuple,

$$\begin{aligned}
(x^3, y^3) &= \left(\begin{bmatrix} 3 \\ 6 \end{bmatrix}, 6 \right). \\
\varepsilon_{31} &= |x_1^3 - v_1^1| = |3-0| = 3 = \varepsilon_f. \\
\varepsilon_{32} &= |x_2^3 - v_2^1| = |6-2| = 4 > \varepsilon_f = 3, \\
\varepsilon_{32} &= 6 > \varepsilon_f = 3.
\end{aligned}$$

Because both the calculated distances are not less than the specified maximum distance ($\varepsilon_f = 3$), we include another two additional clusters for the next rule. We now have two rules, the second of which is based on the third data-tuple. The cluster centers for the second rule are assigned the values equivalent to the input of the third data-tuple, and parameter A_2 is assigned a value equivalent to the output of the third data-tuple while B_2 is assigned a value of 1. The parameters needed for the fuzzy model are displayed below in the updated parameter vector θ:

$$\theta = \left[A_1 = 6, A_2 = 6, B_1 = 2, B_2 = 1, v_1^1 = 0, v_2^1 = 2, v_1^2 = 3, v_2^2 = 6, \sigma = 0.3\right]^{\mathrm{T}}.$$

Learning from Examples

The LFE training procedure relies entirely on a complete specification of the membership functions by the analyst as it only constructs the rules. Again we use the data set Z illustrated in Table 7.1 as the training data set for the fuzzy model. Like the other examples, we initiate the algorithm by designating two rules for this data set, $R = 2$. For this method, we use triangular membership functions rather than Gaussian for both the input and output. This is done to demonstrate the use of other types of membership functions.

We define the expected range of variation in the input and output variables:

$$X_i = \left[x_i^-, x_i^+\right], i = 1, 2 \quad Y = \left[y^-, y^+\right].$$

We designate $x_1^- = 0$, $x_1^+ = 4$, $x_2^- = 0$, $x_2^+ = 8$, $y^- = -1$, and $y^+ = 9$ as a choice for known regions within which all data points lie. Recall that a triangular membership function was defined mathematically in Equation (7.2), which is rewritten here as

$$\mu^c(u) = \begin{cases} \max\left\{0, 1 + \dfrac{u - c}{0.5w}\right\}, & \text{if } u \le c \\ \max\left\{0, 1 + \dfrac{c - u}{0.5w}\right\}, & \text{if } u > c. \end{cases}$$

In the preceding equation, u is the point of interest, c is the membership function center, and w is the base width of the membership function (see Figure 7.3).

Next, the membership functions are defined for each input and output universe of discourse. Our system has two inputs and one output, thus we have X_1^j, X_2^k, and Y^l. It is important to recognize that the number of membership functions on each universe of discourse affects the accuracy of the function approximations and in this example X_1 has fewer membership functions than does X_2. This may occur if the first input is limited to a few values and the second input has more variability. Here, X_1^j, X_2^k, and Y^l denote the fuzzy sets with associated membership functions $\mu_{X_i^j}(x_i)$ and $\mu_{Y^j}(y)$, respectively. These fuzzy sets and their membership functions are illustrated in Figures 7.8–7.10.

In both Figures 7.8 and 7.9, the membership functions are saturated (membership value for the fuzzy set is equal to 1) at the far left and right extremes. In this case we use Equations (7.23) and (7.24) to determine the membership values for the leftmost and rightmost membership functions, respectively:

$$\mu^{\mathrm{L}}(u)=\begin{cases}1, & \text{if } u \le c^{\mathrm{L}}\\ \max\left\{0, 1+\dfrac{c^{\mathrm{L}}-u}{0.5w^{\mathrm{L}}}\right\}, & \text{otherwise.}\end{cases} \tag{7.23}$$

$$\mu^{\mathrm{R}}(u)=\begin{cases}\max\left\{0, 1+\dfrac{u-c^{\mathrm{R}}}{0.5w^{\mathrm{R}}}\right\}, & \text{if } u \le c^{\mathrm{R}}\\ 1, & \text{otherwise.}\end{cases} \tag{7.24}$$

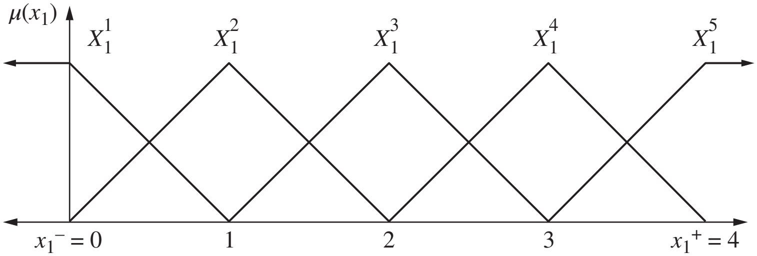

Figure 7.8 Specified triangular input membership functions for x_1.

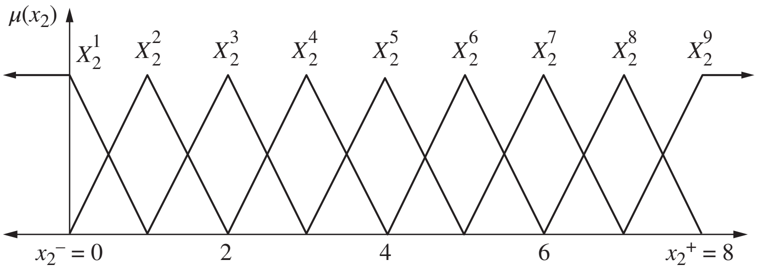

Figure 7.9 Specified triangular input membership functions for x_2.

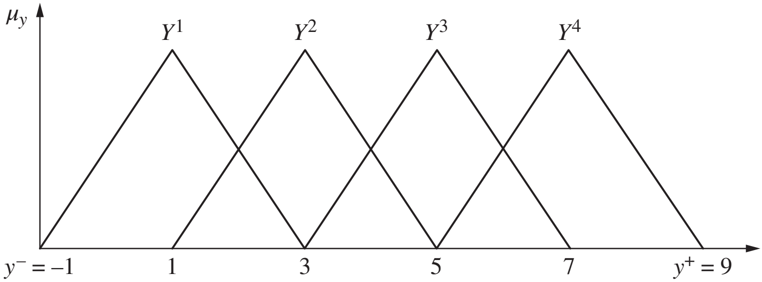

Figure 7.10 Output membership functions for y.

In these equations, c^{L} and c^{R} specify the point on the horizontal axis when full membership begins for the leftmost and rightmost membership functions (μ^{L} and μ^{R}), respectively, whereas w^{L} and w^{R} specify two times the width of the nonunity and nonzero part of μ^{L} and μ^{R}, respectively.

Now for the first training data (d) point for the fuzzy system we have

$$d^{(1)} = \left(x^1, y^1\right) = \left(\begin{bmatrix} 0 \\ 2 \end{bmatrix}, 1\right),$$

which we add to the rule-base since there are currently no other rules in the set:

$$\mathrm{R}_1 := \text{if } x_1 \text{ is } X_1^1 \text{ and } x_2 \text{ is } X_2^3 \text{ then } y \text{ is } Y^1.$$

Using Equation (7.2), (7.23), or (7.24), we can determine the membership values that the first training data point has in the two existing rules. These values are then used to calculate the "degree of attainment" using Equation (7.25). The data-tuple in $d^{(1)}$ has full membership in the leftmost membership of Figure 7.8 (fuzzy value 0 or X_1^1), full membership in the third membership function from the left in Figure 7.9 (fuzzy value 2 or X_2^3), and membership equal to 1 in the output membership function for the determined rule-base:

$$\begin{aligned} \mu_{X_1^1}(x_1=0) &= 1, \\ \mu_{X_2^3}(x_2=2) &= 1, \\ \mu_{Y^1}(y=1) &= 1. \\ \text{degree}(\mathrm{R}_i) &= \mu_{X_1^j}(x_1) * \mu_{X_2^k}(x_2) * \mu_{Y^l}(y). \end{aligned} \tag{7.25}$$

Then,

$$\text{degree}(\mathrm{R}_1) = \mu_{X_1^1}(x_1) * \mu_{X_2^3}(x_2) * \mu_{Y^1}(y) = 1*1*1 = 1.$$

We now move on to the next data-tuple, $d^{(2)}$:

$$d^{(2)} = \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, 5\right).$$

Because the existing rule does not model $d^{(2)}$ we add another rule to the rule-base, which is as follows:

$$\mathrm{R}_2 := \text{if } x_1 \text{ is } X_1^3 \text{ and } x_2 \text{ is } X_2^5 \text{ then } y \text{ is } Y^3.$$

The data-tuple in $d^{(2)}$ has full membership in the center membership function of Figure 7.8 (fuzzy value 2 or X_1^3), full membership in the fifth membership function from the left in Figure 7.9 (fuzzy value 4 or X_2^5), and a membership value of 1 in the output membership function:

$$\mu_{X_1^3}(x_1 = 2) = 1$$
$$\mu_{X_2^5}(x_2 = 4) = 1$$
$$\mu_{Y^3}(y = 5) = 1$$

where we get

$$\text{degree}(\mathrm{R}_1) = \mu_{X_1^1}(x_1) * \mu_{X_2^3}(x_2) * \mu_{Y^1}(y) = 0*0*0 = 0$$
$$\text{degree}(\mathrm{R}_2) = \mu_{X_1^3}(x_1) * \mu_{X_2^5}(x_2) * \mu_{Y^3}(y) = 1*1*1 = 1$$

Of the rules specified, the second training data-tuple has zero membership in Rule 1, R_1, and full membership in Rule 2, R_2. Thus, it was beneficial to add Rule 2 to the rule-base because it represents $d^{(2)}$ better than any other existing rule.

Now we move on to the third data-tuple in the training data set:

$$d^{(3)} = \left(\begin{bmatrix} 3 \\ 6 \end{bmatrix}, 6 \right)$$

Again the existing rules in the rule-base do not model $d^{(3)}$, thus another rule should be added. Because the data set is so small we find that we are including an additional rule for each data-tuple in the data set. Ideally we do not prefer that a rule be specified for every data-tuple in the training set—although, previously, it was mentioned that Z is a representative portion of a larger data set—so in this example there would be more data points in the entire set. If we were to train a fuzzy system with a much larger data set Z, we would find that there will not be a rule for each of the m data-tuples in Z. Some rules will adequately represent more than one data pair. Nevertheless, the system can be improved by designating the fuzzy sets X_1^j, X_2^k, and Y^l differently and perhaps avoiding the addition of unnecessary rules to the rule-base. We do not attempt this correction here but continue with the addition of another rule to the rule-base:

$$\mathrm{R}_3 = \text{if } x_1 \text{ is } X_1^4 \text{ and } x_2 \text{ is } X_2^7 \text{ then } y \text{ is } Y^3.$$
$$\text{degree}(\mathrm{R}_1) = \mu_{X_1^1}(x_1) * \mu_{X_2^3}(x_2) * \mu_{Y^1}(y) = 0*0*0 = 0,$$
$$\text{degree}(\mathrm{R}_2) = \mu_{X_1^3}(x_1) * \mu_{X_2^5}(x_2) * \mu_{Y^3}(y) = 0*0*0.5 = 0,$$
$$\text{degree}(\mathrm{R}_3) = \mu_{X_1^4}(x_1) * \mu_{X_2^7}(x_2) * \mu_{Y^3}(y) = 1*1*0.5 = 0.5,$$

The third data-tuple has 0.5 membership in Rule 3 but has zero membership in the other rules. We now have a complete rule-base and, thus, a fuzzy model for the system. LFE is a useful algorithm for developing a fuzzy model because it can be used as a basis in other algorithms for developing a stronger model. For instance, it can be used to form a rule-base which can then be improved in the RLS algorithm to refine the fuzzy model.

Modified Learning from Examples

Unlike the LFE algorithm which relies entirely on user-specified membership functions, the MLFE calculates both membership functions and rules.

In this example, we again employ delta functions for the outputs and Gaussian membership functions for the inputs. Again, b_i is the output value for the ith rule, c_j^i is the point in the jth input universe of discourse where the membership function for the ith rule achieves a maximum, and $\sigma_j^i(\sigma_j^i > 0)$ is the spread of the membership function for the jth input and the ith rule. Recall that θ is a vector composed of the above parameters, where the dimensions of θ are determined by the number of inputs n and the number of rules R in the rule-base.

Again, we use the data in Table 7.1 to develop a fuzzy model. Let us begin by specifying a "initial fuzzy system" that the MLFE procedure will use to initialize the parameters in θ. We initiate the process by setting the number of rules equal to 1, $R = 1$, and for b_1, c_1^1, and c_2^1 we use the first training data-tuple in Z and the spreads will be assumed to be equal to 0.5. It is important to note that the spreads cannot be set at zero to avoid a division by zero error in the algorithm.

$$b_1 = y^1 = 1,\ c_1^1 = 0,\ c_2^1 = 2,\ \sigma_1^1 = \sigma_2^1 = 0.5.$$

For this example, we would like the fuzzy system to approximate the output to within a tolerance of 0.25, thus we set $\varepsilon_f = 0.25$. We also introduce a weighting factor W which is used to calculate the spreads for the membership functions, as given later in Equation (7.28). The weighting factor W is used to determine the amount of overlap between the membership function of the new rule and that of its nearest neighbor. This is demonstrated later in this section. For this example, we will set the value of W equal to 2.

Following the initial procedure for the first data-tuple, we use the second data-tuple,

$$(x^2, y^2) = \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, 5 \right),$$

and compare the data-tuple output portion y^2 with the existing fuzzy system output value, $f(x^2|\theta)$. The existing fuzzy system contains only the one rule which was previously added to the rule-base:

$$f(x^2|\theta) = \frac{1 * \exp\left[-\frac{1}{2}\left(\frac{2-0}{0.5}\right)^2 \right] * \exp\left[-\frac{1}{2}\left(\frac{4-2}{0.5}\right)^2 \right]}{\exp\left[-\frac{1}{2}\left(\frac{2-0}{0.5}\right)^2 \right] * \exp\left[-\frac{1}{2}\left(\frac{4-2}{0.5}\right)^2 \right]} = 1.$$

Next, we determine how accurate or adequate our fuzzy system is at mapping the information. To do this, we take the absolute value of Equation (7.15), where m denotes the data-tuple $m = 2$:

$$\left| f(x^2|\theta) - y^2 \right| = |1 - 5| = 4.$$

This value is much greater than the tolerance specified, $\varepsilon_f = 0.25$. Thus, we add a rule to the rule-base to represent (x^2, y^2) by modifying the current parameters θ by letting $R = 2$, $b_2 = y^2$, and $c_j^2 = x_j^2$ for $j = 1, 2$:

$$b_2 = 5,\ c_1^2 = 2,\ c_2^2 = 4.$$

If the calculated ε_f had been less than the specified tolerance, there would have been no need to add another rule.

Next, we develop the spreads σ_j^i to adjust the spacing between membership functions for the new rule. MLFE does this in a manner that does not distort what has already been learned, so there is a smooth interpolation between training points. The development of σ_j^i for $i = R$ is accomplished through a series of steps, the first step being the calculation of the distances between the new membership function center and the existing membership function centers for each input,

$$\left| c_j^{i'} - c_j^i \right|. \tag{7.26}$$

These values are then placed in the vector $\overline{h}_j$ below:

$$\overline{h}_j^{i'} = \left\{ \left| c_j^{i'} - c_j^i \right| : i' = 1.2, \ldots, R,\ i' \neq i \right\}.$$

For each input j the smallest nonzero element of this vector of distances $\overline{h}_j^{i'}$ is the nearest neighbor to the newly developed input membership function center for any added rule. This nearest neighbor is then used to create the relative width of the membership function for the new rule. If all the distances are zero in this vector, which means all the membership functions defined before are at the same center value, do not modify the relative width of the new rule but keep the same width defined as for the other input membership functions. The determination of this minimum is accomplished using

$$k_j = \min\left(\overline{h}_j\right), \tag{7.27}$$

where $j = 1, 2, \ldots, n$ and c_j^i are fixed.

For instance, in our example we have the following:

$$\begin{aligned}
i' = 2 \text{ and } j = 1 \quad & \left| c_j^{i'} - c_j^i \right| : i = 1 \\
& \left| c_1^2 - c_1^1 \right| = |2 - 0| = 2, \\
i' = 2 \text{ and } j = 2 \quad & \left| c_j^{i'} - c_j^i \right| : i = 1 \\
& \left| c_2^2 - c_2^1 \right| = |4 - 2| = 2,
\end{aligned}$$

and our vector of distances for each input consists of only one nonzero value:

$$\overline{h}_1^2 = \{2\} \quad \overline{h}_2^2 = \{2\}.$$

Therefore, c_1^1 is closest to c_1^2 and is used to develop the relative width of the membership function for the first input of the new rule and c_2^1 is the nearest neighbor to c_2^2. The weighting factor W is used to determine the amount of overlap between the membership function of the new rule with that of its nearest neighbor. This is accomplished using

$$\sigma_j^{i'} = \frac{1}{W}\left|c_j^{i'} - c_j^{\min}\right|, \tag{7.28}$$

where $c_j^{\min}$ is the nearest membership function center to the new membership function center $c_j^{i'}$.
As mentioned, we select a value of $W=2$ which results in the following relative widths for the membership functions of the new rule:

$$\sigma_1^2 = \frac{1}{W}\left|c_1^2 - c_1^1\right| = \frac{1}{2}|2-0| = 1.$$
$$\sigma_2^2 = \frac{1}{W}\left|c_2^2 - c_2^1\right| = \frac{1}{2}|4-2| = 1.$$

The spread (relative width of the membership function) could be altered by increasing or decreasing W, which results in less or more, respectively, overlapping of the input membership functions.
We now have a rule-base consisting of two rules and θ consists of

$$\theta = \left[b_1, b_2, c_1^1, c_2^1, c_1^2, c_2^2, \sigma_1^1, \sigma_2^1, \sigma_1^2, \sigma_2^2\right]^{\mathrm{T}},$$

where

$$\begin{array}{ll} b_1 = 1 & b_2 = 5, \\ c_1^1 = 0 & c_1^2 = 2, \\ c_2^1 = 2 & c_2^2 = 4, \\ \sigma_1^1 = 0.5 & \sigma_1^2 = 1, \\ \sigma_2^1 = 0.5 & \sigma_2^2 = 1, \end{array}$$

We now move on to the third data-tuple in the training set to determine if our rule-base is adequate and if we have to include any additional rules:

$$f\left(x^3|\theta\right) = \frac{1*\exp\left[-\frac{1}{2}\left(\frac{3-0}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{6-2}{1}\right)^2\right]+5*\exp\left[-\frac{1}{2}\left(\frac{3-2}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{6-4}{1}\right)^2\right]}{\exp\left[-\frac{1}{2}\left(\frac{3-0}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{6-2}{1}\right)^2\right]+\exp\left[-\frac{1}{2}\left(\frac{3-2}{1}\right)^2\right]*\exp\left[-\frac{1}{2}\left(\frac{6-4}{1}\right)^2\right]},$$

$$= \frac{0.410428}{0.08208872} = 5.$$

Again, we determine how accurate or adequate our fuzzy system is at mapping the information:

$$\left|f\left(x^3\middle|\theta\right)-y^3\right| = |5-6| = 1 > \varepsilon_f = 0.25.$$

According to our specified tolerance, another rule is necessary to adequately model our training data:

$$b_3 = 6 \quad c_1^3 = 3 \quad c_2^3 = 6.$$

We now can determine the spreads for the third rule. We begin with Equation (7.26) to form the vector of distances:

$$\begin{aligned}
i' = 3 \text{ and } j = 1 & \left\{\left|c_j^{i'} - c_j^i\right| : i = 1,2\right\}, \\
& \left\{\left|c_1^3 - c_1^1\right|,\ \left|c_1^3 - c_1^2\right|\right\}, \\
& \{3, 1\}. \\
i' = 3 \text{ and } j = 2 & \left\{\left|c_j^{i'} - c_j^i\right| : i = 1,2\right\}, \\
& \left\{\left|c_2^3 - c_2^1\right|,\ \left|c_2^3 - c_2^2\right|\right\}, \\
& \{4.2\}.
\end{aligned}$$

Figure 7.11 shows the graphical meaning of the distances $\overline{h}_1^3$ and $\overline{h}_2^3$ from each respective rule centers. In the distance vector $\overline{h}_1^3$, the minimum nonzero value is 1 and in the distance vector $\overline{h}_2^3$, the minimum nonzero value is 2. Therefore, the nearest neighbor to the first input membership function center is c_1^2, and the nearest neighbor to the second input membership function center is c_2^2, which are both used with the weighting factor and Equation (7.28) to calculate the spreads for each membership function:

$$\sigma_1^3 = \frac{1}{W}\left|c_1^3 - c_1^2\right| = \frac{1}{2}|3-2| = \frac{1}{2}.$$

$$\sigma_2^3 = \frac{1}{W}\left|c_2^3 - c_2^2\right| = \frac{1}{2}|6-4| = 1.$$

The resulting rule-base is the following:

$$\theta = \left[b_1, b_2, b_3, c_1^1, c_2^1, c_1^2, c_2^2, c_1^3, c_2^3, \sigma_1^1, \sigma_2^1, \sigma_1^2, \sigma_2^2, \sigma_1^3, \sigma_2^3\right]^{\mathrm{T}},$$

where

$$\begin{aligned}
&b_1 = 1 \quad b_2 = 5 \quad b_3 = 6, \\
&c_1^1 = 0 \quad c_1^2 = 2 \quad c_1^3 = 3, \\
&c_2^1 = 2 \quad c_2^2 = 4 \quad c_2^3 = 6, \\
&\sigma_1^1 = 0.5 \quad \sigma_1^2 = 1 \quad \sigma_1^3 = 0.5, \\
&\sigma_2^1 = 0.5 \quad \sigma_2^2 = 1 \quad \sigma_2^3 = 1,
\end{aligned}$$

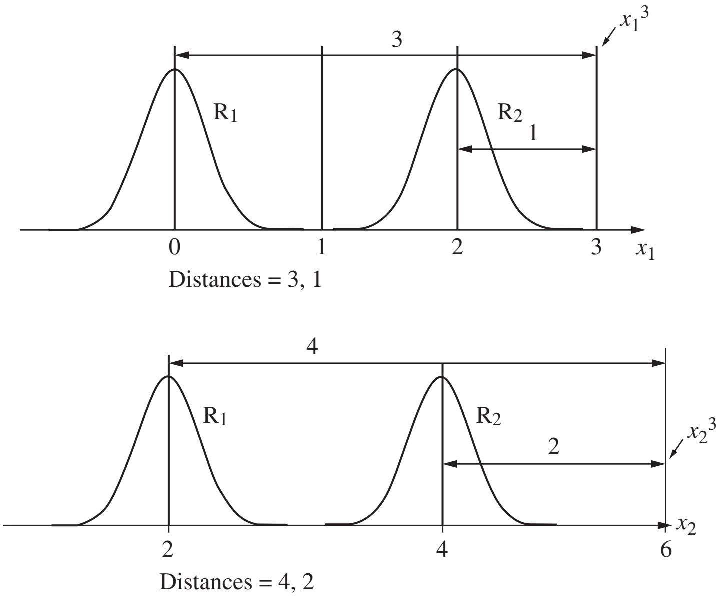

Figure 7.11 Distance measures $\bar{h}_1^3$ and $\bar{h}_2^3$.

Example 7.1

Fuzzy modeling has been employed to generate a rule-base to model a syntactic foam from only limited input–output data obtained through various compressive tests. Owing to their light weight and high compressive strength, foams have been incorporated into the design of many engineering systems. This is especially true in the aircraft industry where these characteristics are extremely important. For instance, aluminum syntactic foam is theorized to have a shear strength that is three times greater than existing aluminum (Erikson, 1999). Syntactic foams are composite materials formed by mechanically combining manufactured material bubbles or microspheres with resin. They are referred to as *syntactic* because the microspheres are arranged together, unlike blown foams that are created by injecting gas into a liquid slurry causing the bubbles to solidify and producing foam (Erikson, 1999). As is often the case with newly developed materials, the cost of preparing the material is high; thus, only limited information is available on the material.

A newly developed syntactic foam has been selected as an encasing material because its light weight, high compressive strength, isotropic behavior, low porosity, and because the material is noncombustible and can be machined to a specific shape and tolerance. Owing to the high costs involved in preparing the syntactic foam, only a limited amount has been made. Fortunately, in addition to the two pieces of syntactic foam specimens prepared for use as the encasing material, four specimens were prepared for conducting triaxial compression tests. The four specimens each had a height of 2.800 ± 0.001 inches and a diameter of 1.400 ± 0.001 inches, as shown in Figure 7.12. The triaxial compression test is capable of applying two different compressive stresses to the sample, a radial (minor principal) and a longitudinal (major principal) stress. In each test performed, the compressive pressure (stress) was gradually increased

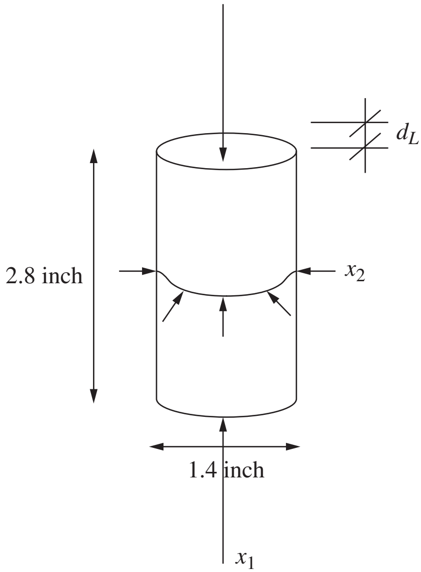

Figure 7.12 Syntactic foam cylinders.

causing the sample to deform. The first test applied a continuous equivalent major and minor principal compressive stress to the specimen (hydrostatic compression) and the yield stress was observed at about 10 000 psi. In the second test, the major and minor principal stresses were gradually increased to a value of 3750 psi; the minor principal stress was then held constant at this value, whereas the major principal stress was continuously increased. To probe a good portion of stress space, the same procedure was followed for the third and fourth tests; however, the minor principal stress was held constant at values of 6500 and 9000 psi, respectively, for these.

In each of the previous tests, the experimentalist found that maintaining the minor principal stress constant was difficult and the minor principal stress was noted to fluctuate by as much as ±200 psi. The experimentalist also noted that at about the yielding stress the syntactic foam began to harden and exhibit nonlinear behavior. A portion of the nonlinear data collected from the triaxial compression tests is provided in Table 7.2. Using the information provided by the experimentalist and the portion of the input–output data we develop a fuzzy model using the MLFE algorithm to obtain a general rule-base for governing the system, followed by the RLS algorithm to fine-tune the rule-base and the model parameters.

We begin by specifying a rule and its parameters that is used by the MLFE algorithm to develop the remainder of the rules in the rule-base. To do this, we use one of the data-tuples from the training set: say we use $x_1 = 12\,600$ psi and $x_2 = 9000$ psi as initial input membership function centers and $\delta_L = 2.8$ inch as the initial output (to simplify the accounting we use 2.8 rather than 0.028). We set the relative width of the first input membership function center at 250 based on the range between similar inputs. For instance, two similar input tuples are $x_1 = 12\,250$ psi and $x_2 = 3750$ psi, and $x_1 = 12\,100$ psi, and $x_2 = 3750$ psi. The relative width of the second input membership function is set at 150 based on the information provided by the experimentalist. For the weighting factor we choose a value of 2.1 and set the test factor equal to 0.45. MLFE uses this to develop the remainder of the membership functions

Table 7.2 Major and minor principal stress and resulting longitudinal deformation.

Training set			Testing set		
Major (x_1 psi)	Minor (x_2 psi)	δ_L (inch)	Major (x_1 psi)	Minor (x_2 psi)	δ_L (inch)
12 250	3750	3.92176E–2	12 911.1	12 927	2.0273E–2
11 500	6500	2.90297E–2	11 092.4	10 966.4	1.59737E–2
11 250	9000	2.51901E–2	14 545.8	14 487.1	2.40827E–2
11 000	11 000	1.63300E–2	13 012.1	12 963.8	2.02157E–2
11 960	9000	2.50463E–2	12 150	3750	3.8150E–2
12 140	6510	3.22360E–2	12 904	3744.8	4.28953E–2
13 000	3750	4.37127E–2	14 000	3770.4	5.05537E–2
13 800	3750	4.91016E–2	11 406	6520.3	2.83967E–2
12 950	6500	3.60650E–2	12 100	6535.5	3.2120E–2
12 600	9000	2.80437E–2	13 109	6525.3	3.69773E–2
11 170	11 110	1.65160E–2	11 017.6	9000.7	2.11967E–2
12 930	13 000	2.02390E–2	12 105.9	8975.1	2.57443E–2
13 000	12 500	2.36600E–2	12 700	900.0	2.8504E–2
11 130	9000	2.19333E–2			
11 250	6500	2.77190E–2			
12 000	3750	3.88243E–2			
12 900	3750	4.28953E–2			
11 990	6500	3.15040E–2			
12 010	9000	2.58113E–2			
12 000	12 000	1.95610E–2			
14 490	14 450	2.41490E–2			
12 900	9000	2.97153E–2			
13 220	6500	3.81633E–2			
14 000	3750	5.05537E–2			

(Figure 7.13a and b) and the rule-base shown in Table 7.3. Note that the actual output is moved two decimal places to the left because of the simplification to the output mentioned previously.

For example,

$$\text{If } \underset{\sim}{A}_1 \text{ and } \underset{\sim}{B}_1 \text{ then } f(x) = 2.80 \times 10^{-2}.$$
$$\text{If } \underset{\sim}{A}_2 \text{ and } \underset{\sim}{B}_2 \text{ then } f(x) = 3.9218 \times 10^{-2}.$$
$$\text{If } \underset{\sim}{A}_3 \text{ and } \underset{\sim}{B}_3 \text{ then } f(x) = 2.903 \times 10^{-2}.$$

We now have the information needed to employ the RLS algorithm to improve our fuzzy model.

Before we move on to specifying the rule parameters needed by the RLS algorithm we point out that the experimentalist, in this example, observed that the minor principal stress fluctuated by as much as 200 psi and the spread for this input developed by MLFE is as high as 1309 psi. This is a direct result of the weighting factor, which forces the adjacent membership functions

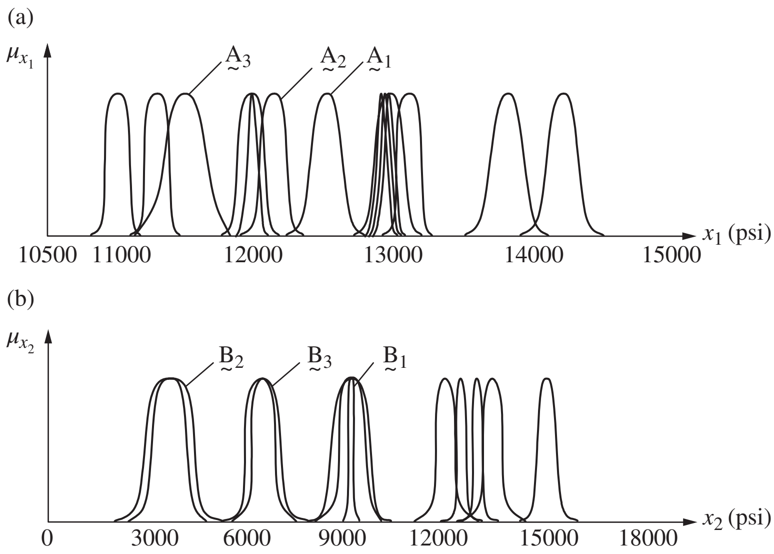

Figure 7.13 (a) Membership functions for major principal stress. (b) Membership functions for minor principal stress.

Table 7.3 Rule-base.

Input parameters				
c_1 (psi)	c_2 (psi)	σ_1	σ_2	Output parameter $f(x)$ (inch)
$\underset{\sim}{A}_1 = 12\,600$	$\underset{\sim}{B}_1 = 9000$	250	150	2.80E–2
$\underset{\sim}{A}_2 = 12\,250$	$\underset{\sim}{B}_2 = 3750$	166.7	2500	3.9218E–2
$\underset{\sim}{A}_3 = 11500$	$\underset{\sim}{B}_3 = 6500$	357.143	1190.476	2.9030E–2
11 250	9000	119.048	1190.476	2.2519E–2
11 000	11 000	119.048	952.381	1.6330E–2
11 960	9000	138.095	952.381	2.5046E–2
13 000	3750	190.476	1309.524	4.3712E–2
13 800	3750	380.952	1309.524	4.9102E–2
12 950	6500	23.810	1190.476	3.6065E–2
12 930	13 000	9.524	952.381	2.0239E–2
13 000	12 500	23.81	238.095	2.3660E–2
12 000	12 000	19.048	238.095	1.9561E–2
14 490	14 450	328.571	690.476	2.4149E–2
13 220	6500	104.762	1166.667	3.8163

to approach one another by increasing the spread. For instance, the spreads of the membership functions with $c_1 = 12\,250$ psi and $c_1 = 12\,000$ psi increase. Because the input membership functions are significantly apart from one another, a low value for the weighting factor forces the spread of the membership functions to be greater. Both of the inputs (x_1 and x_2) are based on

Table 7.4 Rule-base parameters for RLS algorithm.

Input parameters				
c_1 (psi)	c_2 (psi)	σ_1	σ_2	Regression parameter $\hat{\theta}$
12 600	9000	250	250	5
12 250	3750	166.7	250	5
11 500	6500	357.143	250	5
11 250	9000	119.048	250	5
11 000	11 000	119.048	250	5
11 960	9000	138.095	250	5
13 000	3750	190.476	250	5
13 800	3750	380.952	250	5
12 950	6500	23.810	250	5
12 930	13 000	9.524	250	5
13 000	12 500	23.81	250	5
12 000	12 000	19.048	250	5
14 490	14 450	328.571	250	5
13 220	6500	104.762	250	5

the one weighting factor and we can improve the spreads by designating one weighting factor for each input. For now, we simply modify the spread of the second input to about 250 psi and continue with the RLS algorithm.

The rule-base used by the RLS algorithm is shown in Table 7.4; additionally, we use the data-tuples of Table 7.2 as the training set, and the initial $\hat{\theta}$ (a vector, in this case consisting of 14 values for the 14 rules used) with values equal to 5. We use nonweighted least squares regression (so $\lambda = 1$), set $\alpha = 2000$, and cycle through the training set 100 times to develop a fuzzy model.

Once the fuzzy model has been developed we use the testing data of Table 7.2 to verify that the model is working properly. The resulting values included in vector $\hat{\theta}$ are shown in Table 7.5 and the testing set with its predicted output is shown in Table 7.6.

Recent examples of work illustrate the power of the MLFEs approach. MLFE was used to help in learning the Talbot effect, which describes light interference patterns. The proposed MLFE model proved necessary for efficient nanopatterning lithography in the work conducted by Su, Reda Taha, Christodoulou, and El-Kady (2008). Recent investigations by Harp, Reda Taha, and Ross (2009) suggested that robust models based on fuzzy LFEs can be produced if the widths of the membership functions are optimized. Genetic algorithms were suggested to produce the optimal rule-base. The proposed method showed excellent prediction of the bond strength in masonry.

Summary

This chapter has summarized six methods for use in developing fuzzy systems from input–output data. Of these six methods, the LFE, MLFE, and CMs can be used to develop fuzzy systems from such data. The remaining three methods, RLS, BLS, and the GMs, can be used

Table 7.5 Resulting values included in vector $\hat{\theta}$ from RLS algorithm.

$\hat{\theta}$	2.8884	3.9018	3.0122	2.2226	1.6423	2.5246	4.2712	4.9832	3.5991
$\hat{\theta}$	2.0232	2.4142	1.9561	2.4149	3.8163				

Table 7.6 Predicted output values for the testing set.

Testing set			
Major (x_1 psi)	Minor (x_2 psi)	δ_L (inch)	Predicted output δ_L (inch)
12 911.1	12 927	2.0273E–2	2.0231E–2
11 092.4	10 966.4	1.59737E–2	1.6423E–2
14 545.8	14 487.1	2.40827E–2	2.4149E–2
13 012.1	12 963.8	2.02157E–2	2.0820E–2
12 150	3750	3.8150E–2	3.9020E–2
12 904	3744.8	4.28953E–2	4.3184E–2
14 000	3770.4	5.05537E–2	4.9384E–2
11 406	6520.3	2.83967E–2	3.0122E–2
12 100	6535.5	3.2120E–2	3.0122E–2
13 109	6525.3	3.69773E–2	3.8163E–2
11 017.6	9000.7	2.11967E–2	2.2226E–2
12 105.9	8975.1	2.57443E–2	2.5963E–2
12 700	900.0	2.8504E–2	2.8884E–2

to take fuzzy systems that have been developed by the first group of methods and refine them (or *tune* them) with additional training data.

Fuzzy systems are useful in the formulation and quantification of human observations. These observations eventually are manifested in terms of input–output data; these data can take the form of numbers, images, audio records, and so on, but they all can be reduced to unit-based quantities on the real line. We might ask, "in what forms are the human observations manifested?" We could suggest that one form is linguistic, which would be a *conscious* form where humans could express directly their observations as knowledge expressed in the form of rules. Or, the form of knowledge could be *subconscious* in the sense that the observations could be measured or monitored by some device, but where the behavior cannot yet be reduced to rules by a human observer (an example of this would be the parallel parking of a car; humans can do it, but they might have difficulty describing what they do in a canonical rule form). The latter form of information, that from measurable observations, comprises the input–output data that are dealt with in this chapter. Once the rules of the system are derived from the input–output data, they can be combined with other rules, perhaps rules that come from a human, and together (or separately) they provide a contextual meaning to the underlying physics of the problem.

But whether the rules come directly from a human or from methods such as those illustrated in this chapter, the resulting simulations using fuzzy systems theory are the same. A unique property of a fuzzy system is that the rules derived from the observational data provide *knowledge* of the system being simulated; this knowledge is in the form of linguistic rules that are understandable to a human in terms of the underlying physical system behavior. In contrast to this, other model-free methods such as neural networks can also be used in simulation, but the information gleaned about the number of layers, number of neurons in each layer, path weights between neurons, and other features of a neural network reveals to the human almost nothing about the physical process being simulated. This is the power and utility of a fuzzy system model.

References

Erikson, R. (1999). Mechanical engineering. *ASME*, **121** (1), 58–62.

Harp, D., Reda Taha, M. M., and Ross, T. (2009). A robust genetic-fuzzy approach for modeling complex systems with two civil engineering applications. *ASCE: J. Comput. Civil Eng.*, **23** (3), 193–199.

Passino, K., and Yurkovich, S. (1998). *Fuzzy Control* (Menlo Park, CA: Addison Wesley Longman).

Su, M. F., Reda Taha, M. M., Christodoulou, C., and El-Kady, I. (2008). Fuzzy learning of talbot effect guides optimal mask design for proximity field nanopatterning lithography. *IEEE Photonics Technol. Lett.*, **20** (10), 761–763.

Problems

7.1 Previously in this chapter the recursive least squares (RLS) algorithm was demonstrated using training set Z of Table 7.1; however, only one cycle was performed using the input data-tuples. Perform an additional cycle with the input data-tuples to determine if the predicted output changes.

7.2 Previously in this chapter the recursive least squares (RLS) algorithm was demonstrated using training set Z of Table 7.1.

a. For the input membership function centers and spread given below, develop a fuzzy model using the RLS algorithm (perform two steps) using the training set Z given in the table shown.

$$\sigma_j^{\,i} = 2 \quad \text{and} \quad \hat{\theta}(0) = \begin{bmatrix} 0.3647 \\ \\ 8.1775 \end{bmatrix}$$

$$c_1^{\,1} = 1, \qquad c_1^{\,2} = 2.5$$

$$c_1^{\,1} = 2.5, \qquad c_2^{\,2} = 5$$

x_1	x_2	y
1	2	2
2	4	6
4	7	8

b. Modify the input membership function centers for the following values (spread and centers) and develop a fuzzy model for the training set Z of Table 7.1 using the RLS algorithm (perform two steps). Note that the remaining rule-base parameters are the same as those used in the text, for example,

$$\sigma_j^{\,i} = 2 \text{ and } \overline{\theta}(0) = \begin{bmatrix} 0.3647 \\ 8.1775 \end{bmatrix}$$

New centers:

$$c_1^{\,1} = 1, \quad c_1^{\,2} = 3$$
$$c_2^{\,1} = 3, \quad c_2^{\,2} = 6$$

7.3 Using the software provided on the publisher's website (see the preface), improve the output of the RLS model presented in Example 7.1 by modifying the input membership function parameters of Table 7.4 using the gradient method (GM).

7.4 Using the clustering method (CM) develop a fuzzy model for the input-output data presented in Table 7.2

7.5 Using the CM develop a fuzzy model for the input-output data presented in the table in a manner similar to Example 7.1.

c_1 (psi)	c_2 (psi)	σ_3	σ_4
7000	6500	175	175
6850	3200	135	175
6500	4500	121	175
6000	5000	85	175
6200	5000	83.5	175
6315	5800	87	175

x_1 (psi)	x_2 (psi)	δ_L (inch)
6800	6834	1.23 E^{-2}
6500	6492	1.04 E^{-2}
6900	6950	1.34 E^{-2}
6875	6700	1.31 E^{-2}

8

Fuzzy Systems Simulation

> *As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.*
>
> Lotfi Zadeh, professor, systems engineering, 1973

The real world is complex; complexity in the world generally arises from uncertainty in the form of ambiguity. Problems featuring complexity and ambiguity have been addressed subconsciously by humans because they could think; these ubiquitous features pervade most social, technical, and economic problems faced by the human race. Why then are computers, which have been designed by humans after all, not capable of addressing complex and ambiguous issues? How can humans reason about real systems, when the complete description of a real system often requires more detailed data than a human could ever hope to recognize simultaneously and assimilate with understanding? The answer is that humans have the capacity to reason approximately, a capability that computers currently do not have. In reasoning about a complex system, humans reason approximately about its behavior, thereby maintaining only a generic understanding about the problem. Fortunately, this generality and ambiguity are sufficient for human comprehension of complex systems. As the quote from Dr Zadeh's *principle of incompatibility* suggests, complexity and ambiguity (imprecision) are correlated: "The closer one looks at a real-world problem, the fuzzier becomes its solution" (Zadeh, 1973).

As we learn more and more about a system, its complexity decreases and our understanding increases. As complexity decreases, the precision afforded by computational methods becomes more useful in modeling the system. For systems with little complexity and little uncertainty, closed-form mathematical expressions provide precise descriptions of the systems. For systems that are a little more complex, but for which significant data exist, model-free methods, such as artificial neural networks, provide a powerful and robust means to reduce some uncertainty through learning, based on patterns in the available data; unfortunately this learning is shallow. For complex systems where few numerical data exist and where only ambiguous or imprecise information may be available, fuzzy reasoning provides a way to understand system behavior by allowing us to interpolate approximately between observed input and output situations. Finally, for the most complex problems there are required forms of learning because of induction, or combinations of deduction and induction, that are necessary for even a limited level of understanding. This text does not address the most complex forms of learning as a result of induction, but the deductive methods involved in fuzzy reasoning are addressed here in terms of fuzzy systems models.

In constructing a fuzzy system model, Klir and Yuan (1995) describe the relationship among three characteristics that can be thought to maximize a model's usefulness. These characteristics of any model are complexity, credibility, and uncertainty. This relationship is only known in an abstract sense. Uncertainty, of course, plays a crucial role in any effort to maximize a systems model; but this crucial role can only be considered in the context of the other two characteristics. For example, allowing more uncertainty in a model reduces complexity and increases credibility of the resulting model. In developing models of complex systems, one needs to seek a balance of uncertainty and utility; a model that is extremely limited in terms of its robustness is one which cannot accommodate much uncertainty.

All models are mathematical abstractions of the real physical world. The more assumptions one needs to make to get the model into a form where known mathematical structures can be used to address the real problem, the more uncertainty has crept into the modeling process. To ignore this uncertainty is to ignore the real world, and our understanding of it. But, we can make the models robust and credible by addressing the fact that complexity and uncertainty are inextricably related; when one is high, the other tends to be high, just as described by Zadeh's quote.

The illustrations in Figures 8.1 and 8.2 provide some thoughts on the relationship of complexity and uncertainty. In Figure 8.1, we see that the case of ignorance is a situation involving high levels of complexity and uncertainty. Ignorance is the case where we have no specific information and we have no ideas about the physics that might describe the behavior of a system; an example might be our attempt to understand the concept of *infinity*. When we have

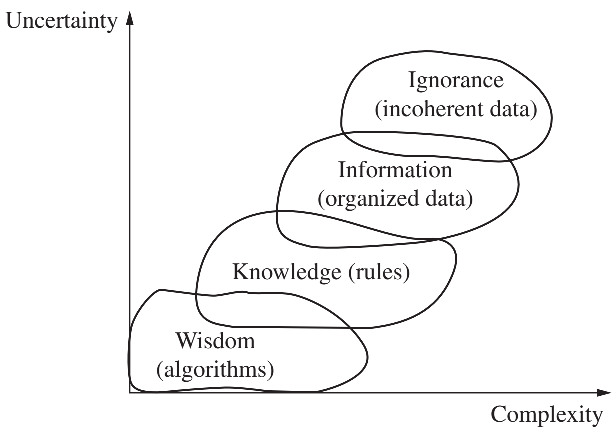

Figure 8.1 Complexity and uncertainty: relationships to ignorance, information, knowledge, and wisdom.

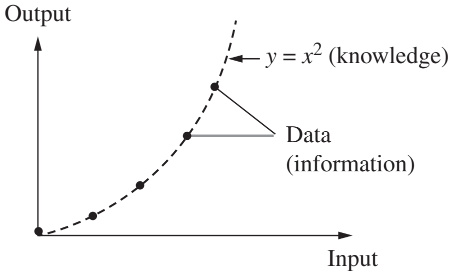

Figure 8.2 Specific example of information and knowledge in a simple one-dimensional relationship.

information about a problem, perhaps described by a random collection of data, we see that our complexity and uncertainty are reduced somewhat, but we still do not have good understanding about a problem (e.g., an attempt to invest in a new technology in the stock market). As uncertainty and complexity diminish further we get to the case of knowledge. In this instance, we have enough information and enough learning about a problem that we can actually pose rules, or algorithms that describe the essential characteristics of a problem (e.g., knowledge of a nonlinear system on the basis of a robust collection of deductive inferences). The final step in understanding I shall term *wisdom*. This is the case where uncertainty and complexity are at their lowest levels. This is because with wisdom we fully understand a problem in all its manifestations and possible configurations. An example of wisdom might be Newton's second law, where we have a mathematical algorithm that describes fully the relationship among mass, acceleration, and force.

In Figure 8.2, this idea is illustrated in a more specific way. We might have a collection of data points that, together as a population of information, is meaningless without some knowledge of how they are related. If we can find an algorithm that relates all of the data, and leaves no data point outside of its description, then we have a case of knowledge; we know the relationship between the inputs and outputs in a limited domain of applicability.

In the process of abstraction from the real world to a model, we need to match the model type with the character of the uncertainty exhibited in the problem. In situations where precision is available in the information, fuzzy systems are less efficient than the more precise algorithms in providing us with the best understanding of the problem. On the other hand, fuzzy systems can focus on models characterized by imprecise or ambiguous information; such models are sometimes termed *nonlinear models*.

Virtually all physical processes in the real world are nonlinear or complex in some way or the other. It is our abstraction of the real world that leads us to the use of linear systems in modeling these processes. The linear systems are simple and understandable, and, in many situations, they provide acceptable simulations of the actual processes that we observe. Unfortunately, only the simplest of systems can be modeled with linear system theory and only a small fraction of the nonlinear systems have verifiable solutions. The bulk of the physical processes that we must address are too complex to be reduced to algorithmic form—linear or nonlinear. Most observable processes have only a small amount of information available with which to develop an algorithmic understanding. The vast majority of information we have on most processes tends to be nonnumeric and nonalgorithmic. Most of the information is fuzzy and linguistic in form.

There is a quote from H. W. Brand (1961) that forms an appropriate introduction to matters in this chapter: "There is no idea or proposition in the field, which cannot be put into mathematical language, although the utility of doing so can very well be doubted." We can always reduce a complicated process to simple mathematical form. And, for a while at least, we may feel comfortable that our model is a useful replicate of the process we seek to understand. However, reliance on simple linear, or nonlinear, models of the algorithmic form can lead to quite disastrous results, as many engineers have found in documented failures of the past.

A classic example in mechanics serves to illustrate the problems encountered in overlooking the simplest of assumptions. In most beginning texts in mechanics, Newton's second law is described by the following equation:

$$\sum F = m \cdot a, \tag{8.1}$$

which states that the motion (acceleration) of a body under an imbalance of external forces acting on the body is equal to the sum of the forces (ΣF) divided by the body's mass (m). Specifically, the forces and acceleration of the body are vectors containing magnitude and direction. Unfortunately, Equation (8.1) is not specifically Newton's second law. Newton hypothesized that the imbalance of forces was equivalent to the rate of change in the momentum ($m \cdot v$) of the body, that is,

$$\sum F = \frac{d(m \cdot v)}{\mathrm{d}t} = m \cdot \frac{\mathrm{d}v}{\mathrm{d}t} + v \cdot \frac{\mathrm{d}m}{\mathrm{d}t}, \tag{8.2}$$

where ν is the velocity of the body and t is time. As one can see, Equations (8.1) and (8.2) are not equivalent unless the body's mass does not change with time (i.e., $dm/dt = 0$). In many mechanics applications, the mass does not change with time, but in other applications, such as in the flight of spacecraft or aircraft, where fuel consumption reduces total system mass, mass most certainly changes over time. It may be asserted that such an oversight has nothing to do with the fact that Newton's second law is not a valid algorithmic model, but rather it is a model that must be applied against an appropriate physical phenomenon. The point is this: algorithmic models are useful only when we understand and can observe all the underlying physics of the process. In the aircraft example, fuel consumption may not have been an observable phenomenon, and Equation (8.1) could have been applied to the model. Most complex problems have only a few observables, and an understanding of all the pertinent physics is usually not available.

If a process can be described algorithmically, we can describe the solution set for a given input set. If the process is not reducible to algorithmic form, perhaps the input–output features of the system are at least observable or measurable. This chapter deals with systems that cannot be simulated with conventional crisp or algorithmic approaches but that can be simulated because of the presence of other information—observed or linguistic—using fuzzy nonlinear simulation methods. However, it must be emphasized here, that this chapter is written, in a sense, in a bit of a *backward* way. We introduce the idea of nonlinear simulation by actually looking at examples where we *already* have the wisdom mentioned; that is, we already know the underlying algorithm (Examples 8.1–8.3). For real, complex processes we will certainly not have the underlying algorithms, but we will have our observations expressed in linguistic rules.

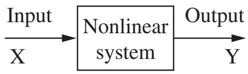

Figure 8.3 A general static physical system with observed inputs and outputs.

This chapter proposes to use fuzzy rule-based systems as suitable representations of simple and complex physical systems. For this purpose, a fuzzy rule-based system consists of (1) a set of rules that represent the engineer's understanding of the behavior of the system, (2) a set of input data observed going into the system, and (3) a set of output data coming from the system. The input and output data can be numerical, or they can be nonnumeric observations. Figure 8.3 shows a general static physical system, which could be a simple mapping from the input space to the output space, an industrial control system, a system identification problem, a pattern recognition process, or a decision-making process.

The input data, rules, and output actions or consequences are generally fuzzy sets expressed by means of appropriate membership functions (MFs) defined on an appropriate universe of discourse. The method of evaluation of rules is known as *approximate reasoning* or *interpolative reasoning* and is commonly represented by the composition of the fuzzy relations that are formed by the IF–THEN rules (see Chapter 5).

Three spaces are present in the general system posed in Figure 8.3 (Ross, 1995):

1. The space of possible conditions of the inputs to the system, which, in general, can be represented by a collection of fuzzy subsets $\underset{\sim}{A}^k$, for $k = 1, 2, \ldots$, which are fuzzy partitions of input space X, expressed by means of membership functions (MFs)

$$\mu_{\underset{\sim}{A}^k}(x), \quad \text{where } k = 1, 2, \ldots \tag{8.3}$$

2. The space of possible output consequences, based on some specific conditions of the inputs, which can be represented by a collection of fuzzy subsets $\underset{\sim}{B}^p$, for $p = 1, 2, \ldots$, which are fuzzy partitions of output space Y, expressed by means of MFs

$$\mu_{\underset{\sim}{B}^p}(y), \quad \text{where } p = 1, 2, \ldots \tag{8.4}$$

3. The space of possible mapping relations from the input space X onto the output space Y. The mapping relations are, in general, represented by fuzzy relations $\underset{\sim}{R}^q$, for $q = 1, 2, \ldots$, and expressed by means of MFs

$$\mu_{\underset{\sim}{R}^q}(x, y), \quad \text{where } q = 1, 2, \ldots \tag{8.5}$$

A human perception of the system shown in Figure 8.3 is based on experience and expertise, empirical observation, intuition, a knowledge of the physics of the system, or a set of subjective preferences and goals. The human observer usually puts this type of knowledge in the form of a set of unconditional as well as conditional propositions in natural language. Our understanding

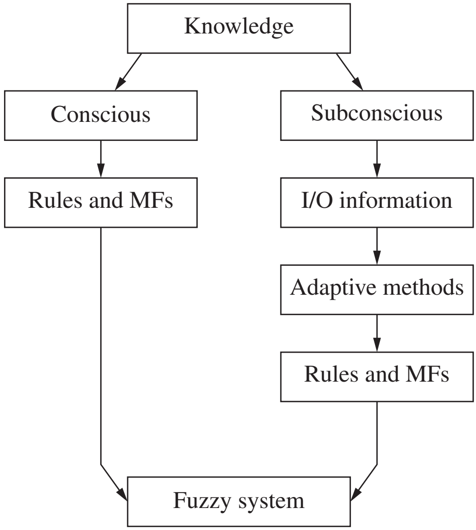

Figure 8.4 Two paths for knowledge to result in a fuzzy system.

of complex systems is at a qualitative and declarative level, based on vague linguistic terms; this is our so-called *fuzzy* level of understanding of the physical system. We have also seen in Chapters 6 and 7 that fuzzy rules and MFs can be derived from a family of input–output data, but this form of the knowledge has the same utility whether it was derived by human understanding or observation or developed in an automated fashion. Figure 8.4 shows this comparison. In the formulation of knowledge, we can have two paths: a conscious path where the rules and MFs are derived intuitively by the human and a subconscious path where we only have input–output (I/O) data or information and we use automated methods, such as those illustrated in Chapters 6 and 7, to derive the rules and MFs. The result of both paths is the construction of a fuzzy system, as shown in Figure 8.4.

Fuzzy Relational Equations

Consider a typical crisp nonlinear function relating elements of a single-input variable, say x, to the elements of a single-output variable, say y, as shown in Figure 8.5. Notice in Figure 8.5 that every x in the domain of the independent variable (each x') is "related" to a y (y') in the dependent variable (we call this relation a "mapping" in Chapter 12). The curve in Figure 8.5 represents a transfer function, which, in generic terms, is a relation. In fact, any continuous-valued function, such as the curve in Figure 8.5, can be discretized and reformulated as a matrix relation.

Example 8.1

For the nonlinear function $y = x^2$, we can formulate a matrix relation to model the mapping imposed by the function. Discretize the independent variable x (the input variable) on the domain $x = -2, -1, 0, 1, 2$. We find that the mapping provides for the dependent variable y

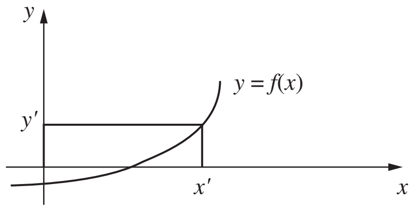

Figure 8.5 A crisp relation represented as a nonlinear function.

(the output variable) to take on the values $y = 0, 1, 4$. This mapping can be represented by a matrix relation, R, or

$$\mathrm{R} = \begin{array}{c} \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 4 \end{array} \\ \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

The elements in the crisp relation, R, are the indicator values as given later in Chapter 12.

We saw in Chapter 5 that a fuzzy relation can also represent a logical inference. The fuzzy implication IF $\underset{\sim}{A}$ THEN $\underset{\sim}{B}$ is known as the *generalized modus ponens* form of inference. There are numerous techniques for obtaining a fuzzy relation $\underset{\sim}{R}$ that will represent this inference in the form of a fuzzy relational equation given as

$$\underset{\sim}{B} = \underset{\sim}{A} \circ \underset{\sim}{R}, \tag{8.6}$$

where $\circ$ represents a general method for composition of fuzzy relations. Equation (8.6) appeared previously in Chapter 5 as the generalized form of approximate reasoning, where Equations (5.4) and (5.5) provided two of the most common forms of determining the fuzzy relation $\underset{\sim}{R}$ from a single rule of the form IF $\underset{\sim}{A}$ THEN $\underset{\sim}{B}$.

Nonlinear Simulation Using Fuzzy Systems

Suppose our knowledge concerning a certain nonlinear process is not algorithmic, like the algorithm $y = x^2$ in Example 8.1, but rather is in some other more complex form. This more complex form could be data observations of measured inputs and measured outputs. Relations can be developed from these data that are analogous to a lookup table, and methods for this step have been given in Chapter 3. Alternatively, the complex form of the knowledge of a nonlinear process could be described with some linguistic rules of the form IF $\underset{\sim}{A}$ THEN $\underset{\sim}{B}$. For example, suppose we are monitoring a thermodynamic process involving an input heat, measured by temperature, and an output variable, pressure. We observe that when we use a "low" temperature, we get out of the process a "low" pressure; when we input a "moderate" temperature, we see a "high" pressure in the system; when we input "high" temperature into the thermodynamics of the system, the output pressure reaches an "extremely high" value; and so on. This process is shown in Figure 8.6, where the inputs are now *not* points in the input universe (heat) and the output universe (pressure), but *patches* of the variables in each universe. These patches represent the fuzziness in describing the variables linguistically. Obviously, the mapping describing this relationship between heat and pressure is fuzzy. That is, patches from the input space map, or relate, to patches in the output space; the relations R^1, R^2, and R^3 in Figure 8.6 represent the fuzziness in this mapping. In general, all the patches, including those representing the relations, overlap because of the ambiguity in their definitions.

Each of the patches in the input space shown in Figure 8.6 could represent a fuzzy set, say $\underset{\sim}{A}$, defined on the input variable, say x; each of the patches in the output space could be represented by a fuzzy set, say $\underset{\sim}{B}$, defined on the output variable, say y; and each of the patches lying on the general nonlinear function path could be represented by a fuzzy relation, say $\underset{\sim}{R}^k$, where $k = 1, 2, \ldots, r$ represents r possible linguistic relationships between input and output. Suppose we have a situation where a fuzzy input, say x, results in a series of fuzzy outputs, say y^k, depending on which fuzzy relation, $\underset{\sim}{R}^k$, is used to determine the mapping. Each of these relationships, as listed in Table 8.1, could be described by what is called a *fuzzy relational equation*, where y^k is the output of the system contributed by the kth rule, and whose MF is given by

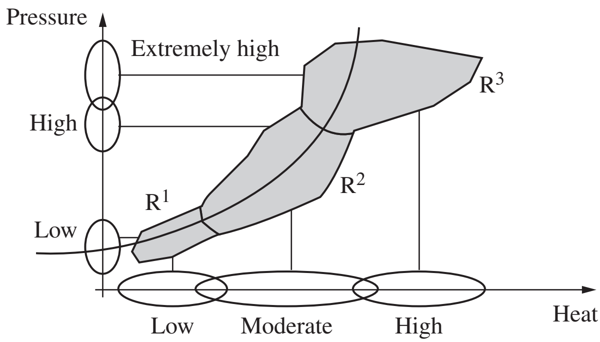

Figure 8.6 A fuzzy nonlinear relation matching patches in the input space to patches in the output space.

Table 8.1 System of fuzzy relational equations.

$\underset{\sim}{R}^1$:	$y^1 = x \circ \underset{\sim}{R}^1$
$\underset{\sim}{R}^2$:	$y^2 = x \circ \underset{\sim}{R}^2$
$\vdots$	$\vdots$
$\underset{\sim}{R}^r$:	$y^r = x \circ \underset{\sim}{R}^r$

$\mu_{y^k}(x)$. Both x and y^k ($k=1,2,\ldots,r$) can be written as single-variable fuzzy relations of dimensions $1\times n$ and $1\times m$, respectively. The unary relations, in this case, are actually similarity relations between the elements of the fuzzy set and a most typical or prototype element, usually with membership value equal to unity.

The system of fuzzy relational equations given in Table 8.1 describes a general fuzzy nonlinear system. If the fuzzy system is described by a system of conjunctive rules, we could decompose the rules into a single aggregated fuzzy relational equation by making use of Equations (5.31) for each input, x, as follows:

$$y=(x\circ \underset{\sim}{R}^1)\ \text{AND}\ (x\circ \underset{\sim}{R}^2)\ \text{AND} \ldots \text{AND}\ (x\circ \underset{\sim}{R}^r)$$

and equivalently,

$$y=x\circ(\underset{\sim}{R}^1\ \text{AND}\ \underset{\sim}{R}^2\ \text{AND}\ldots\text{AND}\ \underset{\sim}{R}^r),$$

and finally

$$y=x\circ \underset{\sim}{R}, \tag{8.7}$$

where $\underset{\sim}{R}$ is defined as

$$\underset{\sim}{R}=\underset{\sim}{R}^1\cap \underset{\sim}{R}^2\cap\cdots\cap \underset{\sim}{R}^r. \tag{8.8}$$

The aggregated fuzzy relation $\underset{\sim}{R}$ in Equation (8.8) is called the *fuzzy system transfer relation* for a single input, x. For the case of a system with n noninteractive fuzzy inputs (Chapter 2), x_i, and a single output, y, described in Equation (5.33), the fuzzy relational Equation (8.7) can be written in the form

$$y=x_1\circ x_2\circ\ldots\circ x_n\circ \underset{\sim}{R}. \tag{8.9}$$

If the fuzzy system is described by a system of disjunctive rules, we could decompose the rules into a single aggregated fuzzy relational equation by making use of Equation (5.32) as follows:

$$y=(x\circ \underset{\sim}{R}^1)\ \text{OR}\ (x\circ \underset{\sim}{R}^2)\ \text{OR}\ldots\text{OR}\ (x\circ \underset{\sim}{R}^r)$$

and equivalently,

$$y=x\circ(\underset{\sim}{R}^1\ \text{OR}\ \underset{\sim}{R}^2\ \text{OR}\ldots\text{OR}\ \underset{\sim}{R}^r),$$

and finally

$$y=x\circ \underset{\sim}{R}, \tag{8.10}$$

where $\underset{\sim}{R}$ is defined as

$$\underset{\sim}{R} = \underset{\sim}{R}^1 \cup \underset{\sim}{R}^2 \cup \ldots \cup \underset{\sim}{R}^r. \tag{8.11}$$

The aggregated fuzzy relation, that is, $\underset{\sim}{R}$, again is called the *fuzzy system transfer relation*.
For the case of a system with n noninteractive (Chapter 2) fuzzy inputs, x_i, and single output, y, described as in Equation (5.33), the fuzzy relational Equation (8.10) can be written in the same form as Equation (8.9).

The fuzzy relation $\underset{\sim}{R}$ is context dependent and therefore has local properties with respect to the Cartesian space of the input and output universes. This fuzzy relation results from the Cartesian product of the fuzzy sets representing the inputs and outputs of the fuzzy nonlinear system. However, before the relation $\underset{\sim}{R}$ can be determined, one must consider the more fundamental question of how to *partition* the input and output spaces (universes of discourse) into meaningful fuzzy sets. Ross (1995) details methods in partitioning that are as a result of human intuition, and Chapters 6 and 7 show how partitioning is a natural consequence of automated methods.

A general nonlinear system, such as that in Figure 8.6, which comprises n inputs and m outputs, can be represented by fuzzy relational equations in the form expressed in Table 8.1. Each of the fuzzy relational equations, that is, $\underset{\sim}{R}^r$, can also be expressed in canonical rule-based form, as shown in Table 8.2.

The rules in Table 8.2 could be connected logically by any of "and," "or," or "else" linguistic connectives, and the variables in Table 8.2, x and y, are the input and output vectors, respectively, of the nonlinear system. Ross (1995) discusses in more detail the various forms of nonlinear systems that can result from a rule-based approach, but this level of detail is not needed in conducting general nonlinear simulations. Only the most general form of a nonlinear system is considered here, shown in Figure 8.7, where the inputs (x) and outputs (y) are considered as fuzzy sets, and where the input–output mappings (R) are considered as fuzzy relations.

Fuzzy Associative Memories (FAMs)

Consider a fuzzy system with n noninteractive (Chapter 2) inputs and a single output. Also assume that each input universe of discourse, that is, $X_1, X_2, \ldots, X_n$, is partitioned into k fuzzy partitions. Based on the canonical fuzzy model given in Table 8.2 for a nonlinear system, the total number of possible rules governing this system is given as

Table 8.2 Canonical rule-based form of fuzzy relational equations.

$\underset{\sim}{R}^1$:	IF x is $\underset{\sim}{A}^1$, THEN y is $\underset{\sim}{B}^1$
$\underset{\sim}{R}^2$:	IF x is $\underset{\sim}{A}^2$, THEN y is $\underset{\sim}{B}^2$
$\vdots$	$\vdots$
$\underset{\sim}{R}^r$:	IF x is $\underset{\sim}{A}^r$, THEN y is $\underset{\sim}{B}^r$

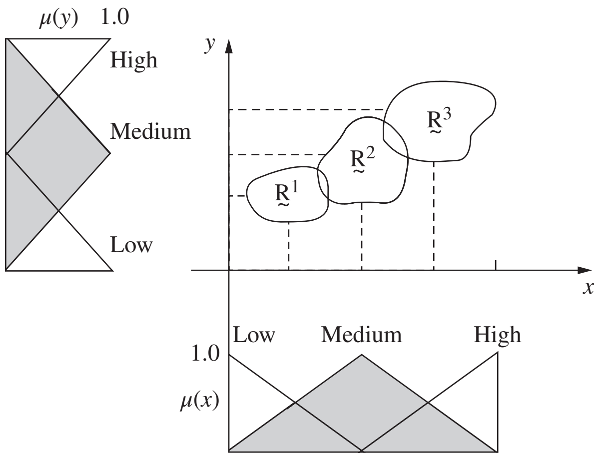

Figure 8.7 Fuzzy set inputs and fuzzy set outputs (the most general case).

$$l = k^n, \tag{8.12a}$$

$$l = (k+1)^n, \tag{8.12b}$$

where l is the maximum possible number of canonical rules. Equation (8.12b) is to be used if the partition "anything" is to be used, otherwise Equation (8.12a) determines the number of possible rules. The actual number of rules, r, necessary to describe a fuzzy system is much less than l, that is, $r \ll l$, because of the interpolative reasoning capability of the fuzzy model and because the fuzzy MFs of the partitions overlap. If each of the n noninteractive inputs is partitioned into a different number of fuzzy partitions, say, X_1 is partitioned into k_1 partitions and X_2 is partitioned into k_2 partitions, and so forth, then the maximum number of rules is given as

$$l = k_1 k_2 k_3 \ldots k_n. \tag{8.13}$$

For a small number of inputs, for example, $n = 1$ or $n = 2$, or $n = 3$, there exists a compact form of representing a fuzzy rule-based system. This form is illustrated for $n = 2$ in Figure 8.8. In the figure, there are seven partitions for input A (A_1 to A_7), five partitions for input B (B_1 to B_5), and four partitions for the output variable C (C_1 to C_4). This compact graphical form is called a *fuzzy associative memory* (FAM) table. As can be seen from the FAM table, the rule-based system actually represents a general nonlinear mapping from the input space of the fuzzy system to the output space of the fuzzy system. In this mapping, the patches of the input space are being applied to the patches in the output space. Each rule or, equivalently, each fuzzy relation from input to the output represents a fuzzy point of data that characterizes the nonlinear mapping from the input to the output.

Input B \ Input A	A_1	A_2	A_3	A_4	A_5	A_6	A_7
B_1	C_1		C_4	C_4		C_3	C_3
B_2		C_1				C_2	
B_3	C_4		C_1			C_2	C_2
B_4	C_3	C_3		C_1		C_1	C_2
B_5	C_3		C_4	C_4	C_1		C_3

Figure 8.8 FAM table for a two-input, single-output fuzzy rule-based system.

In the FAM table in Figure 8.8, we see that the maximum number of rules for this situation, using Equation (8.13), is $l = k_1 k_2 = 7(5) = 35$; but as seen in the figure, the actual number of rules is only $r = 21$.

We will now illustrate the ideas involved in simulation with three examples from various engineering disciplines.

Example 8.2

For the nonlinear function $y = 10 \sin x_1$, we will develop a fuzzy rule-based system using four simple fuzzy rules to approximate the output y. The universe of discourse for the input variable x^1 will be the interval [−180°, 180°], and the universe of discourse for the output variable y is the interval [−10, 10].

First, we will partition the input space x_1 into five simple partitions on the interval [−180°, 180°], and we will partition the output space y on the interval [−10, 10] into three MFs, as shown in Figure 8.9a and 8.9b, respectively. In these figures, the abbreviations NB, NS, Z, PS, and PB refer to the linguistic variables "negative-big," "negative-small," "zero," "positive-small," and "positive-big," respectively.

Second, we develop four simple rules, listed in Table 8.3, that we think emulate the dynamics of the system (in this case the system is the nonlinear equation $y = 10 \sin x_1$ and we are observing the harmonics of this system) and that make use of the linguistic variables in Figure 8.9. The FAM table for these rules is given in Table 8.4.

The FAM table of Table 8.4 is one dimensional because there is only one input variable, x^1. As seen in Table 8.4, all rules listed in Table 8.3 are accommodated. Not all the four rules expressed in Table 8.3 are expressed in canonical form (some have disjunctive antecedents), but if they were transformed into canonical form, they would represent the five rules provided in the FAM table in Table 8.4.

In developing an approximate solution for the output y, we select a few input points and employ a graphical inference method similar to that illustrated in Chapter 5. We will use the centroid method for defuzzification. Let us choose four crisp singletons as the input:

$$x_1 = \{-135°, -45°, 45°, 135°\}.$$

For input $x_1 = -135°$, Rules 3 and 4 are fired, as shown in Figure 8.10c and d. For input $x_1 = -45°$, Rules 1, 3, and 4 are fired. Figure 8.10a and b show the graphical inference for input $x_1 = -45°$ (which fires Rule 1), and for $x_1 = 45°$ (which fires Rule 2), respectively.

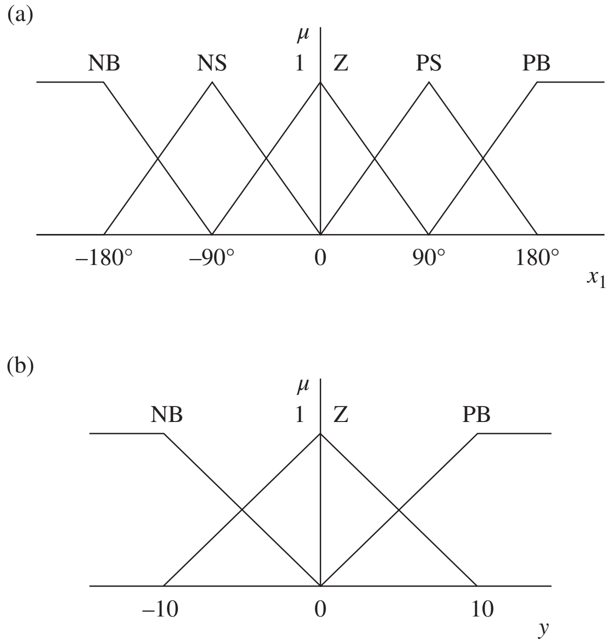

Figure 8.9 Fuzzy membership functions for the input and output spaces: (a) five partitions for the input variable, x_1; (b) three partitions for the output variable, y.

Table 8.3 Four simple rules for $y = 10 \sin x_1$.

1	IF x_1 is Z or PB, THEN y is Z
2	IF x_1 is PS, THEN y is PB
3	IF x_1 is Z or NB, THEN y is Z
4	IF x_1 is NS, THEN y is NB

Table 8.4 FAM for the four simple rules in Table 8.3.

x_i	NB	NS	Z	PS	PB
y	Z	NB	Z	PB	Z

For input $x_1 = -45°$, Rules 3 and 4 are also fired, and we get results similar to those shown in Figure 8.10c and d after defuzzification:

Rule 3: $y = 0$.
Rule 4: $y = -7$.

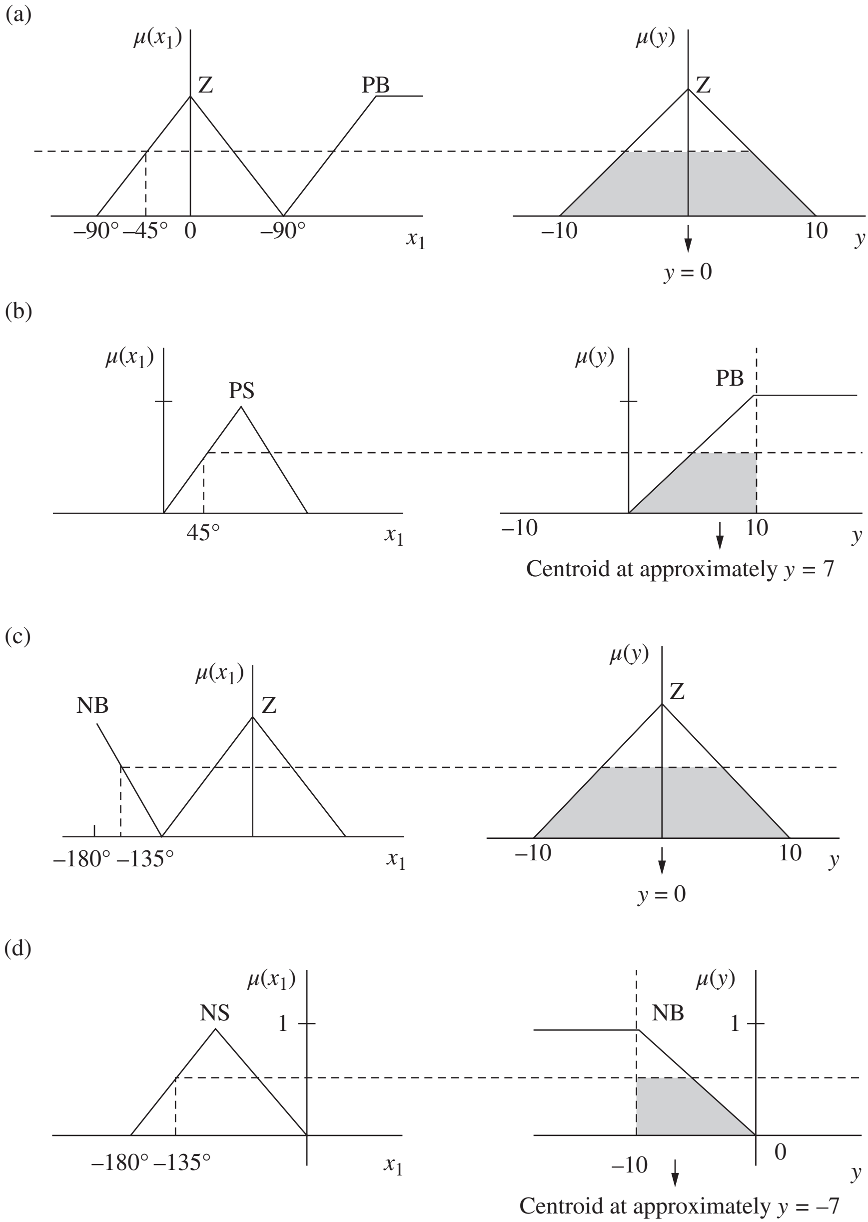

Figure 8.10 Graphical inference method showing membership propagation and defuzzification: (a) input $x_1 = -45°$ fires Rule 1; (b) input $x_1 = 45°$ fires Rule 2; (c) input $x_1 = -135°$ fires Rule 3; (d) input $x_1 = -135°$ fires Rule 4.

For $x_1 = 45°$, Rules 1–3 are fired (see Figure 8.10b for Rule 2), and we get the following results for Rules 1, 2 and 3 after defuzzification:

Rule 1: $y = 0$.
Rule 2: $y = 7$
Rule 3: $y = 0$.

For $x_1 = 135°$, Rules 1 and 2 are fired and we get, after defuzzification, results that are similar to those shown in Figure 8.10b:

Rule 1: $y = 0$.
Rule 2: $y = 7$.

When we combine the results, we get an aggregated result summarized in Table 8.5 and shown graphically in Figure 8.11. The y values in each column of Table 8.5 are the defuzzified results from various rules firing for each of the inputs, x_i. When we aggregate the rules using the union operator (disjunctive rules), the effect is to take the maximum value for y in each of the columns in Table 8.5 (i.e., maximum value irrespective of sign). The plot in Figure 8.11 represents the maximum y for each of the x_i, and it represents a fairly accurate portrayal of the true

Table 8.5 Defuzzified results for simulation of $y = 10 \sin x_1$.

x_1	−135°	−45°	45°	135°
y	0	0	0	0
	−7	0	0	7
		−7	7	

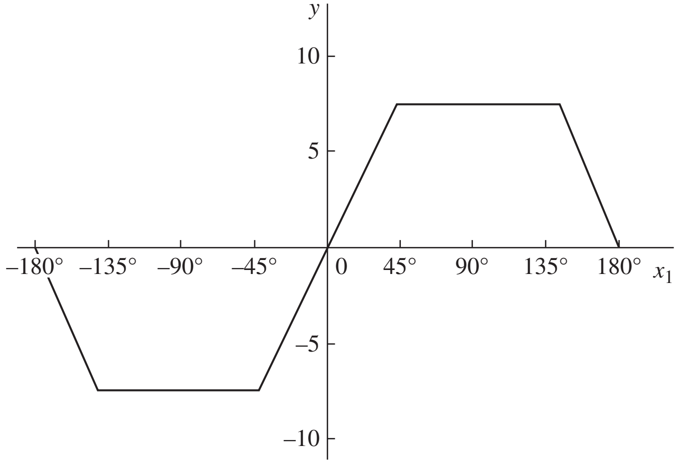

Figure 8.11 Simulation of nonlinear system $y = 10 \sin x_1$ using a four-rule fuzzy rule-base.

solution, given only a crude discretization of four inputs and a simple simulation based on four rules. More rules would result in a closer fit to the true sine curve (Problem 8.6).

Example 8.3

Suppose we want to model a serial transmission of a digital signal over a channel using RS232 format. Packets of information transmitted over the channel are ASCII characters composed of start and stop bits plus the appropriate binary pattern. If we wanted to know whether a valid bit was sent we could test the magnitude of the signal at the receiver using an absolute value function. For example, suppose we have the voltage (V) versus time trace shown in Figure 8.12, a typical pattern. In this pattern, the ranges for a valid mark and a valid space are as follows:

−12 to −3 V or +3 to +12 V   A valid mark (denoted by a one)
−3 to +3   A valid space (denoted by a zero)

The absolute value function used to make this distinction is a nonlinear function, as shown in Figure 8.13. To use this function on the scale of voltages [−12, +12], we will attempt to simulate the nonlinear function $y = 12|x|$, where the range of x is [−1, 1]. First, we partition the input space, $x = [-1, 1]$, into five linguistic partitions as in Figure 8.14. Next, we partition the output space. This task can usually be accomplished by mapping prototypes of input space to corresponding points in output space, if such information is available. Because we know the functional mapping (normally we would not know this for a real, complex, or nonlinear problem), the partitioning can be accomplished readily; we will use three equally spaced output partitions as shown in Figure 8.15.

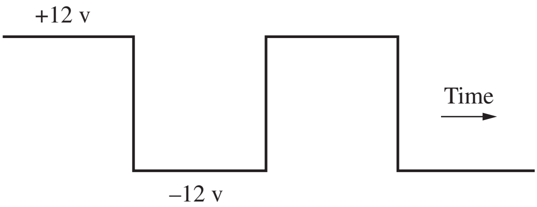

Figure 8.12 Typical pattern of voltage versus time for a valid bit mark.

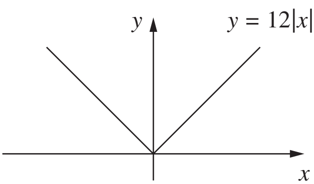

Figure 8.13 Nonlinear function $y = 12|x|$.

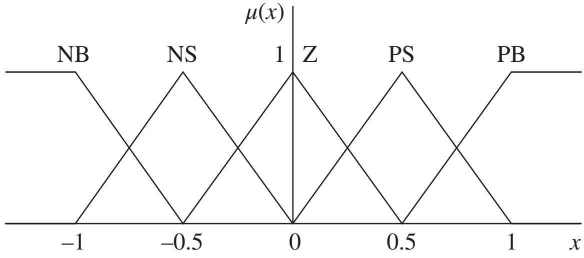

Figure 8.14 Partitions on the input space for $x = [-1, 1]$.

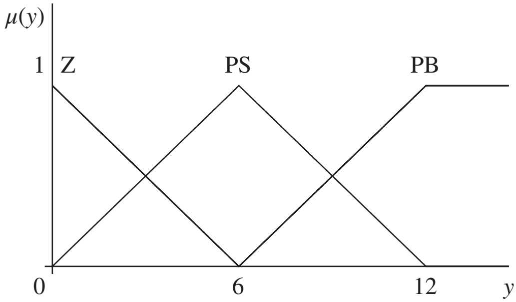

Figure 8.15 Output partitions on the range $y = [0, 12]$.

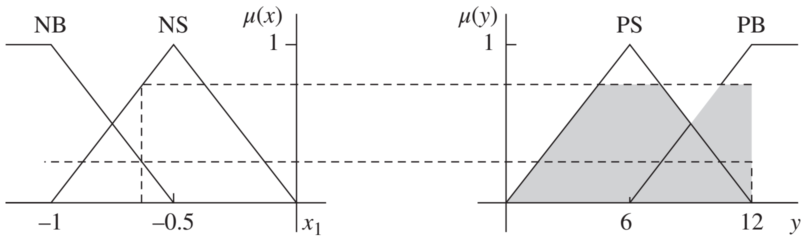

Figure 8.16 Graphical simulation for crisp input $x = -0.6$.

Because this function is simple and nonlinear, we can propose a few simple rules to simulate its behavior:

1. IF x = zero, THEN y = zero.
2. IF x = NS or PS, THEN y = PS.
3. IF x = NB or PB, THEN y = PB.

We can now conduct a graphical simulation of the nonlinear function expressed by these three rules. Let us assume that we have five input values, the crisp singletons $x = -0.6$, -0.3, 0, 0.3, and 0.6. The input $x = -0.6$ invokes (fires) Rules 2 and 3, as shown in Figure 8.16. The defuzzified output, using the centroid method, for the truncated union of the two consequents is approximately 8. The input $x = -0.3$ invokes (fires) Rules 1 and 2, as shown in Figure 8.17. The defuzzified output for the truncated union of the two consequents is approximately 5. The input $x = 0$ invokes (fires) Rule 1 only, as shown in Figure 8.18. The defuzzified output for the truncated consequent ($y = Z$) is a centroidal value of 2. By symmetry it is easy to see that crisp inputs $x = 0.3$ and $x = 0.6$ result in defuzzified values for $y \approx 5$ and $y \approx 8$, respectively.

If we plot these simulated results and compare them to the exact relationship (which, again, we would not normally know), we get the graph in Figure 8.19; the simulation, although approximate, is quite good.

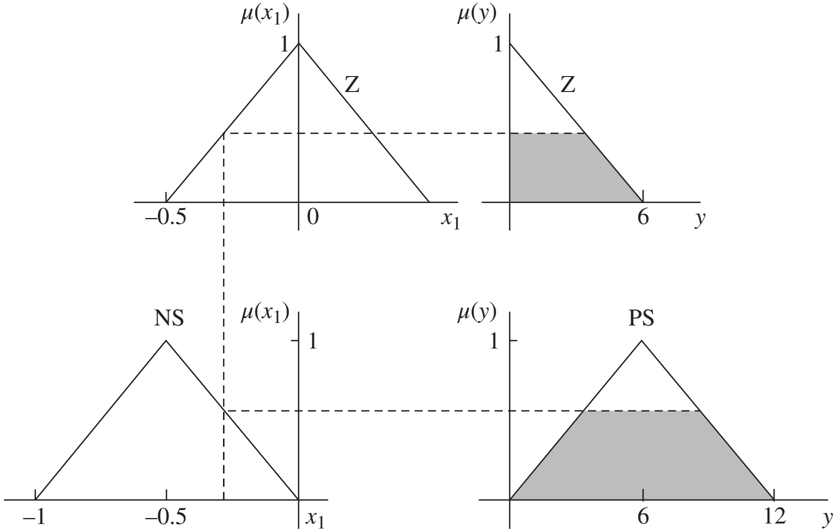

Figure 8.17 Graphical simulation for crisp input $x = -0.3$.

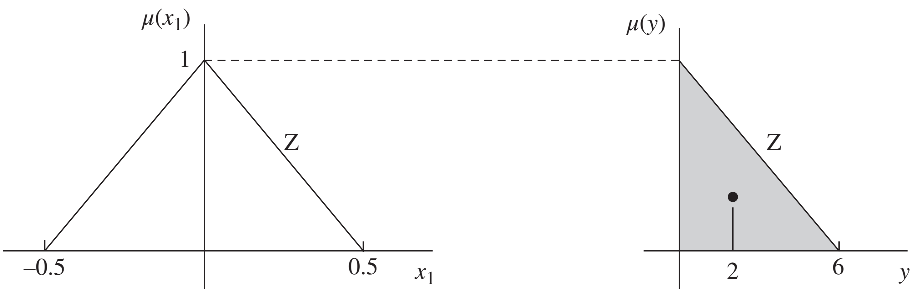

Figure 8.18 Graphical simulation for crisp input $x = 0$.

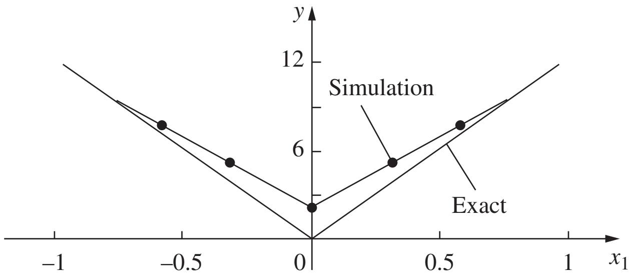

Figure 8.19 Simulated versus exact results for Example 8.3.

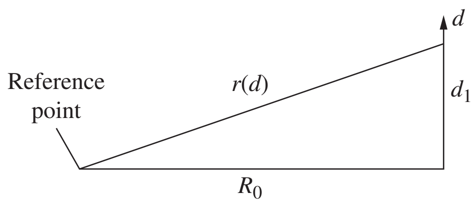

Figure 8.20 Schematic of aircraft SAR problem.

Example 8.4

When an aircraft forms a synthetic aperture radar (SAR) image, the pilot needs to calculate the range to a reference point, based on the position of the craft and the output of an inertial navigator, to within some fraction of a wavelength of the transmitted radar pulse. Assume that at position $d = 0$, the aircraft *knows* that the reference point is distance R_0 off the left broadside (angle = 90°) of the aircraft, and that the aircraft flies in a straight line; see Figure 8.20. The question is: What is the range, $r(d)$, to the reference point when the aircraft is at the position d_1? The exact answer is $r(d) = \left(R_0^2 + d_1^2\right)^{1/2}$; however, the square root operation is nonlinear, cumbersome, and computationally slow to evaluate. In a typical computation, this expression is expanded into a Taylor series. In this example, we wish to use a fuzzy rule-based approach instead.

If we normalize the range, that is, let $d_1/R_0 = k^1 \cdot x_1$, then $r(x_1) = R_0\left(1 + k_1^2 x_1^2\right)^{1/2}$, where now x_1 is a scaled range and k_1 is simply a constant in the scaling process. For example, suppose we are interested in the range $|d_1/R_0| \le 0.2$; then $k_1 = 0.2$ and $|x_1| \le 1$. For this particular problem we will let $R_0 = 10\,000$ m = 10 km; then $r(x_1) = 10000\left[1 + (0.04)x_1^2\right]^{1/2}$. Table 8.6 shows exact values of $r(x_1)$ for typical values of x_1.

Let $y = r(x_1)$ with x_1 partitioned as shown in Figure 8.21, and let the output variable, y, be partitioned as shown in Figure 8.22. In Figure 8.22, the partitions $\underset{\sim}{\mathrm{S}}$ and $\underset{\sim}{\mathrm{L}}$ have symmetrical MFs. We now pose three simple rules that relate the input and output variables:

Rule 1: IF $x \in \underset{\sim}{\mathrm{Z}}$, THEN $y \in \underset{\sim}{\mathrm{S}}$.
Rule 2: IF $x \in \underset{\sim}{\mathrm{PS}}$ or $\underset{\sim}{\mathrm{NS}}$, THEN $y \in \underset{\sim}{\mathrm{M}}$.
Rule 3: IF $x \in \underset{\sim}{\mathrm{PB}}$ or $\underset{\sim}{\mathrm{NB}}$, THEN $y \in \underset{\sim}{\mathrm{L}}$.

Table 8.6 Relationships for distance in SAR problem.

x_1	$r(x_1)$
−1.0	10 198
−0.5	10 050
0.0	10 000
0.5	10 050
1.0	10 198

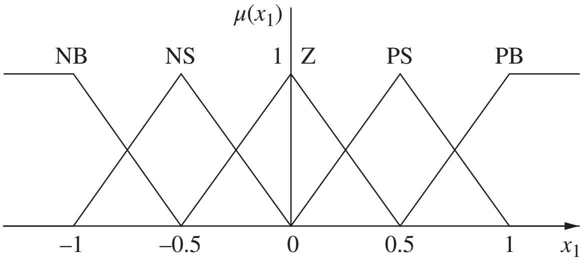

Figure 8.21 Partitioning for the input variable, x_1.

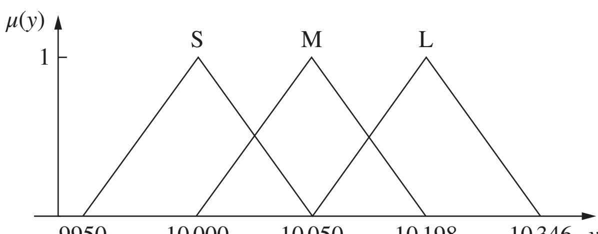

Figure 8.22 Partitioning for the output variable, y.

If we conduct a graphical simulation like that in Example 8.2, we achieve the results shown in Figure 8.23. In this figure, the open circle denotes exact values and the cross denotes the centroidal value of the fuzzy output as determined in the graphical simulation (in some cases the exact value and the fuzzy value coincide; this is represented by an open circle with a cross in it). The "approximate" curve follows the exact curve quite well. As a reminder, we would not normally know the exact values for a real problem whose algorithmic description was not known (this would be the case of knowledge, as described earlier in Figure 8.2).

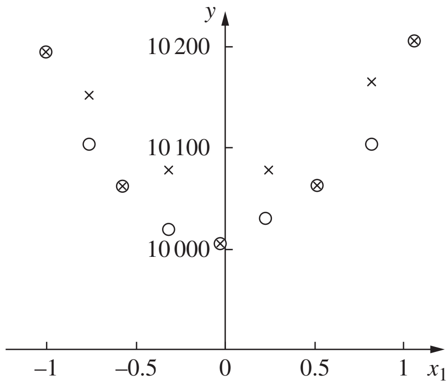

Figure 8.23 Exact and fuzzy values compared for SAR problem.

Summary

A wide class of complex dynamic processes exists where the knowledge regarding the functional relationship between the input and output variables may be established on numerical or nonnumerical information. The numerical information is usually from a limited number of data points and the nonnumerical information is in the form of vague natural language protocols gathered from interviews with humans familiar with the input–output behavior or the real-time control of the system or process. Complexity in the system model arises as a result of many factors such as (1) high dimensionality, (2) too many interacting variables, and (3) unmodeled dynamics such as nonlinearities, time variations, external noise or disturbance, and system perturbations (Ross, 1995). Hence, the information gathered on the system behavior is never complete, sharp, or comprehensive.

It has been shown that fuzzy systems theory is analogous to both a linear and an abstract algebra (Lucero, 2004). The context in which fuzzy systems theory is analogous to linear algebra and to abstract algebra is that they are common for the concepts of mapping and domain. A mapping is intuitively a correspondence between two elements. But when used with an aggregate of various mappings, the simple relations are weighted and the mapping is no longer intuitive. Stated simply, a fuzzy system is a mapping of a state. This state is defined on restricted domains. And the input variables are partitioned using a series of functions (MFs) that transform the variable to a degree on the interval [0, 1]. This degree is used to weigh the importance of a rule. More rules are defined and used, as the complexity of the system requires. The final output is a weighted value. The field of algebra encompasses a vast wealth of theories. In this field are the general disciplines of abstract algebra and linear algebra. Abstract algebra describes sets, relations, algebraic systems in general, and a linear algebra in part. Fuzzy systems do this abstraction as well, with sets which are isomorphic with linguistic knowledge. Linear algebra, as the computational kernel of this theory, contains the actual implementations, analogous to fuzzy compositions and implications. The foundation on which fuzzy systems theory is a universal approximator is based on a fundamental theorem from real analysis, the Stone–Weierstrass theorem (see references and discussion in Chapter 1).

Fuzzy mathematics provides a range of mathematical tools that helps the analyst formalize ill-defined descriptions about complex systems into the form of linguistic rules and then eventually into mathematical equations, which can then be implemented on digital computers. These rules can be represented by FAMs. At the expense of relaxing some of the demands on the requirements for precision in some nonlinear systems, a great deal of simplification, ease of computation, speed, and efficiency are gained when using fuzzy models. The ill-defined nonlinear systems can be described with fuzzy relational equations. These relations are expressed in the form of various fuzzy composition operations, which are carried out on classes of MFs defined on a number of overlapping partitions of the space of possible inputs (antecedents), possible mapping restrictions, and possible output (consequent) responses.

The MFs used to describe linguistic knowledge are enormously subjective and context dependent (Vadiee, 1993). The input variables are assumed to be noninteractive, and the MFs for them are assigned based on the degree of similarity of a corresponding prototypical element. Appropriate nonlinear transformations or sensory integration and fusion on input or output spaces are often used to reduce a complex process to a fuzzy system model. The net effect of this preprocessing on the input data is to decouple and linearize the system dynamics.

This chapter has dealt with the idea of fuzzy nonlinear simulation. The point made in this chapter is not that we can make crude approximations to well-known functions; after all, if we know a function, we certainly do not need fuzzy logic to approximate it. But there are many situations where we can only observe a complicated nonlinear process whose functional relationship we do not know, and whose behavior is known only in the form of linguistic knowledge, such as that expressed for the sine curve example in Table 8.3 or, for more general situations, as that expressed in Table 8.2. Then the power of fuzzy nonlinear simulation is manifested in modeling nonlinear systems whose behavior we can express in the form of input–output data-tuples, or in the form of linguistic rules of knowledge, and whose exact nonlinear specification we do not know. Fuzzy models to address complex systems are being published in the literature at an accelerating pace; see, for example, Huang and Fan (1993) who address complex hazardous waste problems and Sugeno (1985), or Sugeno and Yasukawa (1993) who address problems ranging from a chemical process to a stock price trend model. The ability of fuzzy systems to analyze dynamical systems that are so complex that we do not have a mathematical model is the point made in this chapter. As we learn more about a system, the data eventually become robust enough to pose the model in analytic form; at that point we no longer need a fuzzy model. Recent examples of nonlinear simulations using fuzzy methods can be found in Choi, Sherif, Reda Taha, and Chung, (2009) and Noureldin El-Shafie, and Reda Taha (2007). In both these works, a method similar to the original (see Jang, 1993) artificial neural-fuzzy inference system (ANFIS) is used, where the MFs are modified based on determining the objective function as the error function.

References

Brand, H. W. (1961). *The Fecundity of Mathematical Methods in Economic Theory*, trans. Edwin Holstrom (Dordrecht: D. Reidel).

Choi, K.-K., Sherif, A. G., Reda Taha, M. M., and Chung, L. (2009). Shear strength of slender reinforced concrete beams without web reinforcement: A model using fuzzy set theory. *Eng. Struct.*, **31** (3), 768–777.

Huang, Y., and Fan, L. (1993). A fuzzy logic-based approach to building efficient fuzzy rule-based expert systems, *Comput. Chem. Eng.*, **17** (2), 188–192.

Jang, J.-S.R. (1993). ANFIS: Adaptive-network-based fuzzy inference systems, *IEEE Trans. Syst., Man, Cybern.*, **23** (3), 665–685.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Application* (Upper Saddle River, NJ: Prentice Hall).

Lucero, J. (2004). Fuzzy systems methods in structural engineering, PhD dissertation, Department of Civil Engineering, University of New Mexico: Albuquerque, NM.

Noureldin, A., El-Shafie, A., and Reda Taha, M. M. (2007) Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation, *Eng. Appl. Artif. Intell.*, **20** (2), 37–48.

Ross, T. (1995) *Fuzzy Logic with Engineering Applications* (New York: McGraw-Hill).

Sugeno, M., ed. (1985). *Industrial Applications of Fuzzy Control* (New York: North-Holland).

Sugeno, M., and Yasukawa, T. (1993). A fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. Fuzzy Syst.*, **1** (1), 7–31.

Vadiee, N. (1993). Fuzzy rule-based expert systems—I and II, in M. Jamshidi, N. Vadiee, and T. Ross, eds., *Fuzzy logic and Control: Software and Hardware Applications*, chapters 4 and 5 (Englewood Cliffs, NJ: Prentice Hall).

Zadeh, L. (1973) Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. Syst. Man. Cybern.*, **SMC-3**: 28–44.

Zadeh, L. (1975). The concept of a linguistic variable and its application to approximate reasoning—I, *Inf. Sci.*, **8**, 199–249.

Problems

8.1 A video monitor's cathode-ray tube (CRT) has a nonlinear characteristic between the luminance output and the voltage input. This nonlinear characteristic is $y = x^{1.8}$, where y is the illumination and x is the voltage. The charge-coupled device (CCD) in a video camera has a linear light-in to voltage-out characteristic. To compensate for the nonlinear characteristic of the monitor, a "gamma correction" circuit is usually employed in a CCD camera. This nonlinear circuit has a transfer function of $y = x^{\text{gamma}}$, where the gamma factor is usually 0.45 (i.e., 1/2.2) to compensate for the 1.8 gamma characteristic of the monitor. The net result should be a linear response between the light incident on the CCD and the light produced by the monitor. Figure P8.1 shows the nonlinear gamma characteristic of a CCD camera (y_{actual}). Both the input, x, and the output, y, have a universe of discourse of [0, 1].

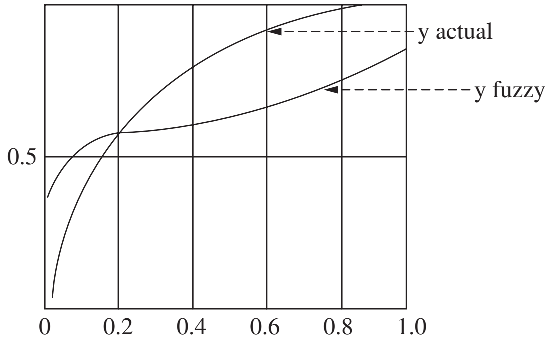

Figure P8.1

Partition the input variable, x, into three partitions, say small, S, medium, M, and big, B, and partition the output variable, y, into two partitions, say small, SM, and large, L. Using your own few simple rules for the nonlinear function $y = x^{0.45}$ and the crisp inputs $x = \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, determine whether your results produce a solution roughly similar to y_{fuzzy} in Figure P8.1 (which was developed with another fuzzy model [Ross, 1995]). Comment on the form of your solution and why it does or does not conform to the actual result.

8.2 A body of mass m is sliding down the slope of angle θ. The frictional force is F = mg∗sin (θ) (Figure P8.2a). Using the partitioning for the input variable θ as shown in Figure P8.2b and the partitioning for the output variable F as shown in Figure P8.2c and the following three simple rules,

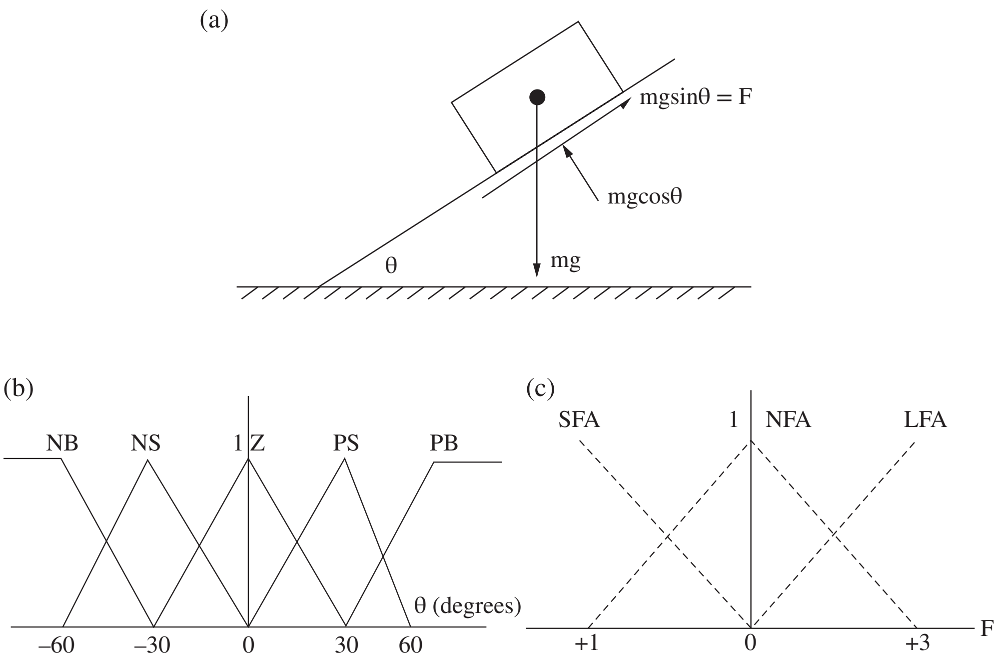

Figure P8.2a, b, c

1. If Z THEN NFA (No Force Acting)
2. If NS or PS THEN SFA (Small Force Acting)
3. If PS or NB THEN LFA (Large Force Acting)

conduct a graphical simulation and plot the results on a graph of F vs θ; show the associated exact solution on this same graph.

8.3 Psycho-acoustic research has shown that *white noise* has different effects on people's moods, depending on the average pitch of the tones that make up the noise. Very high and very low pitches make people nervous, whereas midrange noise has a calming effect.

The annoyance level of white noise can be approximated by a function of the square of the deviance of the average pitch of the noise from the central pitch of the human hearing range, approximately 10 kHz. As shown in Figure P8.3a, the human annoyance level can be modeled by the nonlinear function $y = x^2$, where x = deviance (kilohertz) from 10 kHz. The range of x is $[-10, 10]$; outside that range pitches are not audible to humans.

The partitions for the input variable, x, are the five partitions on the range $[-10, 10]$ kHz, as shown in Figure P8.3b, and the partitions for the output space for $y = x_2$ are shown in Figure P8.3c. Using the following three simple rules,

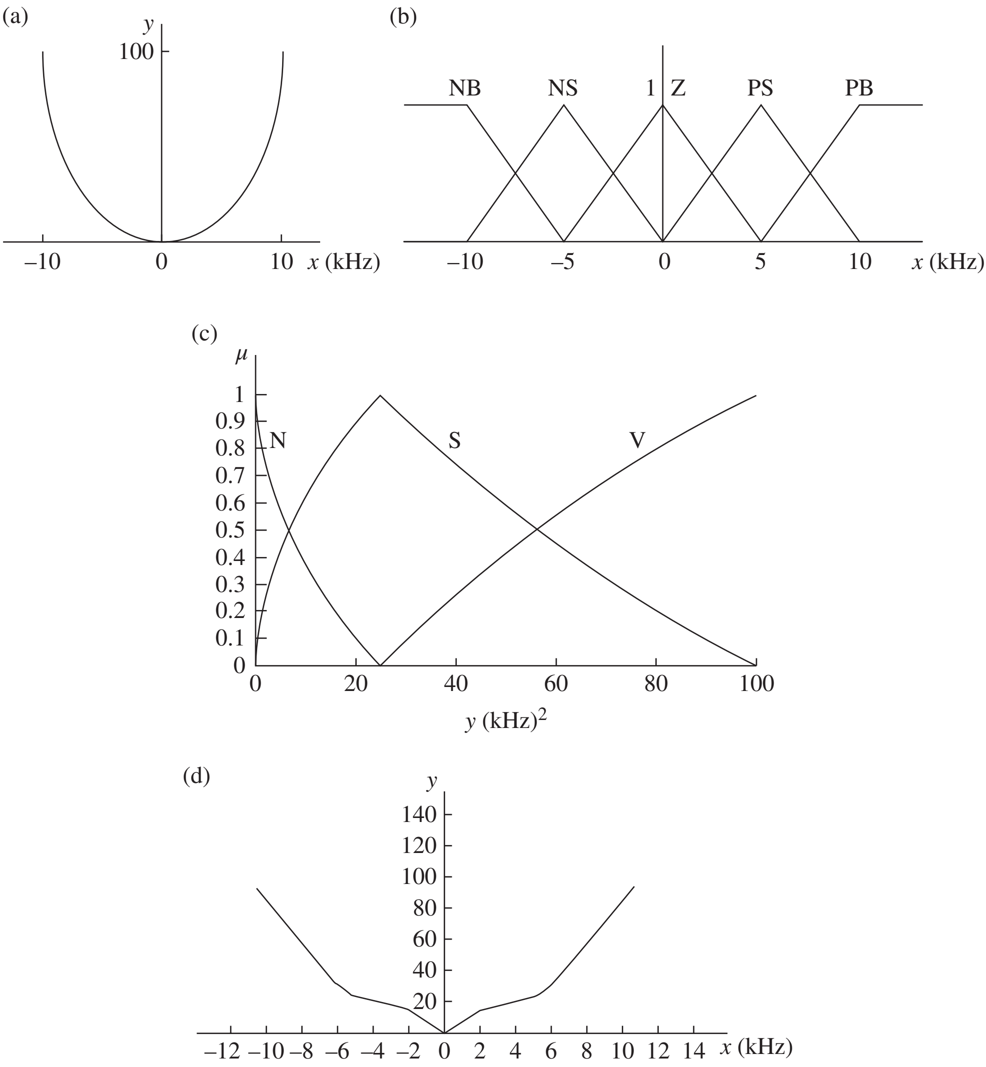

Figure P8.3a, b, c, d

1. IF $x = \mathrm{Z}$, THEN $y = \mathrm{N}$
2. IF x = NS or PS, THEN y = S
3. IF x = NB or PB, THEN y = V

show how a similar plot of fuzzy results, as shown in Figure 8.3d, is determined.

8.4 Let us consider the case of a series motor under the influence of a load and a constant voltage source, as shown in Figure P8.4a. A series motor should always be operated with some load, otherwise the speed of the motor will become excessively high, resulting in damage to the motor. The speed of the motor, N (rpm), is inversely related to the armature current, I_a, (amperes), by the expression $N = k/I_a$, where k is the flux. For this problem, we will estimate the flux parameter based on a motor speed of 1500 rpm at an armature current of 5 A; hence, $k = 5(1500) = 7500$ rpm A. Suppose we consider the armature current to vary in the range $I_a = [-\infty, +\infty]$, and we partition this universe of discourse as shown in Figure 8.4b (note that the extremes at $-\infty$ and $+\infty$ are contained in the partitions NB and PB, respectively). Suppose we also partition the output variable, N, as shown in Figure P8.4c. Using the input and output partitioning provided in Figure P8.4b–c and the following five rules, conduct a graphical numerical simulation for the crisp inputs $I_a = -8, -2, 3, 9$ A. Plot this response on a graph of N versus I_a.

IF I_a is Z, THEN N is HSC (high speed clockwise) or HSAC (high speed counter-clockwise (or anti-clockwise))

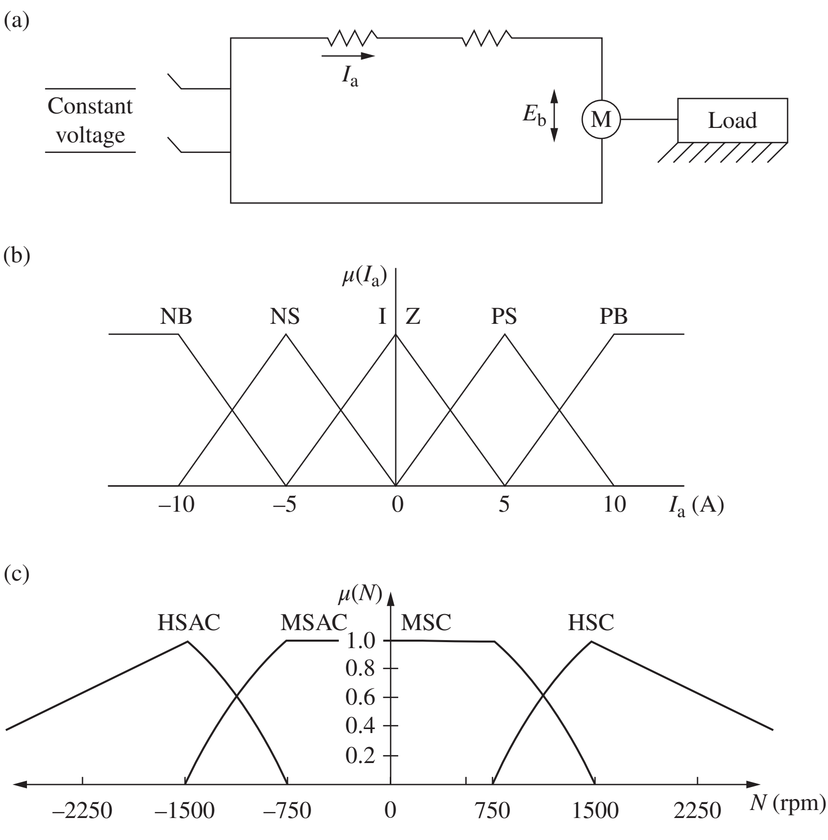

Figure P8.4a, b, c

IF I_a is PS, THEN N is HSC
IF I_a is NS, THEN N is HSAC
IF I_a is PB, THEN N is MSC (medium speed clockwise)
IF I_a is NB, THEN N is MSAC medium speed counter-clockwise (or anticlockwise).

8.5 In the field of image processing, a *limiter* function is used to enhance an image when background lighting is too high. The limiter function is shown in Figure P8.5a.

a. Using the following rules, construct three matrix relations using the input (Figure P8.5b) and output (Figure P8.5c) partitions:

Rule 1: IF $x = \text{Z}$, THEN $y = \text{S}$
Rule 2: IF $x = \text{NS}$ THEN $y = \text{NM}$
Rule 3: IF $x = \text{NB}$, THEN $y = \text{NL}$
Rule 4: IF $x = \text{PS}$, THEN $y = \text{PM}$
Rule 5: IF $x = \text{PB}$, THEN $y = \text{PL}$

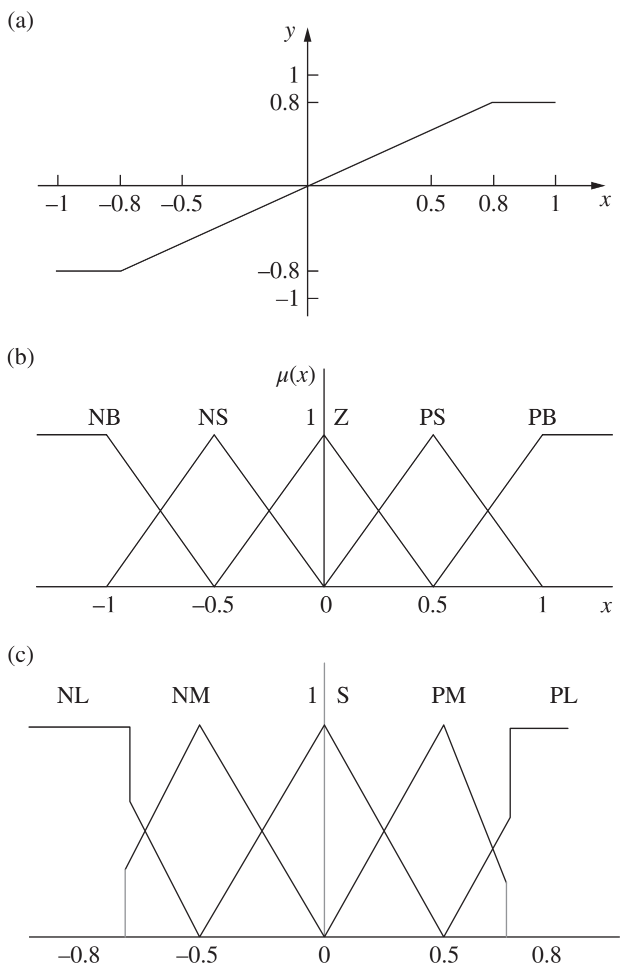

Figure P8.5a, b, c, d

(d)

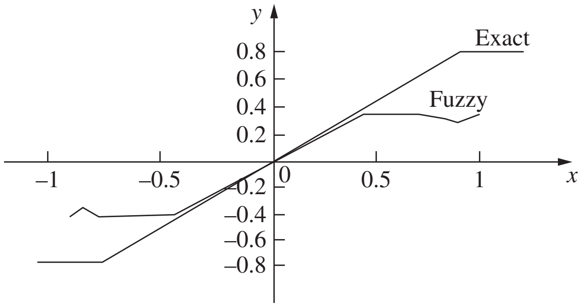

Figure P8.5a, b, c, d (Continued)

b. For crisp input values $x = -1, -0.8, -0.6, -0.4, -0.2$, and 0, use graphical techniques or max–min composition and centroid defuzzification to determine the associated fuzzy outputs. Because of symmetry, values for $0 \le x \le 1$ are equal to $|x|$ for $-1 \le x \le 0$. Verify that these results follow Figure P8.5d.

8.6 Repeat the example problem on the sine curve, Example 8.2, using (a) six rules and (b) eight rules. Does your result look more, or less, like a sine curve than the result in Example 8.2?

8.7 For the nonlinear function, $y = x^3$ develop a fuzzy rule-based system using four simple rules to approximate the output y. To develop the system, partition the range of the input x, [−1, 1], into five triangular MFs, and partition the output range of y, [−5, 5], into three triangular MFs. Use input labels, negative-big, negative-small, zero, positive-small and positive-big. Use output labels, negative-big, zero, and positive-big. Use the following rules for the simulation:

Rule 1: IF x is zero, THEN y is zero.
Rule 2: IF x positive-big, THEN y is positive-big.
Rule 3: IF x is negative-small, THEN y is negative-small.
Conduct a simulation for the inputs: [–2, −1.5, –0.5, 0.5, 1.5, 2.0].

8.8 For the nonlinear function, $y = x^2/2$ develop a fuzzy rule-based system using four simple rules to approximate the output y. To develop the system, partition the range of the input x, [–4, 4] into five triangular MFs, and partition the output range of y, [0, 8] into three triangular MFs. Use input labels, negative-big (NB), negative-small (NS), zero (Z), positive-small (PS), and positive-big (PB). Use output labels zero (Z), positive-small (PS), and positive-big (PB). Use the following rules for the simulation:

Rule 1: IF x is Z, THEN y is Z.
Rule 2: IF x is NB or PB, THEN y is PB.
Rule 3: IF x is NS or PS, THEN y is PS.

Conduct a simulation for the inputs: [–3, −1, 0, 1, 3].

9

Decision Making with Fuzzy Information

To be, or not to be: that is the question: Whether 'tis nobler in the mind to suffer The slings and arrows of outrageous fortune, Or to take arms against a sea of troubles, And by opposing end them.

William Shakespeare, *Hamlet*, Act III, Scene I, 1602

The passage from Shakespeare represents a classic decision situation for humans. It is expressed in natural language—the form of information most used by humans and most ignored in computer-assisted decision making. But as suggested many other times in this text, this is the nature of the problem engineers face every day: how do we embed natural fuzziness into our otherwise crisp engineering paradigms? Shakespeare would undoubtedly rejoice to learn that his question now has a whole range of possibilities available between the extremes of existence that he originally suggested. Ultimately, the decisions may be binary, as originally posed by Shakespeare in this passage from *Hamlet*, but there should certainly be no restrictions on the usefulness of fuzzy information in the *process* of making a decision or of coming to some consensus.

Decision making is a most important scientific, social, and economic endeavor. To be able to make consistent and correct choices is the essence of any decision process imbued with uncertainty. Most issues in life, as trivial as we might consider them, involve decision processes of one form or another. From the moment we wake in the morning to the time we place our bodies at rest at the day's conclusion we make many, many decisions. What should we wear for the day; should we take an umbrella; what should we eat for breakfast, for lunch, for dinner; should we stop by the gas station on the way to work; what route should we take to work; should we attend that seminar at work; should we write the memorandum to our colleagues before we make the reservations for our next trip out of town; should we go to the store on our way home;

should we take the kids to that new museum before, or after, dinner; should we watch the evening news before retiring; and so on and so forth?

We must keep in mind when dealing with decision making under uncertainty that *there is a distinct difference between a good decision and a good outcome*! In any decision process, we weigh the information about an issue or outcome and choose among two or more alternatives for subsequent action. The information affecting the issue is likely incomplete or uncertain; hence, the outcomes are uncertain, irrespective of the decision made or the alternative chosen. We can make a good decision, and the outcome can be adverse. Alternatively, we can make a bad decision, and the outcome can be advantageous. Such are the vagaries of uncertain events. But in the long run, if we consistently make good decisions, advantageous situations will occur more frequently than bad ones.

To illustrate this notion, consider the choice of whether to take an umbrella on a cloudy, dark morning. As a simple binary matter, the outcomes can be rain or no rain. We have two alternatives: take an umbrella or do not. The information we consider in making this decision could be as unsophisticated as our own feelings about the weather on similar days in the past or as sophisticated as a large-scale meteorological analysis from the national weather service. Whatever the source of information, it will be associated with some degree of uncertainty. Suppose we decide to take the umbrella after weighing all the information, and it does not rain. Did we make a bad decision? Perhaps not. Eight times out of 10 in circumstances just like this one, it probably rained. This particular occasion may have been one of the 2 out of 10 situations when it did not.

Despite our formal training in this area and despite our common sense about how clear this notion of uncertainty is, we see it violated every day in the business world. A manager makes a good decision, but the outcome is bad and the manager gets fired. A doctor uses the best established procedures in a medical operation and the patient dies; then the doctor gets sued for malpractice. A boy refuses to accept an unsolicited ride home with a distant neighbor on an inclement day, gets soaking wet on the walk home, ruins his shoes, and is reprimanded by his parent for not accepting the ride. A teenager decides to drive on the highway after consuming too many drinks and arrives home safely without incident. In all of these situations, the outcomes have nothing to do with the quality of the decisions or with the process itself. The best we can do is to make consistently rational decisions every time we are faced with a choice with the knowledge that in the long run the "good" will outweigh the "bad."

The problem in making decisions under uncertainty is that the bulk of the information we have about the possible outcomes, about the value of new information, about the way the conditions change with time (dynamic), about the utility of each outcome–action pair, and about our preferences for each action is typically vague, ambiguous, and otherwise fuzzy. In some situations, the information may be robust enough so that we can characterize it with probability theory.

In making informed and rational decisions, we must remember that individuals are essentially risk averse. When the consequences of an action might result in serious injury, death, economic ruin, or some other dreaded event, humans do not make decisions consistent with rational utility theory (Maes and Faber, 2004). In fact, studies in cognitive psychology show that rationality is a rather weak hypothesis in decision making, easily refuted, and therefore not always useful as an axiomatic explanation of the theory of decision making. Human risk preference in the face of high uncertainty is not easily modeled by its rational methods. In a narrow context of decision making, rational behavior is defined in terms of decision making, which maximizes expected utility (von Neumann and Morgenstern, 1944). Of course, this utility is a function of personal preferences of the decision maker. Although much of the literature addresses decision making in the face of economic and financial risks, engineers are primarily concerned with two types of decisions (Maes and Faber, 2004): operational decisions, where for certain available resources an optimal action is sought to avoid a specific set of hazards, and strategic decisions, which involve decisions regarding one's level of preparedness or anticipation of events in the future. Difficulties in human preference reversal, in using incomplete information, in bias toward one's own experience, and in using epistemic uncertainty (e.g., ambiguity, vagueness, and fuzziness) are among the various issues cited by Maes and Faber (2004) as reasons why the independence axiom of an expected utility analysis (used in rational decision making) is violated by human behavior. Although we do not address these matters in this text, it is nonetheless important to keep them in mind when using any of the methods developed here.

This chapter presents a few paradigms for making decisions within a fuzzy environment. Issues such as personal preferences, multiple objectives, nontransitive reasoning, and group consensus are presented. The chapter concludes with a rather lengthy development of an area known loosely as *fuzzy Bayesian decision making*, so named because it involves the introduction of fuzzy information, fuzzy outcomes, and fuzzy actions into the classical probabilistic method of Bayesian decision making. In developing this, we are able to compare the value and differences of incorporating both fuzzy and random information into the same representational framework. Acceptance of the fuzzy approach is therefore eased by its natural accommodation within a classical, and historically popular, decision-making approach. Moreover, a recent book chapter (Ross, Booker, and Parkinson, 2003) shows how the likelihood function in Bayes's rule has similar properties to a fuzzy membership function. This is not to suggest that Bayesian decision making is accepted universally; Maes and Faber (2004) highlight some problems with Bayesian updating of probabilities and utilities that are currently being debated in the literature.

Fuzzy Synthetic Evaluation

The term *synthetic* is used here to connote the process of evaluation whereby several individual elements and components of an evaluation are synthesized into an aggregate form; the whole is a *synthesis* of the parts. The key here is that the various elements can be numeric or nonnumeric, and the process of fuzzy synthesis is naturally accommodated using synthetic evaluation. In reality, an evaluation of an object, especially an ill-defined one, is often vague and ambiguous. The evaluation is usually described in natural language terms because a numerical evaluation is often too complex, too unacceptable, and too ephemeral (transient). For example, when grading a written examination, the professor might evaluate it from such perspectives as style, grammar, creativity, and so forth. The final grade on the paper might be linguistic instead of numeric, for example, excellent, very good, good, fair, poor, or unsatisfactory. After grading many exams, the professor might develop a relation by which a membership is assigned to the relations between the different perspectives, such as style and grammar, and the linguistic grades, such as fair and excellent. A fuzzy relation, $\underset{\sim}{R}$, such as the following one, might result that summarizes the professor's relationship between pairs of grading factors such as *creativity* and grade evaluations such as *very good*:

$$\underset{\sim}{R} = \begin{array}{c} \\ \text{Creativity} \\ \text{Grammer} \\ \text{Style} \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Excellent} & \text{Very good} & \text{Good} & \text{Fair} & \text{Poor} \end{array} \\ \begin{bmatrix} 0.2 & 0.4 & 0.3 & 0.1 & 0 \\ 0 & 0.2 & 0.5 & 0.3 & 0 \\ 0.1 & 0.6 & 0.3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array}.$$

The professor now wants to assign a grade to each paper. To formalize this approach, let X be a universe of factors and Y be a universe of evaluations, so

$$X = \{x_1,\ x_2, \ldots,\ x_n\} \text{ and } Y = \{y_1,\ y_2, \ldots, y_m\}.$$

Let $\underset{\sim}{R} = [r_{ij}]$ be a fuzzy relation, such as the foregoing grading example, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Suppose we introduce a specific paper into the evaluation process on which the professor has given a set of "scores" (w_i) for each of the n grading factors, and we ensure, for convention, that the sum of the scores is unity. Each of these scores is actually a membership value for each of the factors, x_i, and they can be arranged in a fuzzy vector, $\underset{\sim}{\mathbf{w}}$. So we have

$$\underset{\sim}{\mathbf{w}} = \{w_1, w_2, \ldots w_n\}, \text{ where } \sum_i w_i = 1. \tag{9.1a}$$

The process of determining a grade for a specific paper is equivalent to the process of determining a membership value for the paper in each of the evaluation categories, y_i. This process is implemented through the composition operation

$$\underset{\sim}{\mathbf{e}} = \underset{\sim}{\mathbf{w}} \circ \underset{\sim}{R}, \tag{9.1b}$$

where $\underset{\sim}{\mathbf{e}}$ is a fuzzy vector containing the membership values for the paper in each of the y_i evaluation categories.

Example 9.1

Suppose we want to measure the value of a microprocessor to a potential client. In conducting this evaluation, the client suggests that certain criteria are important. They can include performance (MIPS), cost ($), availability (AV), and software (SW). Performance is measured by millions of instructions per second (MIPS); a minimum requirement is 10 MIPS. Cost is the cost of the microprocessor, and a cost requirement of "not to exceed" 500 has been set. Availability relates to how much time after the placement of an order the microprocessor vendor can deliver the part; a maximum of eight weeks has been set. Software represents the availability of operating systems, languages, compilers, and tools to be used with this microprocessor.

Suppose further that the client is only able to specify a subjective criterion of having "sufficient" software.

A particular microprocessor (CPU) has been introduced into the market. It is measured against these criteria and given ratings categorized as excellent (e), superior (s), adequate (a), and inferior (i). "Excellent" means that the microprocessor is the best available with respect to the particular criterion. "Superior" means that the microprocessor is among the best with respect to this criterion. "Adequate" means that, although not superior, the microprocessor can meet the minimum acceptable requirements for this criterion. "Inferior" means that the microprocessor cannot meet the requirements for the particular criterion. Suppose the microprocessor just introduced has been assigned the following relation based on the consensus of the design team:

$$\underset{\sim}{\mathrm{R}} = \begin{array}{c} \\ \text{MIPS} \\ \$ \\ \text{AV} \\ \text{SW} \end{array} \begin{array}{c} \begin{array}{cccc} e & s & a & i \end{array} \\ \begin{bmatrix} 0.1 & 0.3 & 0.4 & 0.2 \\ 0 & 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.2 & 0.1 \\ 0.1 & 0.4 & 0.3 & 0.2 \end{bmatrix} \end{array}.$$

This relation could have been derived from data using similarity methods such as those discussed in Chapter 3.

If the evaluation team applies a scoring factor of 0.4 for performance, 0.3 for cost, 0.2 for availability, and 0.1 for software, which together form the factor vector, $\underset{\sim}{\mathbf{w}}$, then the composition, $\underset{\sim}{\mathbf{e}} = \underset{\sim}{\mathbf{w}} \circ \underset{\sim}{\mathrm{R}} = \{0.1,\ 0.3,\ 0.4,\ 0.2\}$, results in an evaluation vector that has its highest membership in the category "adequate.

It is important to point out in concluding this section that the relations expressed in this section are not constrained in that their row sums should equal unity. The examples given show the row sums equaling unity, a matter of convenience for illustration. However, because the entries in the synthetic evaluation matrix relations are membership values showing the degree of relation between the factors and the evaluations, these values can take on any number between 0 and 1. Hence, row sums could be larger, or smaller, than unity.

Fuzzy Ordering

Decisions are sometimes made on the basis of rank, or ordinal ranking: which issue is best, which is second best, and so forth. For issues or actions that are deterministic, such as $y_1 = 5, y_2 = 2, y_1 \geq y_2$, there is usually no ambiguity in the ranking; we might call this *crisp ordering*. In situations where the issues or actions are associated with uncertainty, either random or fuzzy, rank ordering may be ambiguous. This ambiguity, or uncertainty, can be demonstrated for both random and fuzzy variables. First, let us assume that the uncertainty in rank is random; we can use probability density functions (pdf) to illustrate the random case. Suppose we have one random variable, x_1, whose uncertainty is characterized by a Gaussian pdf with a mean of μ_1 and a standard deviation of σ_1, and another random variable, x_2, also Gaussian with a mean of μ_2 and standard deviation of σ_2. Suppose further that $\sigma_1 > \sigma_2$ and $\mu_1 > \mu_2$. If we plot the pdfs for these two random variables in Figure 9.1, we see that the question of which variable is greater is not clear.

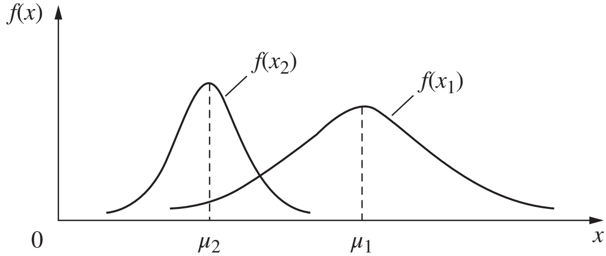

Figure 9.1 Density functions for two Gaussian random variables.

As an example of this uncertain ranking, suppose x_1 is the height of Italians and x_2 is the height of Swedes. Because this uncertainty is of the random kind, we cannot answer the question "Are Swedes taller than Italians?" unless we are dealing with two specific individuals, one each from Sweden and Italy, or we are simply assessing μ_1, average-height Swedes, and μ_2, average-height Italians. But we can ask the question, "How frequently are Swedes taller than Italians?" We can assess this frequency as the probability that one random variable is greater than another, that is, $P(x_1 \geq x_2)$, with

$$P(x_1 \geq x_2) = \int_{-\infty}^{\infty} F_{x_2}(x_1)\mathrm{d}x_1, \tag{9.2a}$$

where F is a cumulative distribution function. Hence, with random variables we can quantify the uncertainty in ordering with a convolution integral, Equation (9.2a).

Second, let us assume that the uncertainty in rank arises because of ambiguity. For example, suppose we are trying to rank people's preferences in colors. In this case the ranking is subjective and not reducible to the elegant form available for some random variables, such as that given in Equation (9.2a). For fuzzy variables we are also able to quantify the uncertainty in ordering, but in this case we must do so with the notion of membership.

A third type of ranking involves the notion of imprecision (Dubois and Prade, 1980). To develop this, suppose we have two fuzzy numbers, $\underset{\sim}{\mathrm{I}}$ and $\underset{\sim}{\mathrm{J}}$. We can use tools provided in Chapter 12 on the extension principle to calculate the truth value of the assertion that fuzzy number $\underset{\sim}{\mathrm{I}}$ is greater than fuzzy number $\underset{\sim}{\mathrm{J}}$ (a fuzzy number is defined in Chapter 4):

$$T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{J}}) = \sup_{x \geq y} \min(\mu_{\underset{\sim}{\mathrm{I}}}(x),\ \mu_{\underset{\sim}{\mathrm{J}}}(y)). \tag{9.2b}$$

Figure 9.2 shows the membership functions for two fuzzy numbers $\underset{\sim}{\mathrm{I}}$ and $\underset{\sim}{\mathrm{J}}$. Equation (9.2b) is an extension of the inequality $x \geq y$ according to the extension principle. It represents the degree of possibility in the sense that if a specific pair (x, y) exists such that $x \geq y$ and $\mu_{\underset{\sim}{\mathrm{I}}}(x) = \mu_{\underset{\sim}{\mathrm{J}}}(y)$, then $T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{J}}) = 1$. Because the fuzzy numbers $\underset{\sim}{\mathrm{I}}$ and $\underset{\sim}{\mathrm{J}}$ are convex, it can be seen from Figure 9.2 that

$$T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{J}}) = 1 \text{ if and only if } \mathrm{I} \geq \mathrm{J}, \tag{9.3a}$$

$$T(\underset{\sim}{\mathrm{J}} \geq \underset{\sim}{\mathrm{I}}) = \text{height}(\underset{\sim}{\mathrm{I}} \cap \underset{\sim}{\mathrm{J}}) = \mu_{\underset{\sim}{\mathrm{I}}}(d) = \mu_{\underset{\sim}{\mathrm{J}}}(d), \tag{9.3b}$$

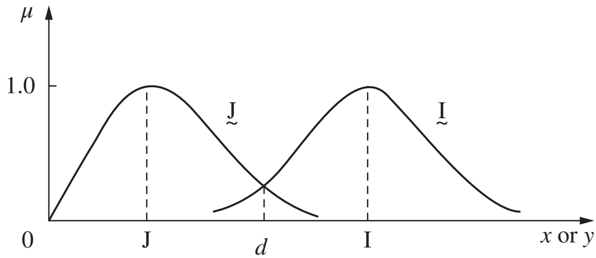

Figure 9.2 Two fuzzy numbers as fuzzy sets on the real line.

where d is the location of the highest intersection point of the two fuzzy numbers. The operation height $(\underset{\sim}{\mathrm{I}} \cap \underset{\sim}{\mathrm{J}})$ in Equation (9.3b) is a good separation metric for two fuzzy numbers; that is, the closer this metric is to unity, the more difficult it is to distinguish which of the two fuzzy numbers is larger. On the other hand, as this metric approaches zero, the easier is the distinction about rank order (which is largest). Unfortunately, the metric given in Equation (9.3a) is not too useful as a ranking metric, because $T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{J}}) = 1$ when I is slightly greater and when I is much greater than J. If we know that $\underset{\sim}{\mathrm{I}}$ and $\underset{\sim}{\mathrm{J}}$ are crisp numbers I and J, the truth value becomes T (I ≥ J) = 1 for I ≥ J and T (I ≥ J) = 0 for I < J.

The definitions expressed in Equations (9.2b) and (9.3) for two fuzzy numbers can be extended to the more general case of many fuzzy sets. Suppose we have k fuzzy sets $\underset{\sim}{\mathrm{I}}_1, \underset{\sim}{\mathrm{I}}_2, \ldots, \underset{\sim}{\mathrm{I}}_k$. Then, the truth value of a specified ordinal ranking is given as

$$T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{I}}_1, \underset{\sim}{\mathrm{I}}_2, \ldots, \underset{\sim}{\mathrm{I}}_k) = T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{I}}_1) \text{ and } T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{I}}_2) \text{ and} \ldots \text{and } T(\underset{\sim}{\mathrm{I}} \geq \underset{\sim}{\mathrm{I}}_k). \tag{9.4}$$

Example 9.2

Suppose we have three fuzzy sets, as described using Zadeh's notation:

$$\underset{\sim}{\mathrm{I}}_1 = \left\{\frac{1}{3} + \frac{0.8}{7}\right\}, \ \underset{\sim}{\mathrm{I}}_2 = \left\{\frac{0.7}{4} + \frac{1.0}{6}\right\}, \text{ and } \underset{\sim}{\mathrm{I}}_3 = \left\{\frac{0.8}{2} + \frac{1}{4} + \frac{0.5}{8}\right\}.$$

We can assess the truth value of the inequality, $\underset{\sim}{\mathrm{I}}_1 \geq \underset{\sim}{\mathrm{I}}_2$, as follows:

$$\begin{aligned} T(\underset{\sim}{\mathrm{I}}_1 \geq \underset{\sim}{\mathrm{I}}_2) &= \max_{x_1 \geq x_2}\left\{\min\left(\mu_{\underset{\sim}{\mathrm{I}}_1}(x_1), \mu_{\underset{\sim}{\mathrm{I}}_2}(x_2)\right)\right\} \\ &= \max\left\{\min\left(\mu_{\underset{\sim}{\mathrm{I}}_1}(7),\ \mu_{\underset{\sim}{\mathrm{I}}_2}(4)\right),\ \min\left(\mu_{\underset{\sim}{\mathrm{I}}_1}(7),\ \mu_{\underset{\sim}{\mathrm{I}}_2}(6)\right)\right\} \\ &= \max\{\min(0.8,\ 0.7),\ \min(0.8,\ 1.0)\} \\ &= 0.8. \end{aligned}$$

Similarly,

$$T\left(\underset{\sim}{I}_1 \geq \underset{\sim}{I}_3\right)=0.8, \quad T\left(\underset{\sim}{I}_2 \geq \underset{\sim}{I}_1\right)=1.0,$$
$$T\left(\underset{\sim}{I}_2 \geq \underset{\sim}{I}_3\right)=1.0, \quad T\left(\underset{\sim}{I}_3 \geq \underset{\sim}{I}_1\right)=1.0,$$
$$T\left(\underset{\sim}{I}_3 \geq \underset{\sim}{I}_2\right)=0.7.$$

Then,

$$T\left(\underset{\sim}{I}_1 \geq \underset{\sim}{I}_2, \underset{\sim}{I}_3\right)=0.8,$$
$$T\left(\underset{\sim}{I}_2 \geq \underset{\sim}{I}_1, \underset{\sim}{I}_3\right)=1.0,$$
$$T\left(\underset{\sim}{I}_3 \geq \underset{\sim}{I}_1, \underset{\sim}{I}_2\right)=0.7.$$

The last three truth values in this example compared one fuzzy set to two others. This calculation is different from pairwise comparisons. To do the former, one makes use of the minimum function, as prescribed by Equation (9.4). For example,

$$T\left(\underset{\sim}{I}_1 \geq \underset{\sim}{I}_2, \underset{\sim}{I}_3\right)=\min\left\{\left(\underset{\sim}{I}_1 \geq \underset{\sim}{I}_2\right),\left(\underset{\sim}{I}_1 \geq \underset{\sim}{I}_3\right)\right\}.$$

Equation (9.4) can be used similarly to obtain the other multiple comparisons. Based on the foregoing ordering, the overall ordering for the three fuzzy sets would be $\underset{\sim}{I}_2$ first, $\underset{\sim}{I}_1$ second, and $\underset{\sim}{I}_3$ last.

Another procedure to compare two fuzzy numbers and find the larger or smaller of them is given by Klir and Yuan (1995). In this case, we make use of what is known as the *extended MIN and MAX operations*; these are operations on fuzzy numbers, whereas the standard min and max are operations on real numbers. The reader is referred to the literature for more discussion on this.

Nontransitive Ranking

When we compare objects that are fuzzy, ambiguous, or vague, we may well encounter a situation where there is a contradiction in the classical notions of ordinal ranking and transitivity in the ranking. For example, suppose we are ordering on the preference of colors. When comparing red to blue, we prefer red; when comparing blue to yellow, we prefer blue; but when comparing red and yellow we might prefer yellow. In this case, transitivity of sets representing preference in colors (red > blue and blue > yellow does *not* yield red > yellow) is not maintained.

To accommodate this form of nontransitive ranking (which can be quite normal for noncardinal-type concepts), we introduce a special notion of relativity (Shimura, 1973). Let x and y be variables defined on universe X. We define a pairwise function

$$f_y(x) \text{ as the membership value of } x \text{ with respect to } y,$$

and we define another pairwise function

$$f_x(y) \text{ as the membership value of } y \text{ with respect to } x.$$

Then, the relativity function given as

$$f(x|y) = \frac{f_y(x)}{\max\left[f_y(x), f_x(y)\right]} \tag{9.5}$$

is a measurement of the membership value of choosing x over y. The relativity function $f(x \mid y)$ can be thought of as the membership of preferring variable x over variable y. Note that the function in Equation (9.5) uses arithmetic division.

To develop the general case of Equation (9.5) for many variables, define variables $x_1, x_2, \ldots, x_i, x_{i+1}, \ldots, x_n$ all defined on universe X, and let these variables be collected in a set A, that is, $A = \{x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n\}$. We then define a set identical to set A except this new set will be missing one element, x_i, and this set will be termed A′. The relativity function then becomes

$$\begin{aligned} f(x_i|A') &= f(x_i|\{x_1,\ x_2, \ldots, x_{i-1},\ x_{i+1}, \ldots, x_n\}) \\ &= \min\{f(x_i|x_1), f(x_i|x_2), \ldots, f(x_i|x_{i-1}),\ f(x_i|x_{i+1}), \ldots, f(x_i|x_n)\}, \end{aligned} \tag{9.6}$$

which is the fuzzy measurement of choosing x_i over all elements in the set A′. The expression in Equation (9.6) involves the logical intersection of several variables; hence, the minimum function is used. Because the relativity function of one variable with respect to itself is identity, that is,

$$f(x_i|x_i) = 1, \tag{9.7}$$

then

$$f(x_i|A') = f(x_i|A) \tag{9.8}$$

We can now form a matrix of relativity values, $f(x_i \mid x_j)$, where $i, j = 1, 2, \ldots, n$, and where x_i and x_j are defined on a universe X. This matrix will be square and of order n, and will be termed the C matrix (C for comparison). The C matrix can be used to rank many different fuzzy sets.

To determine the overall ranking, we need to find the smallest value in each of the rows of the C matrix; that is,

$$C_i' = \min f(x_i|X), i = 1, 2, \ldots, n, \tag{9.9}$$

where C_i' is the membership ranking value for the ith variable. We use the minimum function because this value will have the lowest weight for ranking purposes; further, the maximum function often returns a value of 1 and there would be ambivalence in the ranking process (i.e., ties will result).

Example 9.3

In manufacturing, we often try to compare the capabilities of various microprocessors for their appropriateness to certain applications. For instance, suppose we are trying to select from among four microprocessors the one that is best suited for image-processing applications. Because many factors, including performance, cost, availability, and software, can affect this decision, coming up with a crisp mathematical model for all these attributes is complicated. Another consideration is that it is much easier to compare these microprocessors subjectively in pairs rather than all four at one time.

Suppose the design team is polled to determine which of the four microprocessors, labeled x_1, x_2, x_3, and x_4, is the most preferred when considered as a group rather than when considered as pairs. First, pairwise membership functions are determined. These represent the subjective measurement of the appropriateness of each microprocessor when compared only to one another. The following pairwise functions are determined:

$$\begin{array}{llll} f_{x_1}(x_1)=1, & f_{x_1}(x_2)=0.5, & f_{x_1}(x_3)=0.3, & f_{x_1}(x_4)=0.2. \\ f_{x_2}(x_1)=0.7, & f_{x_2}(x_2)=1, & f_{x_2}(x_3)=0.8, & f_{x_2}(x_4)=0.9. \\ f_{x_3}(x_1)=0.5, & f_{x_3}(x_2)=0.3, & f_{x_3}(x_3)=1, & f_{x_3}(x_4)=0.7. \\ f_{x_4}(x_1)=0.3, & f_{x_4}(x_2)=0.1, & f_{x_4}(x_3)=0.3, & f_{x_4}(x_4)=1. \end{array}$$

For example, microprocessor x_2 has membership 0.5 with respect to microprocessor x_1. Note that if these values were arranged into a matrix, it would not be symmetric. These membership values do not express similarity or relation; they represent membership values of ordering when considered in a particular order. If we now employ Equation (9.5) to calculate all of the relativity values, the matrix shown expresses these calculations; this is the so-called C matrix. In the notation for the relativity values, we have $f(x_i \mid x_j) = f$ (ith row | jth column). For example,

$$f(x_2|x_1) = \frac{f_{x_1}(x_2)}{\max[f_{x_1}(x_2), f_{x_2}(x_1)]} = \frac{0.5}{\max[0.5, 0.7]} = 0.71.$$

$$\begin{array}{cc} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \qquad \min = f(x_i|X) \\ \mathrm{C} = \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \end{array} & \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.71 & 1 & 0.38 & 0.11 \\ 0.6 & 1 & 1 & 0.43 \\ 0.67 & 1 & 1 & 1 \end{bmatrix} \qquad \begin{array}{c} 1 \\ 0.11 \\ 0.43 \\ 0.67 \end{array} \end{array}.$$

The extra column to the right of the foregoing C matrix is the minimum value for each of the rows, that is, for C_i', $i = 1, 2, 3, 4$, in Equation (9.9). For this example problem, the order from best to worst is x_1, x_4, x_3, and x_2. This ranking is much more easily attained with this fuzzy approach than it would have been with some other method where the attributes of each microprocessor are assigned a value measurement and these values are somehow combined. This fuzzy method also contains the subjectiveness inherent in comparing one microprocessor to another. If all four are considered at once, a person's own bias might skew the value assessments to favor a particular microprocessor. By using pairwise comparisons, each microprocessor is compared individually against its peers, which should allow a more fair and less biased comparison.

Preference and Consensus

The goal of group decision making typically is to arrive at a consensus concerning a desired action or alternative from among those considered in the decision process. In this context, consensus is usually taken to mean a unanimous agreement by all those in the group concerning their choice. Despite the simplicity in defining consensus, it is another matter altogether to quantify this notion. Most traditional mathematical developments of consensus have used individual preference ranking as their primary feature. In these developments, the individual preferences of those in the decision group are collected to form a group metric whose properties are used to produce a scalar measure of "degree of consensus." However, the underlying axiomatic structure of many of these classical approaches is based on classical set theory. The argument given in this text is that the crisp set approach is too restrictive for the variables relevant to a group decision process. The information in the previous section showed individual preference to be a fuzzy relation.

There can be numerous outcomes of decision groups in developing consensus about a universe, X, of n possible alternatives, that is, $X = \{x_1, x_2, \ldots, x_n\}$. To start the development, we define a *reciprocal* relation as a fuzzy relation, $\underset{\sim}{R}$, of order n, whose individual elements r_{ij} have the following properties (Bezdek, Spillman, and Spillman, 1978):

$$r_{ii} = 0, \quad \text{for } 1 \leq i \leq n. \tag{9.10}$$

$$r_{ij} + r_{ji} = 1, \quad \text{for } i \neq j. \tag{9.11}$$

This reciprocal relation, $\underset{\sim}{R}$, Equations (9.10) and (9.11), can be interpreted as follows: r_{ij} is the preference accorded to x_i relative to x_j. Thus, $r_{ij} = 1$ (hence, $r_{ji} = 0$) implies that alternative i is *definitely* preferred to alternative j; this is the crisp case in preference. At the other extreme, we have maximal fuzziness, where $r_{ij} = r_{ji} = 0.5$, and there is equal preference, pairwise. A definite choice of alternative i to all others is manifested in $\underset{\sim}{R}$, as the ith row being all ones (except $r_{ii} = 0$), or the ith column being all zeros.

Two common measures of preference are defined here as *average fuzziness* in $\underset{\sim}{R}$, Equation (9.12), and *average certainty* in $\underset{\sim}{R}$, Equation (9.13):

$$F(\underset{\sim}{R}) = \frac{\mathrm{tr}(\underset{\sim}{R}^2)}{n(n-1)/2}. \tag{9.12}$$

$$C(\underset{\sim}{R}) = \frac{\mathrm{tr}\left(\underset{\sim}{R}\underset{\sim}{R}^{T}\right)}{n(n-1)/2}. \tag{9.13}$$

In Equations (9.12) and (9.13), tr() and $()^T$ denote the trace and transpose, respectively, and matrix multiplication is the algebraic kind. Recall that the trace of a matrix is simply the algebraic sum of the diagonal elements, that is,

$$\mathrm{tr}(\underset{\sim}{R}) = \sum_{i=1}^{n} r_{ii}.$$

The measure, $F(\underset{\sim}{R})$, averages the joint preferences in $\underset{\sim}{R}$ over all distinct pairs in the Cartesian space, X × X. Each term maximizes the measure when $r_{ij} = r_{ji} = 0.5$ and minimizes the measure when $r_{ij} = 1$ and $r_{ji} = 0$; consequently, $F(\underset{\sim}{R})$ is proportional to the fuzziness or uncertainty (also, confusion) about pairwise rankings exhibited by the fuzzy preference relation, $\underset{\sim}{R}$. Conversely, the measure, $C(\underset{\sim}{R})$, averages the individual dominance (assertiveness) of each distinct pair of rankings in the sense that each term maximizes the measure when $r_{ij} = 1$ and $r_{ji} = 0$ and minimizes the measure when $r_{ij} = r_{ji} = 0.5$; hence, $C(\underset{\sim}{R})$ is proportional to the overall certainty in $\underset{\sim}{R}$. The two measures are dependent; they are both on the interval [0, 1]; and it can be shown (Bezdek, et al., 1978) that

$$F(\underset{\sim}{R}) + C(\underset{\sim}{R}) = 1. \tag{9.14}$$

It can further be shown that C is a minimum and F is a maximum at $r_{ij} = r_{ji} = 0.5$, and that C is a maximum and F is a minimum at $r_{ij} = 1$, $r_{ji} = 0$. Also, at the state of maximum fuzziness ($r_{ij} = r_{ji} = 0.5$) we get $F(\underset{\sim}{R}) = C(\underset{\sim}{R}) = \frac{1}{2}$, and at the state of no uncertainty ($r_{ij} = 1$, $r_{ji} = 0$) we get $F(\underset{\sim}{R}) = 0$ and $C(\underset{\sim}{R}) = 1$. Moreover, the ranges for these two measures are $0 \le F(\underset{\sim}{R}) \le \frac{1}{2}$ and $\frac{1}{2} \le C(\underset{\sim}{R}) \le 1$.

Measures of preference can be useful in determining consensus. There are different forms of consensus. We have discussed the antithesis of consensus: complete ambivalence, or the maximally fuzzy case where all alternatives are rated equally; we call this type of consensus M_1. For M_1, we have a matrix $\underset{\sim}{R}$ where all nondiagonal elements are equal to $\frac{1}{2}$. We have also discussed the converse of M_1, which is the nonfuzzy (crisp) preference where every pair of alternatives is definitely ranked; we call this case M_2. In M_2, all nondiagonal elements in $\underset{\sim}{R}$ are equal to 1 or 0; however, there may not be a clear consensus. Consider the following reciprocity relation, M_2:

$$M_2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Here, the clear pairwise choices are as follows: alternative 1 over alternative 2, alternative 1 over alternative 4, alternative 2 over alternative 3, and alternative 3 over alternative 4. However, we do not have consensus because alternative 3 is preferred over alternative 1 and alternative 4 is preferred over alternative 2! So, for relation M_1, we cannot have consensus and for relation M_2 we may not have consensus.

Three types of consensus, however, arise from considerations of the matrix $\underset{\sim}{R}$. The first type, known as *Type I consensus*, M_1^*, is a consensus in which there is one clear choice, say alternative i (the ith column is all zeros), and the remaining $(n-1)$ alternatives all have equal secondary preference (i.e., $r_{kj} = \frac{1}{2}$, where $k \neq j$), as shown in the following example, where alternative 2 has a clear consensus:

$$M_1^* = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 \\ 1 & 0 & 1 & 1 \\ 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix}.$$

In the second type of consensus, called a *Type II consensus*, M_2^*, there is one clear choice, say alternative i (the ith column is all zeros), but the remaining $(n-1)$ alternatives all have definite secondary preference (i.e., $r_{kj} = 1$, where $k \neq i$), as shown in this example:

$$M_2^* = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

where alternative 2 has a clear consensus, but where there is no clear ordering after the first choice because alternative 1 is preferred to alternative 3, 3 to 4, but alternative 4 is preferred to alternative 1. There can be clear ordering after the first choice in Type II consensus matrices, but it is not a requirement.

Finally, the third type of consensus, called a Type *fuzzy* consensus, M_f^*, occurs where there is a unanimous decision for the most preferred choice, say alternative i again, but the remaining $(n-1)$ alternatives have infinitely many fuzzy secondary preferences. The matrix shown here has a clear choice for alternative 2, but the other secondary preferences are fuzzy to various degrees:

$$M_f^* = \begin{bmatrix} 0 & 0 & 0.5 & 0.6 \\ 1 & 0 & 1 & 1 \\ 0.5 & 0 & 0 & 0.3 \\ 0.4 & 0 & 0.7 & 0 \end{bmatrix}.$$

Mathematically, relations M_1 and M_2 are logical opposites, as are consensus relations M_1^* and M_2^* (Bezdek, et al., 1978). It is interesting to discuss the cardinality of these various preference and consensus relations. In this case, the cardinality of a relation is the number of possible combinations of that matrix type. It is obvious that there is only one possible matrix for M_1. The cardinality of all the preference or consensus relations discussed here is given in Equation (9.15), where the symbol || denotes cardinality of the particular relation:

$$\begin{aligned}
&|M_1| = 1 \\
&|M_2| = 2^{n(n-1)/2} \\
&|M_1^*| = n && \text{(Type I)} \\
&|M_2^*| = \left(2^{(n^2-3n+2)/2}\right)(n) && \text{(Type II)} \\
&|M_f^*| = \infty && (\text{Type}\,\textit{fuzzy})
\end{aligned} \quad . \tag{9.15}$$

So, for the examples previously illustrated for n = 4 alternatives, there are 64 (2^6) possible forms of the M_2 preference matrix, there are only four Type I (M_1^*) consensus matrices, and there are 32 ($2^3 \cdot 4$) possible Type II (M_2^*) consensus matrices.

From the *degree of preference* measures given in Equations (9.12) and (9.13), we can construct a *distance to consensus* metric, defined as

$$m(\underset{\sim}{R}) = 1 - \left(2C\left(\underset{\sim}{R}\right) - 1\right)^{1/2}, \tag{9.16}$$

where

$$\begin{aligned}
&m(\underset{\sim}{R}) = 1 \text{ for an } M_1 \text{ preference relation} \\
&m(\underset{\sim}{R}) = 0 \text{ for an } M_2 \text{ preference relation} \\
&m(\underset{\sim}{R}) = 1 - (2/n)^{1/2} \text{ for a Type I } \left(M_1^*\right) \text{ consensus relation} \\
&m(\underset{\sim}{R}) = 0 \text{ for a Type II } \left(M_2^*\right) \text{ consensus relation}
\end{aligned}$$

We can think of this metric, $m(\underset{\sim}{R})$, as being a distance between the points M_1 (1.0) and M_2 (0.0) in n-dimensional space. We see that $m\left(M_1^*\right) = m\left(M_2^*\right)$ for the case where we have only two ($n = 2$) alternatives. For the more general case, where $n > 2$, the distance between Type I and Type II consensus increases with n as it becomes increasingly difficult to develop a consensus choice and simultaneously rank the remaining pairs of alternatives.

Example 9.4

Suppose a reciprocal fuzzy relation, $\underset{\sim}{R}$, is developed by a small group of people in their initial assessments for pairwise preferences for a decision process involving four alternatives, $n = 4$, as shown here:

$$\underset{\sim}{R} = \begin{bmatrix} 0 & 1 & 0.5 & 0.2 \\ 0 & 0 & 0.3 & 0.9 \\ 0.5 & 0.7 & 0 & 0.6 \\ 0.8 & 0.1 & 0.4 & 0 \end{bmatrix}.$$

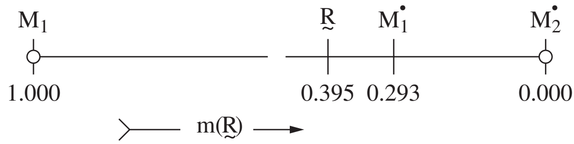

Figure 9.3 Illustration of *distance to consensus*. Adapted from Bezdek *et al*., 1978.

Notice that this matrix carries none of the properties of a consensus type; that is, the group does not reach consensus on their first attempt at ranking the alternatives. However, the group can assess their "degree of consensus" and they can measure how "far" they are from consensus prior to subsequent discussions in the decision process. So, for example, alternative 1 is *definitely* preferred to alternative 2 and alternative 1 is rated *equal* to alternative 3. For this matrix, $C(\underset{\sim}{R}) = 0.683$ (Equation (9.13)), $m(\underset{\sim}{R}) = 0.395$ (Equation (9.16)), and $m\left(M_1^*\right) = 1-(2/n)^{1/2} = 0.293$. For their first attempt at ranking the four alternatives, the group has an average certainty of 0.683 [recall a value of 0.5 is completely ambivalent (uncertain) and a value of 1.0 is completely certain]. Moreover, the group is $1 - 0.395 = 0.605$ or 60.5% of the way from complete ambivalence (M_1) toward a Type II consensus or they are $0.605/(1 - 0.293) = 85.5\%$ of the way toward a Type I consensus. These ideas are shown graphically in Figure 9.3. The value of the *distance to consensus*, $m(\underset{\sim}{R})$, is its use in quantifying the dynamic evolution of a group as the group refines its preferences and moves closer to a Type I or Type II consensus. It should be noted that the vast majority of group preference situations eventually develop into a *fuzzy* consensus; Types I and II are typically useful only as boundary conditions.

Multiobjective Decision Making

Many simple decision processes are based on a single objective, such as minimizing cost, maximizing profit, and minimizing run time. Often, however, decisions must be made in an environment where more than one objective constrains the problem, and the relative value of each of these objectives is different. For example, suppose we are designing a new computer, and simultaneously we want to minimize cost, maximize CPU, maximize random access memory (RAM), and maximize reliability. Moreover, suppose cost is the most important of our objectives and the other three (CPU, RAM, and reliability) carry lesser but equal weight when compared with cost. Two primary issues in multiobjective decision making are to acquire meaningful information regarding the satisfaction of the objectives by the various choices (alternatives) and to rank or weight the relative importance of each of the objectives. The approach illustrated in this section defines a decision calculus that requires only *ordinal* information on the ranking of preferences and importance weights (Yager, 1981).

The typical multiobjective decision problem involves the selection of one alternative, a_i, from a universe of alternatives A given a collection, or set, say {O}, of criteria or objectives that are important to the decision maker. We want to evaluate how well each alternative, or choice, satisfies each objective, and we wish to combine the weighted objectives into an overall decision function in some plausible way. This decision function essentially represents a mapping of the alternatives in A to an ordinal set of ranks. This process naturally requires subjective information from the decision authority concerning the importance of each objective. Ordinal orderings of this importance are usually the easiest to obtain. Numerical values, ratios, and intervals expressing the importance of each objective are difficult to extract and, if attempted and then subsequently altered, can often lead to results inconsistent with the intuition of the decision maker.

To develop this calculus we require some definitions. Define a universe of n alternatives, $A=\{a_1,a_2,\ldots,a_n\}$, and a set of r objectives, $O=\{O_1,O_2,\ldots,O_r\}$. Let O_i indicate the ith objective. Then the degree of membership of alternative a in O_i, denoted $\mu_{O_i}(a)$, is the degree to which alternative a satisfies the criteria specified for this objective. We seek a decision function that simultaneously satisfies all of the decision objectives; hence, the decision function, D, is given by the intersection of all the objective sets,

$$D=O_1\cap O_2\cap\cdots\cap O_r. \tag{9.17}$$

Therefore, the grade of membership that the decision function, D, has for each alternative a is given as

$$\mu_D(a)=\ \min\left[\mu_{O_1}(a),\ \mu_{O_2}(a),\ldots,\mu_{O_r}(a)\right]. \tag{9.18}$$

The optimum decision, a^*, will then be the alternative that satisfies

$$\mu_D(a^*)=\ \max_{a\in A}\left(\mu_D(a)\right). \tag{9.19}$$

We now define a set of preferences, $\{P\}$, which we will constrain to being linear and ordinal. Elements of this preference set can be linguistic values such as none, low, medium, high, absolute, or perfect; or they could be values on the interval [0, 1]; or they could be values on any other linearly ordered scale, for example, [−1, 1] and [1, 10]. These preferences will be attached to each of the objectives to quantify the decision maker's feelings about the influence that each objective should have on the chosen alternative. Let the parameter, b_i, be contained on the set of preferences, $\{P\}$, where $i=1,2,\ldots,r$. Hence, we have for each objective a measure of how important it is to the decision maker for a given decision.

The decision function, D, now takes on a more general form when each objective is associated with a weight expressing its importance to the decision maker. This function is represented as the intersection of r-tuples, denoted as a decision measure, $M(O_i, b_i)$, involving objectives and preferences,

$$D=M(O_1,b_1)\cap M(O_2,b_2)\cap\cdots\cap M(O_r,b_r) \tag{9.20}$$

A key question is what operation should relate each objective, O_i, and its importance, b_i, that preserves the linear ordering required of the preference set, and at the same time relates the two quantities in a logical way where negation is also accommodated. It turns out that the classical implication operator satisfies all of these requirements. Hence, the decision measure for a particular alternative, a, can be replaced with a classical implication of the form

$$M(O_i(a),b_i)=b_i\rightarrow O_i(a)=\overline{b_i}\vee O_i(a). \tag{9.21}$$

Justification of the implication as an appropriate measure can be developed using an intuitive argument (Yager, 1981). The statement "b_i implies O_i" indicates a unique relationship between a preference and its associated objective. Although various objectives can have the same preference weighting in a cardinal sense, they will be unique in an ordinal sense even though the equality situation $b_i = b_j$ for $i \neq j$ can exist for some objectives. Ordering will be preserved because $b_i \geq b_j$ will contain the equality case as a subset. Therefore, a reasonable decision model will be the joint intersection of r decision measures,

$$\mathrm{D} = \bigcap_{i=1}^{r} \left(\overline{b_i} \cup \mathrm{O}_i\right), \tag{9.22}$$

and the optimum solution, a^*, is the alternative that maximizes D. If we define

$$\mathrm{C}_i = \overline{b_i} \cup \mathrm{O}_i, \quad \text{hence} \quad \mu_{\mathrm{C}_i}(a) = \max\left[\mu_{\overline{b_i}}(a), \mu_{\mathrm{O}_i}(a)\right], \tag{9.23}$$

then the optimum solution, expressed in membership form, is given by

$$\mu_D(a^*) = \max_{a \in \mathrm{A}} \left[\min\left\{\mu_{\mathrm{C}_1}(a), \mu_{\mathrm{C}_2}(a), \ldots, \mu_{\mathrm{C}_r}(a)\right\}\right]. \tag{9.24}$$

This model is intuitive in the following manner. As the ith objective becomes more important in the final decision, b_i increases, causing $\overline{b_i}$ to decrease, which in turn causes $\mathrm{C}_i(a)$ to decrease, thereby increasing the likelihood that $\mathrm{C}_i(a) = \mathrm{O}_i(a)$, where now $\mathrm{O}_i(a)$ will be the value of the decision function, D, representing alternative a (Equation (9.22)). As we repeat this process for other alternatives, a, Equation (9.24) reveals that the largest value $\mathrm{O}_i(a)$ for other alternatives will eventually result in the choice of the optimum solution, a^*. This is exactly how we would want the process to work.

Yager (1981) gives a good explanation of the value of this approach. For a particular objective, the negation of its importance (preference) acts as a barrier such that all ratings of alternatives below that barrier become equal to the value of that barrier. Here, we disregard all distinctions less than the barrier while keeping distinctions above this barrier. This process is similar to the grading practice of academics who lump all students whose class averages fall below 60% into the F category, while keeping distinctions of A, B, C, and D for students above this percentile. However, in the decision model developed here this barrier varies, depending on the preference (importance) of the objective to the decision maker. The more important is the objective, the lower is the barrier, and thus the more levels of distinction there are. As an objective becomes less important, the distinction barrier increases, which lessens the penalty to the objective. In the limit, if the objective becomes totally unimportant, then the barrier is raised to its highest level and all alternatives are given the same weight and no distinction is made. Conversely, if the objective becomes the most important, all distinctions remain. In sum, the more important an objective is in the decision process, the more significant its effect on the decision function, D.

A special procedure (Yager, 1981) should be followed in the event of a numerical tie between two or more alternatives. If two alternatives, x and y, are tied, their respective decision values are equal, that is, $\mathrm{D}(x) = \mathrm{D}(y) = \max_{a \in A}[\mathrm{D}(a)]$, where $a = x = y$. Because $\mathrm{D}(a) = \min_i[\mathrm{C}_i(a)]$, there exists some alternative k such that $C_k(x) = D(x)$ and some alternative g such that $C_g(y) = D(y)$. Let

$$\hat{D}(x) = \min_{i \neq k} [C_i(x)] \quad \text{and} \quad \hat{D}(y) = \min_{i \neq g} [C_i(y)]. \tag{9.25}$$

Then, we compare $\hat{D}(x)$ and $\hat{D}(y)$ and if, for example, $\hat{D}(x) > \hat{D}(y)$, we select x as our optimum alternative. However, if a tie still persists, that is, if $\hat{D}(x) = \hat{D}(y)$, then there exist some other alternatives j and h such that $\hat{D}(x) = C_j(x) = \hat{D}(y) = C_h(y)$. Then, we formulate

$$\hat{\hat{D}}(x) = \min_{i \neq k,j} [C_i(x)] \quad \text{and} \quad \hat{\hat{D}}(y) = \min_{i \neq g,h} [C_i(y)] \tag{9.26}$$

and compare $\hat{\hat{D}}(x)$ and $\hat{\hat{D}}(y)$. The tie-breaking procedure continues in this manner until an unambiguous optimum alternative emerges or all of the alternatives have been exhausted. In the latter case where a tie still results, some other tie-breaking procedure, such as a refinement in the preference scales, can be used.

Example 9.5

A geotechnical engineer on a construction project must prevent a large mass of soil from sliding into a building site during construction and must retain this mass of soil indefinitely after construction to maintain stability of the area around a new facility to be constructed on the site (Adams, 1994). The engineer therefore must decide which type of retaining wall design to select for the project. Among the many alternative designs available, the engineer reduces the list of candidate retaining wall designs to three: (1) a mechanically stabilized embankment (MSE) wall, (2) a mass concrete spread wall (Conc), and (3) a gabion (Gab) wall. The owner of the facility (the decision maker) has defined four objectives that impact the decision: (1) the cost of the wall (Cost), (2) the maintainability (Main) of the wall, (3) whether the design is a standard one (SD), and (4) the environmental (Env) impact of the wall. Moreover, the owner also decides to rank the preferences for these objectives on the unit interval. Hence, the engineer sets up the problem as follows:

$$\begin{aligned} &\mathrm{A} = \{\mathrm{MSE}, \mathrm{Conc}, \mathrm{Gab}\} = \{a_1, a_2, a_3\}. \\ &\mathrm{O} = \{\mathrm{Cost}, \mathrm{Main}, \mathrm{SD}, \mathrm{Env}\} = \{\mathrm{O}_1, \mathrm{O}_2, \mathrm{O}_3, \mathrm{O}_4\}. \\ &\mathrm{P} = \{b_1, b_2, b_3, b_4\} \to [0, 1]. \end{aligned}$$

From previous experience with various wall designs, the engineer first rates the retaining walls with respect to the objectives, given here. These ratings are fuzzy sets expressed in Zadeh’s notation.

$$\begin{aligned} \underset{\sim}{O}_1 &= \left\{ \frac{0.4}{\mathrm{MSE}} + \frac{1}{\mathrm{Conc}} + \frac{0.1}{\mathrm{Gab}} \right\}. \\ \underset{\sim}{O}_2 &= \left\{ \frac{0.7}{\mathrm{MSE}} + \frac{0.8}{\mathrm{Conc}} + \frac{0.4}{\mathrm{Gab}} \right\}. \\ \underset{\sim}{O}_3 &= \left\{ \frac{0.2}{\mathrm{MSE}} + \frac{0.4}{\mathrm{Conc}} + \frac{1}{\mathrm{Gab}} \right\}. \\ \underset{\sim}{O}_4 &= \left\{ \frac{1}{\mathrm{MSE}} + \frac{0.5}{\mathrm{Conc}} + \frac{0.5}{\mathrm{Gab}} \right\}. \end{aligned}$$

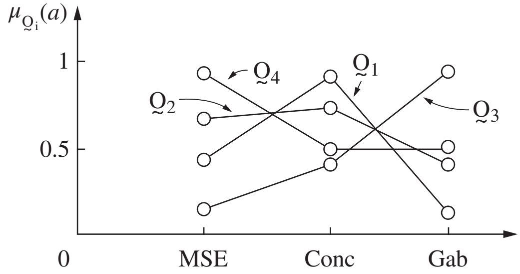

Figure 9.4 Membership for each alternative with respect to the objectives.

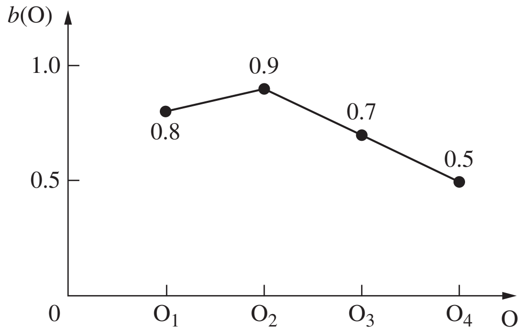

Figure 9.5 Preferences in the first scenario.

These membership functions for each of the alternatives are shown graphically in Figure 9.4.

The engineer wishes to investigate two decision scenarios. Each scenario propagates a different set of preferences from the owner, who wishes to determine the sensitivity of the optimum solutions to the preference ratings. In the first scenario, the owner lists the preferences for each of the four objectives, as shown in Figure 9.5. From these preference values, the following calculations result:

$$b_1 = 0.8, \quad b_2 = 0.9, \quad b_3 = 0.7, \quad b_4 = 0.5.$$

$$\overline{b}_1 = 0.2, \quad \overline{b}_2 = 0.1, \quad \overline{b}_3 = 0.3, \quad \overline{b}_4 = 0.5.$$

$$\begin{aligned}
\mathrm{D}(a_1) = \mathrm{D}(\mathrm{MSE}) &= \left(\overline{b}_1 \cup \mathrm{O}_1\right) \cap \left(\overline{b}_2 \cup \mathrm{O}_2\right) \cap \left(\overline{b}_3 \cup \mathrm{O}_3\right) \cap \left(\overline{b}_4 \cup \mathrm{O}_4\right) \\
&= (0.2 \vee 0.4) \wedge (0.1 \vee 0.7) \wedge (0.3 \vee 0.2) \wedge (0.5 \vee 1) \\
&= 0.4 \wedge 0.7 \wedge 0.3 \wedge 1 = 0.3. \\
\mathrm{D}(a_2) = \mathrm{D}(\mathrm{Conc}) &= (0.2 \vee 1) \wedge (0.1 \vee 0.8) \wedge (0.3 \vee 0.4) \wedge (0.5 \vee 0.5) \\
&= 1 \wedge 0.8 \wedge 0.4 \wedge 0.5 = 0.4. \\
\mathrm{D}(a_3) = \mathrm{D}(\mathrm{Gab}) &= (0.2 \vee 0.1) \wedge (0.1 \vee 0.4) \wedge (0.3 \vee 1) \wedge (0.5 \vee 0.5) \\
&= 0.2 \wedge 0.4 \wedge 1 \wedge 0.5 = 0.2.
\end{aligned}$$

$$\mathrm{D}* = \max\{\mathrm{D}(a_1), \mathrm{D}(a_2), \mathrm{D}(a_3)\} = \max\{0.3, 0.4, 0.2\} = 0.4.$$

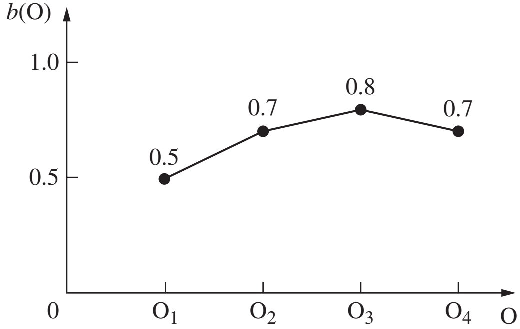

Figure 9.6 Preferences in the second scenario.

Thus, the engineer chooses the second alternative, a_2, a concrete (Conc) wall as the retaining design under preference scenario 1.

Now, in the second scenario the engineer was given a different set of preferences by the owner, as shown in Figure 9.6. From the preference values in Figure 9.6, the following calculations result:

$$b_1 = 0.5, \quad b_2 = 0.7, \quad b_3 = 0.8, \quad b_4 = 0.7.$$

$$\overline{b}_1 = 0.5, \quad \overline{b}_2 = 0.3, \quad \overline{b}_3 = 0.2, \quad \overline{b}_4 = 0.3.$$

$$\begin{aligned}
\mathrm{D}(a_1) = \mathrm{D}(\mathrm{MSE}) &= \left(\overline{b}_1 \cup \mathrm{O}_1\right) \cap \left(\overline{b}_2 \cup \mathrm{O}_2\right) \cap \left(\overline{b}_3 \cup \mathrm{O}_3\right) \cap \left(\overline{b}_4 \cup \mathrm{O}_4\right) \\
&= (0.5 \vee 0.4) \wedge (0.3 \vee 0.7) \wedge (0.2 \vee 0.2) \wedge (0.3 \vee 1) \\
&= 0.5 \wedge 0.7 \wedge 0.2 \wedge 1 = 0.2. \\
\mathrm{D}(a_2) = \mathrm{D}(\mathrm{Conc}) &= (0.5 \vee 1) \wedge (0.3 \vee 0.8) \wedge (0.2 \vee 0.4) \wedge (0.3 \vee 0.5) \\
&= 1 \wedge 0.8 \wedge 0.4 \wedge 0.5 = 0.4. \\
\mathrm{D}(a_3) = \mathrm{D}(\mathrm{Gab}) &= (0.5 \vee 0.1) \wedge (0.3 \vee 0.4) \wedge (0.2 \vee 1) \wedge (0.3 \vee 0.5) \\
&= 0.5 \wedge 0.4 \wedge 1 \wedge 0.5 = 0.4.
\end{aligned}$$

Therefore, $\mathrm{D}^* = \max\{\mathrm{D}(a1), \mathrm{D}(a2), \mathrm{D}(a3)\} = \max\{0.2, 0.4, 0.4\} = 0.4$. But there is a tie between alternative a_2 and a_3. To resolve this tie, the engineer implements Equation (9.25). The engineer looks closely at D(a_2) and D(a_3) and notes that the decision value of 0.4 for D(a_2) came from the third term (i.e., C_3 (a_2); hence $k = 3$ in Equation (9.25)), and that the decision value of 0.4 for D(a_3) came from the second term (i.e., C_2 (a_3); hence $g = 2$ in Equation (9.25)). Then, the calculations proceed again between the tied choices a_2 and a_3:

$$\begin{aligned}
\hat{\mathrm{D}}(a_2) = \hat{\mathrm{D}}(\mathrm{Conc}) &= (0.5 \vee 1) \wedge (0.3 \vee 0.8) \wedge (0.3 \vee 0.5) \\
&= 1 \wedge 0.8 \wedge 0.5 = 0.5. \\
\hat{\hat{\mathrm{D}}}(a_3) = \hat{\mathrm{D}}(\mathrm{Gab}) &= (0.5 \vee 0.1) \wedge (0.2 \vee 1) \wedge (0.3 \vee 0.5) \\
&= 0.5 \wedge 1 \wedge 0.5 = 0.5.
\end{aligned}$$

Then $\mathrm{D}^* = \max\{\hat{\mathrm{D}}(a_2),\ \hat{\mathrm{D}}(a_3)\} = \max\{0.5,\ 0.5\} = 0.5$, and there is still a tie between alternatives a_2 and a_3. To resolve this second tie, the engineer implements Equation (9.26). The engineer looks closely at $\hat{\mathrm{D}}(a_2)$ and $\hat{\mathrm{D}}(a_3)$ and notes that the decision value of 0.5 for $\hat{\mathrm{D}}(a_2)$ came from the third term (i.e., $\mathrm{C}_3\ (a_2)$; hence $j = 3$ in Equation (9.26)), and that the decision value of 0.5 for $\hat{\mathrm{D}}(a_3)$ came from the first term *and* the third term (i.e., $\mathrm{C}_1(a_3) = \mathrm{C}_3(a_3)$; hence $h = 1$ and $h = 3$ in Equation (9.26)). Then, the calculations proceed again between the tied choices a_2 and a_3:

$$\hat{\hat{\mathrm{D}}}(a_2) = \hat{\hat{\mathrm{D}}}(\mathrm{Conc}) = (0.5 \vee 1) \wedge (0.3 \vee 0.8) = 0.8.$$

$$\hat{\hat{\mathrm{D}}}(a_3) = \hat{\hat{\mathrm{D}}}(\mathrm{Gab}) = (0.2 \vee 1) = 1.$$

From these results, $\mathrm{D}^* = \max\{\hat{\mathrm{D}}(a_2),\ \hat{\mathrm{D}}(a_3)\} = 1$; hence, the tie is finally broken and the engineer chooses retaining wall a_3, a gabion wall, for the design under preference scenario 2.

Fuzzy Bayesian Decision Method

Classical statistical decision making involves the notion that the uncertainty in the future can be characterized probabilistically as discussed in the introduction to this chapter. When we want to make a decision among various alternatives, our choice is predicated on information about the future, which is normally discretized into various "states of nature." If we knew with certainty the future states of nature, we would not need an analytic method to assess the likelihood of a given outcome. Unfortunately, we do not know what the future will entail so we have devised methods to make the best choices given an uncertain environment. Classical Bayesian decision methods presume that future states of nature can be characterized as probability events. For example, consider the condition of "cloudiness" in tomorrow's weather by discretizing the state space into three levels and assessing each level probabilistically: the chance of a very cloudy day is 0.5, a partly cloudy day is 0.2, and a sunny (no clouds) day is 0.3. By convention the probabilities sum to unity. The problem with the Bayesian scheme here is that the events are vague and ambiguous. How many clouds does it take to transition between very cloudy and cloudy? If there is one small cloud in the sky, does this mean it is not sunny? This is the classic sorites paradox discussed in Chapter 5.

The following material first presents Bayesian decision making and then starts to consider ambiguity in the value of new information, in the states of nature, and in the alternatives in the decision process (Terano, Asai, and Sugeno, 1992). Examples will illustrate these points.

First, we shall consider the formation of probabilistic decision analysis. Let $\mathrm{S} = \{s_1, s_2, \ldots, s_n\}$ be a set of possible states of nature, and the probabilities that these states will occur are listed in a vector,

$$\mathbf{P} = \{p(s_1), p(s_2), \ldots, p(s_n)\}, \quad \text{where} \quad \sum_{i=1}^{n} p(s_i) = 1. \tag{9.27}$$

The probabilities expressed in Equation (9.27) are called *prior probabilities* in Bayesian jargon because they express prior knowledge about the true states of nature. Assume that the decision maker can choose among m alternatives, $\mathrm{A} = \{a_1, a_2, \ldots, a_m\}$, and for a given

Table 9.1 Utility matrix.

	States s_i			
Action a_j	s_1	s_2	$\cdots$	s_n
a_1	u_{11}	u_{12}	$\cdots$	u_{1n}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
a_m	u_{m1}	u_{m2}	$\cdots$	u_{mn}

alternative a_j we assign a utility value, u_{ji}, if the future state of nature turns out to be state s_i. These utility values should be determined by the decision maker because they express value, or cost, for each alternative-state pair, that is, for each a_j–s_i combination. The utility values are usually arranged in a matrix of the form shown in Table 9.1. The expected utility associated with the jth alternative would be

$$E(u_j) = \sum_{i=1}^{n} u_{ji} p(s_i). \tag{9.28}$$

The most common decision criterion is the *maximum* expected utility among all the alternatives, that is,

$$E(u^*) = \max_j E(u_j), \tag{9.29}$$

which leads to the selection of alternative a_k if $u^* = E(u_k)$.

Example 9.6

Suppose you are a geological engineer who has been asked by the chief executive officer (CEO) of a large oil firm to help make a decision about whether to drill for natural gas in a particular geographic region of northwestern New Mexico. You determine for your first attempt at the decision process that there are only two states of nature regarding the existence of natural gas in the region:

$$s_1 = \text{there is natural gas}$$
$$s_2 = \text{there is no natural gas}$$

and you are able to find from previous drilling information that the prior probabilities for each of these states is

$$p(s_1) = 0.5,$$
$$p(s_2) = 0.5.$$

Note these probabilities sum to unity. You suggest that there are two alternatives in this decision:

$$a_1 = \text{drill for gas},$$
$$a_2 = \overline{a}_1 = \text{do not drill for gas}.$$

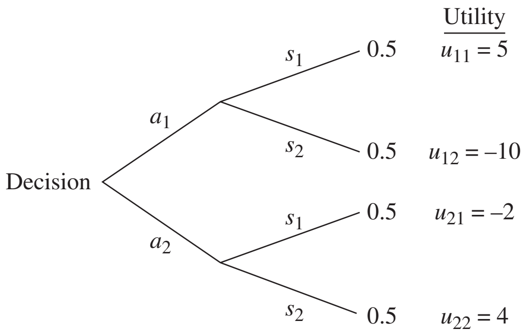

Figure 9.7 Decision tree for the two-alternative, two-state problem of Example 9.6.

The decision maker (the CEO) helps you assemble a utility matrix to get the process started. The CEO tells you that the best situation for the firm is to decide to drill for gas, and subsequently find that gas, indeed, was in the geologic formation. The CEO assesses this value (u_{11}) as +5 in nondimensional units; in this case, the CEO would have gambled (drilling costs big money) and won. Moreover, the CEO feels that the worst possible situation would be to drill for gas, and subsequently find that there was no gas in the area. Because this would cost time and money, the CEO determines that the value for this would be $u_{12} = -10$ units; the CEO would have gambled and lost—big. The other two utilities are assessed by the decision maker in nondimensional units as $u_{21} = -2$ and $u_{22} = 4$. Hence, the utility matrix for this situation is given as

$$\mathrm{U} = \begin{bmatrix} 5 & -10 \\ -2 & 4 \end{bmatrix}.$$

Figure 9.7 shows the decision tree for this problem, of which the two initial branches correspond to the two possible alternatives and the second layer of branches corresponds to the two possible states of nature. Superposed on the tree branches are the prior probabilities. The expected utility, in nondimensional units, for each alternative a_1 and a_2 is, from Equation (9.28),

$$E(a_1) = (0.5)(5) + (0.5)(-10) = -2.5,$$
$$E(a_2) = (0.5)(-2) + (0.5)(4) = 1.0,$$

and we see that the maximum utility, using Equation (9.29), is 1.0, which comes from alternative a_2; hence, only on the basis of prior information (prior probabilities) the CEO decides not to drill for natural gas (alternative a_2).

In many decision situations an intermediate issue arises: Should you get more information about the true states of nature prior to deciding? Suppose some new information regarding the true states of nature S is available from r experiments or other observations and is collected in a data vector, $\mathbf{X} = \{x_1, x_2, \ldots, x_r\}$. This information can be used in the Bayesian approach to update the prior probabilities, $p(s_i)$, in the following manner. First, the new information is expressed in the form of conditional probabilities, where the probability of each piece of data, x_k, where $k = 1, 2, \ldots, r$, is assessed according to whether the true state of nature, s_i, is known (not uncertain); these probabilities are presumptions of the future because they are equivalent to the following statement: *Given that we know that the true state of nature is* s_i, *the probability that the piece of new information* x_k *confirms that the true state is* s_i *is* $p(x_k|s_i)$. In the literature, these conditional probabilities, denoted $p(x_k|s_i)$, are also called *likelihood values*. The likelihood values are then used as weights on the previous information, the prior probabilities $p(s_i)$, to find updated probabilities, known as *posterior probabilities*, denoted $p(s_i|x_k)$. The posterior probabilities are equivalent to this statement: *Given that the piece of new information* x_k *is true, the probability that the true state of nature is* s_i *is* $p(s_i|x_k)$. These updated probabilities are determined by Bayes's rule,

$$p(s_i|x_k) = \frac{p(x_k|s_i)}{p(x_k)} p(s_i), \tag{9.30}$$

where the term in the denominator of Equation (9.30), $p(x_k)$, is the marginal probability of the data x_k and is determined using the total probability theorem

$$p(x_k) = \sum_{i=1}^{n} p(x_k|s_i) \cdot p(s_i). \tag{9.31}$$

Now the expected utility for the jth alternative, given the data x_k, is determined from the posterior probabilities (instead of the priors),

$$E(u_j|x_k) = \sum_{i=1}^{n} u_{ji} p(s_i|x_k), \tag{9.32}$$

and the maximum expected utility, given the new data x_k, is now given as

$$E(u^*|x_k) = \max_j E(u_j|x_k). \tag{9.33}$$

To determine the unconditional maximum expected utility, we need to weight each of the r conditional expected utilities given in Equation (9.33) by the respective marginal probabilities for each datum x_k, that is, by $p(x_k)$, given in Equation (9.34) as

$$E(u_x^*) = \sum_{k=1}^{r} E(u^*|x_k) \cdot p(x_k). \tag{9.34}$$

We can now introduce a new notion in the decision-making process, called the *value* of information, $V(x)$. In the case we have just introduced where there is some uncertainty about the new information, $X = \{x_1, x_2, \ldots, x_r\}$, we call the information *imperfect* information. The value of this imperfect information, $V(x)$, can be assessed by taking the difference between the maximum expected utility without any new information, Equation (9.29), and the maximum expected utility with the new information, Equation (9.34), that is,

$$V(x) = E\left(u_x^*\right) - E(u^*). \tag{9.35}$$

We now introduce yet another notion in this process, called *perfect* information. This exercise is an attempt to develop a boundary condition for our problem, one that is altruistic in nature, that is, can never be achieved in reality, but nonetheless, is quite useful in a mathematical sense to give us some scale on which to assess the value of imperfect information. If information is considered to be perfect (i.e., can predict the future states of nature precisely), we can say that the conditional probabilities are free of dissonance. That is, each new piece of information, or data, predicts only one state of nature; hence, there is no ambivalence about what state is predicted by the data. However, if there is more than one piece of information, the probabilities for a particular state of nature have to be shared by all the data. Mathematically, perfect information is represented by posterior probabilities of 0 or 1, that is,

$$p(s_i|x_k) = \begin{cases} 1 \\ 0 \end{cases}. \tag{9.36}$$

We call this perfect information x_p. For perfect information, the maximum expected utility becomes (Example 9.7)

$$E(u_{x_p}^*) = \sum_{k=1}^{r} E(u_{x_p}^* | x_k) p(x_k), \tag{9.37}$$

and the value of perfect information becomes

$$V(x_p) = E(u_{x_p}^*) - E(u^*). \tag{9.38}$$

Example 9.7

We continue with our gas exploration problem from Example 9.6. We had two states of nature—gas, s_1, and no gas, s_2—and two alternatives—drill, a_1, and no drill, a_2. The prior probabilities were uniform,

$$p(s_1) = 0.5,$$
$$p(s_2) = 0.5.$$

Now, let us suppose the CEO of the natural gas company wants to reconsider the utility values. The CEO provides the utility matrix of Table 9.2 in the same form as Table 9.1. Further, the CEO has asked you to collect new information by taking eight geological boring samples from the region being considered for drilling. You have a natural gas expert examine the results of these eight tests, and get the expert's opinions about the conditional probabilities in the form of a matrix, given in Table 9.3. Moreover, you ask the natural gas expert for an assessment

Table 9.2 Utility matrix for natural gas example.

u_{ji}	s_1	s_2
a_1	4	−2
a_2	−1	2

Table 9.3 Conditional probabilities for *imperfect* information.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
$p(x_k\|s_1)$	0	0.05	0.1	0.1	0.2	0.4	0.1	0.05	Σ row = 1
$p(x_k\|s_2)$	0.05	0.1	0.4	0.2	0.1	0.1	0.05	0	Σ row = 1

Table 9.4 Conditional probabilities for *perfect* information.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
$p(x_k\|s_1)$	0	0	0	0	0.2	0.5	0.2	0.1	Σ row = 1
$p(x_k\|s_2)$	0.1	0.2	0.5	0.2	0	0	0	0	Σ row = 1

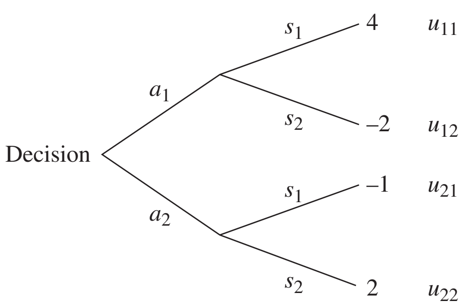

Figure 9.8 Decision tree showing utility values.

about how the conditional probabilities might change if they were perfect tests capable of providing perfect information. The expert gives you the matrix shown in Table 9.4.

As an engineer assisting the CEO, you now conduct a decision analysis. Because the CEO changed the utility values, you have to recalculate the expected utility of making the decision on the basis of just the prior probabilities, before any new information is acquired. The decision tree for this situation is shown in Figure 9.8.

The expected utilities and maximum expected utility, based just on prior probabilities, are

$$E(a_1) = (4)(0.5) + (-2)(0.5) = 1.0.$$
$$E(a_2) = (-1)(0.5) + (2)(0.5) = 0.5.$$
$$E(u^*) = 1;\ \text{hence, you choose alternative}\ a_1,\ \text{drill of natural gas.}$$

Table 9.5 Posterior probabilities based on imperfect information.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(s_1 \mid x_k)$	0	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{5}$	$\frac{2}{3}$	1
$p(s_2 \mid x_k)$	1	$\frac{2}{3}$	$\frac{4}{5}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{3}$	0
$p(x_k)$	0.025	0.075	0.25	0.15	0.15	0.25	0.075	0.025
$E(u^* \mid x_k)$	2	1	$\frac{7}{5}$	1	2	$\frac{14}{5}$	2	4
$a_j \mid x_k$	a_2	a_2	a_2	a_2	a_1	a_1	a_1	a_1

You are now ready to assess the changes in this decision process by considering additional information, both imperfect and perfect. Table 9.5 summarizes your calculations for the new prior probabilities, $p(s_1 \mid x_k)$ and $p(s_2 \mid x_k)$, the marginal probabilities for the new information, $p(x_k)$, the expected conditional utilities, $E(u^* \mid x_k)$, and the expected alternatives, $a_j \mid x_k$.

Typical calculations for the values in Table 9.5 are provided here. For the marginal probabilities for the new *imperfect* information, use Equation (9.31), the conditional probabilities from Table 9.3, and the prior probabilities:

$$p(x_1) = (0)(0.5) + (0.05)(0.5) = 0.025.$$
$$p(x_4) = (0.1)(0.5) + (0.2)(0.5) = 0.15.$$

The posterior probabilities are calculated with the use of Equation (9.30), using the conditional probabilities from Table 9.3, the prior probabilities, and the marginal probabilities, $p(x_k)$, just determined and summarized in Table 9.5 (third row); for example,

$$p(s_1|x_2) = \frac{0.05(0.5)}{0.075} = \frac{1}{3}, \quad p(s_2|x_2) = \frac{0.1(0.5)}{0.075} = \frac{2}{3}, \quad \ldots$$
$$p(s_1|x_6) = \frac{0.4(0.5)}{0.25} = \frac{4}{5}, \quad p(s_2|x_6) = \frac{0.1(0.5)}{0.25} = \frac{1}{5}, \quad \ldots$$

The conditional expected utilities, $E(u^* \mid x_k)$, are calculated using first Equation (9.32) and then Equation (9.33); for example,

$$E(u_1|x_3) = \left(\frac{1}{5}\right)(4) + \left(\frac{4}{5}\right)(-2) = -\frac{4}{5} \quad \text{and} \quad E(u_2|x_3) = \left(\frac{1}{5}\right)(-1) + \left(\frac{4}{5}\right)(2) = \frac{7}{5}.$$

Hence, $E(u^*|x_3) = \max\left(-\frac{4}{5}, \frac{7}{5}\right) = \frac{7}{5}$ (choose alternative a_2):

$$E(u_1|x_8) = (1)(4) + (0)(-2) = -4 \quad \text{and} \quad E(u_2|x_8) = (1)(-1) + (0)(2) = -1$$

Hence, $E(u^* \mid x_8)$ = max (4, –1) = 4 (choose alternative a_1).

Table 9.6 Posterior probabilities based on perfect information.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$p(s_1 \vert x_k)$	0	0	0	0	1	1	1	1
$p(s_2 \vert x_k)$	1	1	1	1	0	0	0	0
$p(x_k)$	0.05	0.1	0.25	0.1	0.1	0.25	0.1	0.05
$E(u^* \vert x_k)$	2	2	2	2	4	4	4	4
$a_j \vert x_k$	a_2	a_2	a_2	a_2	a_1	a_1	a_1	a_1

Now use Equation (9.34) to calculate the overall unconditional expected utility for *imperfect* information, which is actually the sum of pairwise products of the values in the third and fourth rows of Table 9.5, for example,

$$E\left(u_x^*\right) = (0.025)(2) + (0.075)(1) + \cdots + (0.025)(4) = 1.875,$$

and the value of the new *imperfect* information, using Equation (9.35), is

$$V(x) = E\left(u_x^*\right) - E(u^*) = 1.875 - 1 = 0.875.$$

To decide what alternative to choose, notice in Table 9.5 that the total utility favoring a_1 is $10.8\left(2 + \frac{14}{5} + 2 + 4\right)$ and the total utility favoring a_2 is $5.4\left(2 + 1 + \frac{7}{5} + 1\right)$. Hence, the CEO chooses alternative a_1, to drill for gas. In effect, the new information has not changed the CEO's mind about drilling.

You begin the process of assessing the changes because of the consideration of the hypothetical *perfect* information. Table 9.6 summarizes your calculations for the new prior probabilities, $p(s_1|x_k)$ and $p(s_2|x_k)$, the marginal probabilities for the perfect information, $p(x_k)$, the expected conditional utilities, $E(u_{x_p}^*|x_k)$, and the expected alternatives, $a_j|x_k$. These are calculated in the same way as those in Table 9.5, except that you make use of the *perfect* conditional probabilities of Table 9.4.

Equation (9.37) is used to calculate the overall unconditional expected utility for *perfect* information, which is actually the sum of pairwise products of the values in the third and fourth rows of Table 9.6, for example,

$$E\left(u_{x_p}^*\right) = (0.05)(2) + (0.1)(2) + \cdots + (0.05)(4),$$

and the value of the new *perfect* information, using Equation (9.38), is

$$V(x_p) = E\left(u_{x_p}^*\right) - E(u^*) = 3 - 1 = 2.0.$$

Alternative a_1 is still the choice here. We note that the hypothetical information has a value of 2 and the imperfect information has a value of less than half of this, 0.875. This difference can be used to assess the value of the imperfect information compared to both no information (1) and perfect information (3).

We now discuss the fact that the new information might be inherently fuzzy (Okuda, Tanaka, and Asai,1974, 1978). Suppose the new information, $\mathrm{X} = \{x_1, x_2, \ldots, x_r\}$, is a universe of discourse in the units appropriate for the new information. Then, we can define fuzzy events, $\underset{\sim}{\mathrm{M}}$, on this information, such as "good" information, "moderate" information, and "poor" information. The fuzzy event will have membership function $\mu_{\underset{\sim}{\mathrm{M}}}(x_k),\ k = 1, 2, \ldots, r$. We can now define the idea of a "probability of a fuzzy event," that is, the probability of $\underset{\sim}{\mathrm{M}}$, as

$$P(\underset{\sim}{\mathrm{M}}) = \sum_{k=1}^{r} \mu_{\underset{\sim}{\mathrm{M}}}(x_k) p(x_k). \tag{9.39}$$

We note in Equation (9.39) that if the fuzzy event is, in fact, crisp, that is, $\underset{\sim}{\mathrm{M}} = \mathrm{M}$, then the probability reduces to

$$P(\mathrm{M}) = \sum_{x_k \in \mathrm{M}} p(x_k), \qquad \mu_{\mathrm{M}} = \begin{cases} 1, & x_k \in \mathrm{M}, \\ 0, & \text{otherwise}, \end{cases} \tag{9.40}$$

where Equation (9.40) describes the probability of a crisp event simply as the sum of the marginal probabilities of those data points, x_k, that are defined to be in the event, M. On the basis of this, the posterior probability of s_i, given fuzzy information $\underset{\sim}{\mathrm{M}}$, is

$$P(s_i|\underset{\sim}{\mathrm{M}}) = \frac{\sum_{k=1}^{r} p(x_k|s_i) \mu_{\underset{\sim}{\mathrm{M}}}(x_k) p(s_i)}{P(\underset{\sim}{\mathrm{M}})} = \frac{P(\underset{\sim}{\mathrm{M}}|s_i) p(s_i)}{P(\underset{\sim}{\mathrm{M}})}, \tag{9.41}$$

where

$$p(\underset{\sim}{\mathrm{M}}|s_i) = \sum_{k=1}^{r} p(x_k|s_i) \mu_{\underset{\sim}{\mathrm{M}}}(x_k). \tag{9.42}$$

We now define the collection of all the fuzzy events describing fuzzy information as an *orthogonal* fuzzy information system, $\Phi = \{\underset{\sim}{\mathrm{M}}_1, \underset{\sim}{\mathrm{M}}_2, \ldots, \underset{\sim}{\mathrm{M}}_g,\}$, where by *orthogonal* we mean that the sum of the membership values for each fuzzy event, $\underset{\sim}{\mathrm{M}}_i$, for every data point in the information universe, x_k, equals unity (Tanaka, Okuda, and Asai, 1976). That is,

$$\sum_{t=1}^{g} \mu_{\underset{\sim}{\mathrm{M}}_t}(x_k) = 1, \qquad \text{for all}\, x_k \in \mathrm{X}. \tag{9.43}$$

If the fuzzy events on the new information universe are *orthogonal*, we can extend the Bayesian approach to consider fuzzy information. The fuzzy equivalents of Equations (9.32) to (9.34) become, for a fuzzy event $\underset{\sim}{\mathrm{M}}_t$,

$$E(u_j|\underset{\sim}{M}_t) = \sum_{i=1}^{n} u_{ij} \cdot p(s_i|\underset{\sim}{M}_t). \tag{9.44}$$

$$E(u^*|\underset{\sim}{M}_t) = \max_j E(u_j|\underset{\sim}{M}_t). \tag{9.45}$$

$$E\left(u^*_\Phi\right) = \sum_{t=1}^{g} E(u^*|\underset{\sim}{M}_t) \cdot p(\underset{\sim}{M}_t). \tag{9.46}$$

Now the value of fuzzy information can be determined in an analogous manner as

$$V(\Phi) = E\left(u^*_\Phi\right) - E(u^*). \tag{9.47}$$

Example 9.8 (continuation of Example 9.7).
Suppose the eight data samples are from overlapping, ill-defined parcels within the drilling property. We define an orthogonal fuzzy information system, Φ,

$$\Phi = \left\{\underset{\sim}{M}_1, \underset{\sim}{M}_2, \underset{\sim}{M}_3\right\} = \{\text{fuzzy parcel 1, fuzzy parcel 2, fuzzy parcel 3}\}$$

with membership functions in Table 9.7. The fourth row of Table 9.7 repeats the marginal probabilities for each data, x_k, from Table 9.5. As can be seen in Table 9.7, the sum of the membership values in each column (the first three rows) equals unity, as required for *orthogonal* fuzzy sets.

As before, we use Equation (9.39) to determine the marginal probabilities for each fuzzy event,

$$p(\underset{\sim}{M}_1) = 0.225, \quad p(\underset{\sim}{M}_2) = 0.55, \quad p(\underset{\sim}{M}_3) = 0.225;$$

Equation (9.42) to determine the fuzzy conditional probabilities,

$$p(\underset{\sim}{M}_1|s_1) = 0.1, \quad p(\underset{\sim}{M}_2|s_1) = 0.55, \quad p(\underset{\sim}{M}_3|s_1) = 0.35;$$
$$p(\underset{\sim}{M}_1|s_2) = 0.35, \quad p(\underset{\sim}{M}_2|s_2) = 0.55, \quad p(\underset{\sim}{M}_3|s_2) = 0.1;$$

Table 9.7 Orthogonal membership functions for orthogonal fuzzy events.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
$\mu_{\underset{\sim}{M}_1}(x_k)$	1	1	0.5	0	0	0	0	0
$\mu_{\underset{\sim}{M}_2}(x_k)$	0	0	0.5	1	1	0.5	0	0
$\mu_{\underset{\sim}{M}_3}(x_k)$	0	0	0	0	0	0.5	1	1
$P(x_k)$	0.025	0.075	0.25	0.15	0.15	0.25	0.075	0.025

and Equation (9.41) to determine the fuzzy posterior probabilities,

$$p(s_1|\underset{\sim}{M}_1)=0.222,\quad p(s_1|\underset{\sim}{M}_2)=0.5,\quad p(s_1|\underset{\sim}{M}_3)=0.778;$$
$$p(s_2|\underset{\sim}{M}_1)=0.778,\quad p(s_2|\underset{\sim}{M}_2)=0.5,\quad p(s_2|\underset{\sim}{M}_3)=0.222.$$

Now the conditional fuzzy expected utilities can be determined using Equation (9.44),

$$\begin{aligned}
\underset{\sim}{M}_1: E(u_1|\underset{\sim}{M}_1) &= (4)(0.222)+(-2)(0.778)=-0.668\\
E(u_2|\underset{\sim}{M}_1) &= (-1)(0.222)+(2)(0.778)=1.334;\\
\underset{\sim}{M}_2: E(u_1|\underset{\sim}{M}_2) &= (4)(0.5)+(-2)(0.5)=1.0\\
E(u_2|\underset{\sim}{M}_2) &= (-1)(0.5)+(2)(0.5)=0.5;\\
\underset{\sim}{M}_3: E(u_1|\underset{\sim}{M}_3) &= (4)(0.778)+(-2)(0.222)=2.668\\
E(u_2|\underset{\sim}{M}_3) &= (-1)(0.778)+(2)(0.222)=-0.334;
\end{aligned}$$

the maximum expected utility from Equation (9.46), using each of the foregoing three maximum conditional probabilities,

$$E\left(u_\Phi^*\right)=(0.225)(1.334)+(0.55)(1)+(0.225)(2.668)=1.45;$$

and the value of the fuzzy information from Equation (9.47),

$$V(\Phi)=1.45-1=0.45.$$

Here, we see that the value of the fuzzy information is less than the value of the perfect information (2.0), and less than the value of the imperfect information (0.875). However, it may turn out that fuzzy information is *far less costly* (remember, precision costs) than either the imperfect or perfect (hypothetical) information. Although not developed in this text, this analysis could be extended to consider *cost* of information.

Decision Making under Fuzzy States and Fuzzy Actions

The Bayesian method can be further extended to include the possibility that the states of nature are fuzzy and the decision makers' alternatives are also fuzzy (Tanaka, Okuda, and Asai, 1976). For example, suppose your company wants to expand and you are considering three fuzzy alternatives in terms of the size of a new facility:

$\underset{\sim}{A}_1$ = small-scale project
$\underset{\sim}{A}_2$ = middle-scale project
$\underset{\sim}{A}_3$ = large-scale project.

Just as all fuzzy sets are defined on some universe of discourse, continuous or discrete, the fuzzy alternatives (actions) would also be defined on a universe of discourse, say values of square footage of floor space on a continuous scale of areas in some appropriate units of area. Moreover, suppose further that the economic climate in the future is fuzzy and you pose the following three possible fuzzy states of nature ($\underset{\sim}{F}_s, s=1,2,3$):

$\underset{\sim}{F}_1$ = low rate of economic growth
$\underset{\sim}{F}_2$ = medium rate of economic growth
$\underset{\sim}{F}_3$ = high rate of economic growth,

all of which are defined on a universe of numerical rates of economic growth, say S, where S = $\{s_1, s_2, \ldots, s_n\}$ is a discrete universe of economic growth rates (e.g., –4%, –3%,…, 0%, 1%, 2%,…). The fuzzy states $\underset{\sim}{F}_s$ will be required to be orthogonal fuzzy sets, for us to continue to use the Bayesian framework. This orthogonal condition on the fuzzy states will be the same constraint as illustrated in Equation (9.43), that is,

$$\sum_{s=1}^{3} \mu_{\underset{\sim}{F}_s}(s_i) = 1, \qquad i = 1,2,\ldots,n. \tag{9.48}$$

Further, as we need utility values to express the relationship between crisp alternative-state pairs, we still need a utility matrix to express the value of all the fuzzy alternative-state pairings. Such a matrix will have the form shown in Table 9.8.

Proceeding as before, but now with fuzzy states of nature, the expected utility of fuzzy alternative $\underset{\sim}{A}_j$ is

$$E(u_j) = \sum_{s=1}^{3} \mu_{js} p(\underset{\sim}{F}_s), \tag{9.49}$$

where

$$p(\underset{\sim}{F}_s) = \sum_{i=1}^{n} \mu_{\underset{\sim}{F}_s}(s_i) p(s_i) \tag{9.50}$$

and the maximum utility is

$$E(u^*) = \max_j E(u_j). \tag{9.51}$$

We can have crisp or fuzzy information on a universe of information X = $\{x_1, x_2, \ldots, x_r\}$, for example, rate of increase of gross national product. Our fuzzy information will again reside on a

Table 9.8 Utility values for fuzzy states and fuzzy alternatives.

	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
$\underset{\sim}{A}_1$	u_{11}	u_{12}	u_{13}
$\underset{\sim}{A}_2$	u_{21}	u_{22}	u_{23}
$\underset{\sim}{A}_3$	u_{31}	u_{32}	u_{33}

collection of orthogonal fuzzy sets on X, $\Phi = \{\underset{\sim}{M}_1, \underset{\sim}{M}_2, \ldots \underset{\sim}{M}_g\}$, that are defined on X. We can now derive the posterior probabilities of fuzzy states $\underset{\sim}{F}_s$, given probabilistic information (Equation [9.52a]), x_r, and fuzzy information, $\underset{\sim}{M}_t$ (Equation [9.52b]), as follows:

$$p(\underset{\sim}{F}_s|x_k) = \frac{\sum_{i=1}^{n} \mu_{\underset{\sim}{F}_s}(s_i) p(x_k|s_i) p(s_i)}{p(x_k)}. \tag{9.52a}$$

$$p(\underset{\sim}{F}_s|\underset{\sim}{M}_t) = \frac{\sum_{i=1}^{n} \sum_{k=1}^{r} \mu_{\underset{\sim}{F}_s}(s_i) \mu_{\underset{\sim}{M}} t(x_k) p(x_k|s_i) p(s_i)}{\sum_{i=1}^{n} \sum_{k=1}^{r} \mu_{\underset{\sim}{M}_t}(x_k) p(x_k).} \tag{9.52b}$$

Similarly, the expected utility, given the probabilistic (Equation (9.53a)) and fuzzy (Equation (9.53b)) information, is then

$$E\left(u_j|x_k\right) = \sum_{s=1}^{3} u_{js} p(\underset{\sim}{F}_s|x_k), \tag{9.53a}$$

$$E\left(u_j|x_k\right) = \sum_{s=1}^{3} u_{js} p(\underset{\sim}{F}_s|\underset{\sim}{M}_t), \tag{9.53b}$$

where the maximum conditional expected utility for probabilistic (Equation (9.54a)) and fuzzy (Equation (9.54b)) information is

$$E(u_{x_k}^*) = \max_j E\left(u_j|x_k\right). \tag{9.54a}$$

$$E(u_{\underset{\sim}{M}_t}^*) = \max_j E(u_j|\underset{\sim}{M}_t). \tag{9.54b}$$

Finally, the unconditional expected utilities for fuzzy states and probabilistic information (Equation (9.55a)), or fuzzy information (Equation (9.55b)), will be

$$E\left(u_x^*\right) = \sum_{k=1}^{r} E(u_{x_k}^*) p(x_k). \tag{9.55a}$$

$$E\left(u_\Phi^*\right) = \sum_{t=1}^{g} E(u_{\underset{\sim}{M}_t}^*) p(\underset{\sim}{M}_t). \tag{9.55b}$$

The expected utility given in Equation (9.55) now enables us to compute the value of the fuzzy information, within the context of fuzzy states of nature, for probabilistic information (Equation (9.35)) and fuzzy information (Equation (9.56)):

$$V(\Phi) = E\left(u_\Phi^*\right) - E(u^*). \tag{9.56}$$

If the new fuzzy information is hypothetically *perfect* (because this would represent the ability to predict a fuzzy state $\underset{\sim}{F}_s$ without dissonance, it is admittedly an untenable boundary condition on the analysis), denoted Φ_p, then we can compute the maximum expected utility of fuzzy *perfect* information using Equation (9.57). The expected utility of ith alternative $\underset{\sim}{A}_i$ for fuzzy perfect information on state $\underset{\sim}{F}_s$ becomes (from Table 9.8)

$$u(\underset{\sim}{A}_i|\underset{\sim}{F}_s) = u(\underset{\sim}{A}_i, \underset{\sim}{F}_s). \tag{9.57}$$

Therefore, the optimum fuzzy action, $\underset{\sim}{A}^*_{\underset{\sim}{F}_s}$, is defined as

$$u(\underset{\sim}{A}^*_{\underset{\sim}{F}_s}|\underset{\sim}{F}_s) = \max_i u(\underset{\sim}{A}_i, \underset{\sim}{F}_s). \tag{9.58}$$

Hence, the total expected utility for fuzzy perfect information is

$$E(u^*_{\Phi_p}) = \sum_{j=1}^{3} u(\underset{\sim}{A}^*_{\underset{\sim}{F}_s}|\underset{\sim}{F}_s) p(\underset{\sim}{F}_s), \tag{9.59}$$

where $p(\underset{\sim}{F}_s)$ are the prior probabilities of the fuzzy states of nature given by Equation (9.50). The result of Equation (9.55b) for the fuzzy *perfect* case would be denoted $E(u^*_{\Phi_p})$, and the value of the fuzzy *perfect* information would be

$$V(\Phi_p) = E(u^*_{\Phi_p}) - E(u^*). \tag{9.60}$$

Tanaka Okuda, and Asai (1976) have proved that the various values of information conform to the following inequality expression:

$$V(\Phi_p) \geq V(x_p) \geq V(x) \geq V(\Phi) \geq 0. \tag{9.61}$$

The inequalities in Equation (9.61) are consistent with our intuition. The ordering, $V(x) \geq V(\Phi)$, is as a result of the fact that information Φ is characterized by fuzziness *and* randomness. The ordering, $V(x_p) \geq V(x)$, is true because x_p is better information than x; it is *perfect*. The ordering, $V(x_p) \geq V(x_p)$, is created by the fact that the uncertainty expressed by the probability $P(\underset{\sim}{F}_i)$ still remains, even if we know the true state, s_i; hence, our interest is not in the crisp states of nature, S, but rather in the fuzzy states, $\underset{\sim}{F}$, that are defined on S.

To illustrate the development of Equations (9.48) to (9.61) in expanding a decision problem to consider fuzzy information, fuzzy states, and fuzzy actions in the Bayesian decision framework, the following example in computer engineering is provided.

Example 9.9

One of the decisions your project team faces with each new computer product is what type of printed circuit board (PCB) will be required for the unit. Depending on the density of tracks (metal interconnect traces on the PCB that act like wire to connect components together), which is related to the density of the components, we may use a single-layer PCB, a double-layer PCB, a four-layer PCB, or a six-layer PCB. A PCB layer is a two-dimensional plane of interconnecting tracks. The number of layers on a PCB is the number of parallel interconnection layers in the PCB. The greater the density of the interconnections in the design, the greater the number of layers required to fit the design onto a PCB of given size. One measure of board track density is the number of nodes required in the design. A node is created at a location in the circuit where two or more lines (wires, tracks) meet. The decision process will comprise the following steps.

1. *Define the fuzzy states of nature:* The density of the PCB is defined as three fuzzy sets on the singleton states S = $(s_1, s_2, s_3, s_4, s_5) = (s_i)$, $i = 1, 2, \ldots, 5$, where i defines the states in terms of a percentage of our most dense (in terms of components and interconnections) PCB. So, your team defines $s_1 = 20\%$, $s_2 = 40\%$, $s_3 = 60\%$, $s_4 = 80\%$, and $s_5 = 100\%$ of the density of the densest PCB; these are singletons on the universe of relative densities. Further, you define the following three fuzzy states that are defined on the universe of relative density states S:

   $\underset{\sim}{F}_1$ = low-density PCB
   $\underset{\sim}{F}_2$ = medium-density PCB
   $\underset{\sim}{F}_3$ = high-density PCB.

2. *Define fuzzy alternatives:* Your decision alternative will represent the type of the PCB we decide to use as follows (these actions are admittedly not very fuzzy, but in general they can be):

   $\underset{\sim}{A}_1$ = use a 2-layer PCB for the new design
   $\underset{\sim}{A}_2$ = use a 4-layer PCB for the new design
   $\underset{\sim}{A}_3$ = use a 6-layer PCB for the new design.

3. *Define new data samples* (*information*): The universe X = $(x_1, x_2, \ldots, x_5)$ represents the "measured number of nodes in the PCB schematic"; that is, the additional information is the measured number of nodes of the schematic, which can be calculated by a *schematic capture system*. You propose the following discrete values for number of nodes:

   $x_1 = 100$ nodes
   $x_2 = 200$ nodes
   $x_3 = 300$ nodes
   $x_4 = 400$ nodes
   $x_5 = 500$ nodes.

4. *Define orthogonal fuzzy information system:* You determine that the ambiguity in defining the density of nodes can be characterized by three linguistic information sets as $(\underset{\sim}{M}_1, \underset{\sim}{M}_2, \underset{\sim}{M}_3)$, where

$\underset{\sim}{M}_1$ = low number of nodes on PCB [generally < 300 nodes]
$\underset{\sim}{M}_2$ = average (medium) number of nodes on PCB [about 300 nodes]
$\underset{\sim}{M}_3$ = high number of nodes on PCB [generally > 300 nodes].

5. *Define the prior probabilities:* The prior probabilities of the singleton densities (states) are as follows:

$p(s_1) = 0.2$
$p(s_2) = 0.3$
$p(s_3) = 0.3$
$p(s_4) = 0.1$
$p(s_5) = 0.1.$

The preceding numbers indicate that moderately dense boards are the most probable, followed by low-density boards, and high- to very high–density boards are the least probable.
6. *Identify the utility values:* You propose the nondimensional utility values shown in Table 9.9 to represent the fuzzy alternative-fuzzy state relationships. The highest utility in Table 9.9 is achieved by the selection of a six-layer PCB for a high-density PCB because the board layout is achievable. The same high-utility level of 10 is also achieved by selecting the two-layer PCB in conjunction with the low-density PCB because a two-layer PCB is cheaper than a four- or six-layer PCB. The lowest utility is achieved by the selection of a two-layer PCB for a high-density PCB; because the layout cannot be done, it will not fit. The second to lowest utility is achieved when a six-layer PCB is chosen, but the design is of low density, so you are wasting money.
7. *Define membership values for each orthogonal fuzzy state:* The fuzzy sets in Table 9.10 satisfy the orthogonality condition, for the sum of each column equals 1, $\sum \text{column} = \sum_s \mu_{\underset{\sim}{F}_s}(s_i) = 1$.
8. *Define membership values for each orthogonal fuzzy set on the fuzzy information system:* In Table 9.11, $\sum \text{column} = \sum_t \mu_{\underset{\sim}{M}_t}(x_i) = 1$; hence, the fuzzy sets are orthogonal.

Table 9.9 Utilities for fuzzy states and alternatives.

	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
$\underset{\sim}{A}_1$	10	3	0
$\underset{\sim}{A}_2$	4	9	6
$\underset{\sim}{A}_3$	1	7	10

Table 9.10 Orthogonal fuzzy sets for fuzzy states.

	s_1	s_2	s_3	s_4	s_5
$\underset{\sim}{F}_1$	1	0.5	0	0	0
$\underset{\sim}{F}_2$	0	0.5	1	0.5	0
$\underset{\sim}{F}_3$	0	0	0	0.5	1

Table 9.11 Orthogonal fuzzy sets for fuzzy information.

	x_1	x_2	x_3	x_4	x_5
$\underset{\sim}{M}_1$	1	0.4	0	0	0
$\underset{\sim}{M}_2$	0	0.6	1	0.6	0
$\underset{\sim}{M}_3$	0	0	0	0.4	1

Table 9.12 Conditional probabilities $p(x_k|s_i)$ for uncertain information.

	x_1	x_2	x_3	x_4	x_5
$p(x_k|s_1)$	0.44	0.35	0.17	0.04	0
$p(x_k|s_2)$	0.26	0.32	0.26	0.13	0.03
$p(x_k|s_3)$	0.12	0.23	0.30	0.23	0.12
$p(x_k|s_4)$	0.03	0.13	0.26	0.32	0.26
$p(x_k|s_5)$	0	0.04	0.17	0.35	0.44

Table 9.13 Conditional probabilities $p(x_k|s_i)$ for fuzzy perfect information.

	x_1	x_2	x_3	x_4	x_5
$p(x_k|s_1)$	1	0	0	0	0
$p(x_k|s_2)$	0	1	0	0	0
$p(x_k|s_3)$	0	0	1	0	0
$p(x_k|s_4)$	0	0	0	1	0
$p(x_k|s_5)$	0	0	0	0	1

9. *Define the conditional probabilities (likelihood values) for the uncertain information:* Table 9.12 shows the conditional probabilities for uncertain (probabilistic) information; note that the sum of elements in each row equals unity.
10. *Define the conditional probabilities (likelihood values) for the probabilistic perfect information:* Table 9.13 shows the conditional probabilities for probabilistic perfect information; note that the sum of elements in each row equals unity and that each column has only one entry (i.e., no dissonance).

You are now ready to compute the values of information for this decision process involving fuzzy states, fuzzy alternatives, and fuzzy information.

Case 1: Crisp States and Actions

1. **Utility and optimum decision given no information.** Before initiating the noinformation case, we must define the nondimensional utility values for this nonfuzzy state situation. Note that the utility values given in Table 9.14 compare the fuzzy alternatives to the singleton states (s_i), as opposed to the fuzzy states for which the utility is defined in Table 9.9. The expected values for this case are determined by using Equation (9.28), for example,

$$\begin{aligned} E(u_1) &= (10)(0.2) + (8)(0.3) + \cdots + (2)(0.1) \\ &= 6.4. \end{aligned}$$

   Similarly, $E(u_2) = 6.3$ and $E(u_3) = 4.4$. Hence, the optimum decision, given no information and with crisp (singleton) states, is alternative 1, $\underset{\sim}{A}_1$, that is, $E(u_1) = 6.4$.
2. **Utility and optimal decision given uncertain and perfect information.**
   a. *Probabilistic (uncertain) information:* Table 9.15 summarizes the values of the marginal probability $p(x_k)$, the posterior probabilities, and the maximum expected values for the uncertain case. The values in Table 9.15 have been calculated as in the preceding computation. For example, the marginal probabilities are calculated using Equation (9.31):

$$\begin{aligned} p(x_1) &= (0.44)(0.2) + (0.26)(0.3) + (0.12)(0.3) + (0.03)(0.1) \\ &= 0.205. \\ p(x_3) &= (0.17)(0.2) + (0.26)(0.3) + (0.3)(0.3) + (0.26)(0.1) + (0.17)(0.1) \\ &= 0.245. \end{aligned}$$

Table 9.14 Utility values for crisp states.

	s_1	s_2	s_3	s_4	s_5
$\underset{\sim}{A}_1$	10	8	6	2	0
$\underset{\sim}{A}_2$	4	6	9	6	4
$\underset{\sim}{A}_3$	1	2	6	8	10

Table 9.15 Computed values for uncertain case (nonfuzzy states).

	x_1	x_2	x_3	x_4	x_5
$p(x_k)$	0.205	0.252	0.245	0.183	0.115
$p(s_1 \mid x_k)$	0.429	0.278	0.139	0.044	0.0
$p(s_2 \mid x_k)$	0.380	0.381	0.318	0.213	0.078
$p(s_3 \mid x_k)$	0.176	0.274	0.367	0.377	0.313
$p(s_4 \mid x_k)$	0.015	0.052	0.106	0.175	0.226
$p(s_5 \mid x_k)$	0.0	0.016	0.069	0.191	0.383
$E(u^* \mid x_k)$	8.42	7.47	6.68	6.66	7.67
$a_j \mid a_k$	1	1	2	2	3

The posterior probabilities are calculated using Equation (9.30), the conditional probabilities, and the prior probabilities; for example,

$$p(s_1|x_3) = \frac{(0.17)(0.2)}{(0.245)} = 0.139.$$

$$p(s_3|x_2) = \frac{(0.23)(0.3)}{(0.245)} = 0.274.$$

$$p(s_5|x_4) = \frac{(0.35)(0.1)}{(0.183)} = 0.191.$$

The conditional expected utilities are calculated using Equations (9.32) and (9.33); for example, for datum x_1,

$$\begin{aligned} E(u_1|x_1) &= (0.429)(10) + (0.380)(8) + (0.176)(6) + (0.015)(2) + (0)(0) \\ &= 8.42. \\ E(u_2|x_1) &= (0.429)(4) + (0.380)(6) + (0.176)(9) + (0.015)(6) + (0)(4) \\ &= 5.67. \\ E(u_3|x_1) &= (0.429)(1) + (0.380)(2) + (0.176)(6) + (0.015)(8) + (0)(10) \\ &= 2.36. \end{aligned}$$

Therefore, the optimum decision for datum x_1, given uncertain information with crisp states, is

$$E(u^*|x_1) = \max(8.42,\ 5.67,\ 2.36. = 8.42(\text{choose action } \underset{\sim}{A}_1).$$

Now, using Equation (9.34) to calculate the overall (for all data x_i) unconditional expected utility for the uncertain information, we get

$$\begin{aligned} E\left(u_x^*\right) &= (8.42)(0.205) + (7.47)(0.252) + (6.68)(0.245) + \cdots + (7.67)(0.115) \\ &= 7.37. \end{aligned}$$

The value of the uncertain information, using Equation (9.35), is

$$V(x) = 7.37 - 6.4 = 0.97.$$

b. *Probabilistic perfect information:* Using the same utility values as before, and conditional probabilities as defined in Table 9.12, the marginal probabilities, posterior probabilities, and the expected values are shown in Table 9.16. The unconditional expected utility for probabilistic perfect information is given as

$$\begin{aligned} E\left(u_{x_p}^*\right) &= (10)(0.2) + (8)(0.3) + \cdots + (10)(0.1) \\ &= 8.9, \end{aligned}$$

Table 9.16 Computed quantities for perfect information and crisp states.

	x_1	x_2	x_3	x_4	x_5
$p(x_k)$	0.20	0.30	0.30	0.10	0.10
$p(s_1\|x_k)$	1.0	0.0	0.0	0.0	0.0
$p(s_2\|x_k)$	0.0	1.0	0.0	0.0	0.0
$p(s_3\|x_k)$	0.0	0.0	1.0	0.0	0.0
$p(s_4\|x_k)$	0.0	0.0	0.0	1.0	0.0
$p(s_5\|x_k)$	0.0	0.0	0.0	0.0	1.0
$E(u^*\|x_k)$	10.0	8.0	9.0	8.0	10.0
$a_j\|a_k$	1	1	2	3	3

and the value of the probabilistic perfect information from Equation (9.35) is

$$V\left(x_p\right) = 8.9 - 6.4 = 2.5.$$

Case 2: Fuzzy States and Actions

1. **Utility and optimum decision given no information.** The utility values for this case are shown in Table 9.9. We calculate the prior probabilities for the fuzzy states using Equation (9.50). For example,

$$\begin{aligned} p(\underset{\sim}{F}_1) &= (1)(0.2) + (0.5)(0.3) + (0)(0.3) + (0)(0.1) + (0)(0.1) \\ &= 0.35. \end{aligned}$$

Similarly, $p(\underset{\sim}{F}_2) = 0.5$ and $p(\underset{\sim}{F}_3) = 0.15$. Therefore, the expected utility is given by Equation (9.49) as

$$E\left(u_j\right) = \begin{bmatrix} 5 \\ 6.8 \\ 5.35 \end{bmatrix}.$$

The optimum expected utility of the fuzzy alternatives (actions) for the case of no information using Equation (9.51) is

$$E(u^*) = 6.8,$$

so alternative $\underset{\sim}{A}_2$ is the optimum choice.

2. **Utility and optimum decision given uncertain and perfect information.**
   a. *Probabilistic (uncertain) information:* Table 9.17 lists the posterior probabilities as determined by Equation (9.52a). For example,

$$p(\underset{\sim}{F}_1|x_1) = \frac{(1)(0.44)(0.2) + (0.5)(0.26)(0.3)}{0.205} = 0.620.$$

Table 9.17 Posterior probabilities for probabilistic information with fuzzy states.

	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
x_1	0.620	0.373	0.007
x_2	0.468	0.49	0.042
x_3	0.298	0.58	0.122
x_4	0.15	0.571	0.279
x_5	0.039	0.465	0.496

Table 9.18 Expected utilities for fuzzy alternatives with probabilistic information.

	$\underset{\sim}{A}_1$	$\underset{\sim}{A}_2$	$\underset{\sim}{A}_3$
x_1	7.315	5.880	3.305
x_2	6.153	6.534	4.315
x_3	4.718	7.143	5.58
x_4	3.216	7.413	6.934
x_5	1.787	7.317	8.252

The other values are calculated in a similar manner and are shown in Table 9.17. The expected utility values for each of the x_k are now calculated using Equation (9.53a), and these values are given in Table 9.18. The optimum expected utilities for each alternative are found by using Equation (9.54a),

$$E(u^*_{x_k}) = \max_j E\left(u_j|x_k\right) = \{7.315,\ 6.534,\ 7.143,\ 7.413,\ 8.252\},$$

where the optimum choice associated with this value is obviously alternative $\underset{\sim}{A}_3$. Finally, the expected utility, given by Equation (9.55), is calculated to be

$$\begin{aligned} E\left(u^*_{\Phi}\right) &= \sum_{k=1}^{r} E(u^*_{x_k})p(x_k) \\ &= (7.315)(0.205) + (6.534)(0.252) + (7.143)(0.245) \\ &\quad + (7.413)(0.183) + (8.252)(0.115) = 7.202. \end{aligned}$$

The value of the probabilistic uncertain information for fuzzy states is

$$V(x) = 7.202 - 6.8 = 0.402.$$

Table 9.19 Posterior probabilities for probabilistic *perfect* information with fuzzy states.

	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
x_1	1.0	0.0	0.0
x_2	0.5	0.5	0.0
x_3	0.0	1.0	0.0
x_4	0.0	0.5	0.5
x_5	0.0	0.0	1.0

Table 9.20 Expected utilities for fuzzy alternatives with probabilistic *perfect* information.

	$\underset{\sim}{A}_1$	$\underset{\sim}{A}_2$	$\underset{\sim}{A}_3$
x_1	10.0	4.0	1.0
x_2	6.5	6.5	4.0
x_3	3.0	9.0	7.0
x_4	1.5	7.5	8.5
x_5	0.0	6.0	10.0

b. *Probabilistic perfect information:* Table 9.19 lists the posterior probabilities as determined in Equation (9.52a). For example,

$$p(\underset{\sim}{F}_1|x_1) = [(1)(1)(0.2) + (0.5)(0)(0.3)/(0.2) = 1.0.$$

The other values are calculated in a similar manner and are shown in Table 9.19. The expected utility values for each of the x_k are now calculated using Equation (9.53a), and these values are given in Table 9.20. The optimum expected utilities for each alternative are found by using Equation (9.54a),

$$E(u^*_{x_k}) = \max_j E\left(u_j|x_k\right) = \{10.0,\ 6.5,\ 9.0,\ 8.5,\ 10.0\},$$

where it is not clear which alternative is the optimum choice (i.e., there is a tie between alternatives 1 and 3). Finally, the expected utility, given by Equation (9.37), is calculated to be

$$\begin{aligned} E(u^*_{x_p}) &= \sum_{k=1}^{r} E\left(u^*_{x_p}|x_k\right)p(x_k) \\ &= (10.0)(0.2) + (6.5)(0.3) + (9.0)(0.3) + (8.5)(0.1) + (10.0)(0.1) \\ &= 8.5. \end{aligned}$$

The value of the probabilistic *perfect* information for fuzzy states is

$$V\left(x_p\right) = 8.5 - 6.8 = 1.7.$$

c. *Fuzzy information:* For the hypothetical fuzzy information, Table 9.21 summarizes the results of the calculations using Equation (9.52b). An example calculation is shown here:

Table 9.21 Posterior probabilities for fuzzy information with fuzzy states.

	$\underset{\sim}{M}_1$	$\underset{\sim}{M}_2$	$\underset{\sim}{M}_3$
$\underset{\sim}{F}_1$	0.570	0.317	0.082
$\underset{\sim}{F}_2$	0.412	0.551	0.506
$\underset{\sim}{F}_3$	0.019	0.132	0.411

Table 9.22 Posterior probabilities for fuzzy alternatives with fuzzy information.

	$\underset{\sim}{M}_1$	$\underset{\sim}{M}_2$	$\underset{\sim}{M}_3$
$\underset{\sim}{A}_1$	6.932	4.821	2.343
$\underset{\sim}{A}_2$	6.096	7.019	7.354
$\underset{\sim}{A}_3$	3.638	5.496	7.740

$$p(\underset{\sim}{F}_1|\underset{\sim}{M}_1) = [(1)(1)(0.44)(0.2) + (1)(0.4)(0.35)(0.2) + (0.5)(1)(0.26)(0.3) \\ + (0.5)(0.4)(0.32)(0.3)] \div [(1)(0.205) + (0.4)(0.252)] = 0.57.$$

Similarly, Table 9.22 summarizes the calculations of the expected utilities using Equation (9.53b).

Now, using Equation (9.54b), we find that the optimum expected utility for each of the fuzzy states is

$$E(u^*_{\underset{\sim}{M}_t}) = \max_j E(u_j|\underset{\sim}{M}_t) = \{6.932,\ 7.019,\ 7.740\},$$

where the optimum choice is again $\underset{\sim}{A}_3$. The marginal probabilities of the fuzzy information sets are calculated using Equation (9.39); for example, using the marginal probabilities from Table 9.15 and the fuzzy information from Table 9.11, we find

$$p(\underset{\sim}{M}_1) = (1.0)(0.205) + (0.4)(0.252) = 0.306,$$

and, along with the other two marginal probabilities, we get

$$p(\underset{\sim}{M}_t) = \begin{bmatrix} 0.306 \\ 0.506 \\ 0.188 \end{bmatrix}.$$

The unconditional expected utility using Equation (9.55b) is

$$E\left(u^*_\Phi\right) = \sum_{t=1}^{g} E(u^*_{\underset{\sim}{M}_t})p(\underset{\sim}{M}_t) = 7.128.$$

and the value of the perfect information for fuzzy states is $V(\Phi) = 7.128 - 6.8 = 0.328$

Table 9.23 Expected utilities for fuzzy alternatives with fuzzy perfect information.

	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
$\underset{\sim}{A}_1$	10.0	3.0	0.0
$\underset{\sim}{A}_2$	4.0	9.0	6.0
$\underset{\sim}{A}_3$	1.0	7.0	10.0

d. *Fuzzy perfect information:* Table 9.23 summarizes the calculations of the expected utilities using Equation (9.57). Note in Table 9.23 that the expected utilities are the same as the utilities in Table 9.9; this identity arises because the information is presumed perfect, and the conditional probability matrix, $p(\underset{\sim}{F}_s|\underset{\sim}{M}_t)$, in Equation (9.53b), is the identity matrix.

Now, using Equation (9.58), we find that the optimum expected utility for each of the fuzzy states is

$$u(A^*_{\underset{\sim}{F}_s}|\underset{\sim}{F}_s) = \max_i u(\underset{\sim}{A}_i, \underset{\sim}{F}_s) = \{10.0,\ 9.0,\ 10.0\}.$$

Finally, using the previously determined prior probabilities of the fuzzy states, $p(\underset{\sim}{F}_s)$—see section (i)—we see that the unconditional expected utility using Equation (9.59) is

$$E(u^*_{\Phi_p}) = \sum_{j=1}^{3} u(\mathrm{A}^*_{\underset{\sim}{F}_s}|\underset{\sim}{F}_s)p(\underset{\sim}{F}_s) = 10(0.35) + 9(0.5) + 10(0.15) = 9.5,$$

where the value of the fuzzy perfect information for fuzzy states is

$$V(\Phi_p) = 9.5 - 6.8 = 2.7.$$

Example Summary

A typical decision problem is to decide on a basic policy in a fuzzy environment. This basic policy can be thought of as a fuzzy action. The attributes of such a problem are that there are many states, feasible policy alternatives, and available information. Usually, the utilities for all the states and all the alternatives cannot be formulated because of insufficient data, because of the high cost of obtaining this information, and because of time constraints. On the other hand, a decision maker in top management is generally not concerned with the detail of each element in the decision problem. Mostly, top managers want to decide roughly what alternatives to select as indicators of policy directions. Hence, an approach that can be based on fuzzy states and fuzzy alternatives and that can accommodate fuzzy information is a very powerful tool for making preliminary policy decisions.

The expected utilities and the value of information for the five cases, that is, for no information, probabilistic (uncertain) information, probabilistic perfect information, fuzzy probabilistic (uncertain) information, and fuzzy perfect information, are summarized in Table 9.24. We can see from this table that the ordering of values of information is in accordance with that described in Equation (9.61), that is, $V(\Phi_p) \geq V(x_p) \geq V(x) \geq V(\Phi) \geq 0$. The probabilistic perfect

Table 9.24 Summary of expected utility and value of information for fuzzy states and actions for the example.

Information	Expected utility	Value of information
No information	6.8	–
Probabilistic information, $V(x)$	7.20	0.40
Perfect information, $V(x_p)$	8.5	1.7
Fuzzy probabilistic information, $V(\Phi)$	7.13	0.33
Fuzzy perfect information, $V(\Phi_p)$	9.5	2.7

information ($V(x_p) = 1.70$) has a value much higher than the probabilistic information ($V(x) = 0.40$). The decision maker needs to ascertain the cost of the hypothetical perfect information when compared with the cost of the uncertain information, the latter being more realistic. On the other hand, there is little difference between the value of fuzzy probabilistic information ($V(\Phi) = 0.33$) and that of probabilistic information ($V(x) = 0.40$). This result suggests that the fuzzy probabilistic information is sufficiently valuable compared with the probabilistic information for this problem because fuzzy information generally costs far less than probabilistic (uncertain) information. Finally, the fact that the fuzzy perfect information ($V(\Phi_p) = 2.70$) holds more value than the probabilistic perfect information ($V(x_p) = 1.70$) confirms that our interest is more in the fuzzy states than the crisp states. When utility values, prior probabilities, conditional probabilities, and orthogonal membership values change for any of these scenarios, the elements in Table 9.24 will change and the conclusions derived from them will portray a different situation. The power of this approach is its ability to measure on an ordinal basis the value of the approximate information used in a decision-making problem. When the value of approximate (fuzzy) information approaches that of either probabilistic or perfect information, there is the potential for significant cost savings without reducing the quality of the decision itself.

Summary

The literature is rich with references in the area of fuzzy decision making. This chapter has presented only a few rudimentary ideas in the hope of making the readers to continue their learning in this important area. One of the decision metrics in this chapter represents a philosophical approach where an existing crisp theory, Bayesian decision making, is reinterpreted to accept both fuzzy and random uncertainty. It is important to note that there have been significant challenges to the maximum expected utility theory on which Bayesian decision making is founded. Three violations of the independence axiom of this theory (the Allias paradox, the Bergen paradox, and sensitivity to tail affects) and one difficulty in representing epistemic uncertainty as a probabilistic belief (the Ellsburg paradox) have been reported in the literature (Maes and Faber, 2004). One key problem in Bayesian decision making is that the updating (updating the priors to become posteriors) is not always applied correctly. Psychometric studies have shown (Tversky and Kahneman, 1974) that too little weight is given to prior information and too much importance is given to new data (likelihood function). Recent information tends to take precedence over long-accumulated prior knowledge.

Theoretical developments are expanding the field of fuzzy decision making; for example, multiobjective situations represent an interesting class of problems that plague optimization in decision making (Sakawa, 1993) as do multiattribute decision problems (Baas and Kwakernaak, 1977). This philosophical approach has been extended further where fuzzy utilities have been addressed with fuzzy states (Jain, 1976), and where fuzzy utilities are determined in the presence of probabilistic states (Jain, 1978; Watson, Weiss, and Donnell 1979). Häage (1978) extended the Bayesian scheme to include the possibility distributions (see Chapter 13 for definition of possibility) for the consequences of the decision actions. The other metrics in this chapter extend some specific problems to deal with issues such as fuzzy preference relations, fuzzy objective functions, fuzzy ordering, and fuzzy consensus. In all of these, there is a compelling need to incorporate fuzziness in human decision making, as originally proposed by Bellman and Zadeh (1970). In most decision situations, the goals, constraints, and consequences of the proposed alternatives are not known with precision. Much of this imprecision is not measurable, and not random. The imprecision can be as a result of vague, ambiguous, or fuzzy information. Methods to address this form of imprecision are necessary to deal with many of the uncertainties we deal with in humanistic systems.

References

Adams, T. (1994). Retaining structure selection with unequal fuzzy project-level objectives. *J. Intell. Fuzzy Syst.*, **2** (3), 251–266.

Baas, S., and Kwakernaak, H. (1977). Rating and ranking of multiple-aspect alternatives using fuzzy sets. *Automatica*, **13**, 47–58.

Bellman, R., and Zadeh, L. (1970). Decision making in a fuzzy environment. *Manage. Sci.*, **17**, 141–164.

Bezdek, J., Spillman, B., and Spillman, R. (1978). A fuzzy relation space for group decision theory. *Fuzzy Sets Syst.*, **1**, 255–268.

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications* (New York: Academic Press).

Häage, C. (1978). Possibility and cost in decision analysis. *Fuzzy Sets Syst.*, **1** (2), 81–86.

Jain, R. (1976). Decision-making in the presence of fuzzy variables. *IEEE Trans. Syst., Man, Cybern.*, **6** (10), 698–703.

Jain, R. (1978). *Decision-making in the Presence of Fuzziness and Uncertainty*, in proceedings of the IEEE Conference on Decision Control, New Orleans, pp. 1318–1323.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic* (Upper Saddle River, NJ: Prentice Hall).

Maes, M., and Faber, M. (2004). Issues in utility modeling and rational decision making, in M. Maes and L. Huyse, eds., *Proceedings of 11th IFIP WG 7.5 Reliability and Optimization of Structural Systems*, pp. 95–104. London: Balkema Publishers.

Okuda, T., Tanaka, H., and Asai, K. (1974). Decision making and information in fuzzy events. *Bull. Univ. Osaka Prefect., Ser. A*, **23** (2), 193–202.

Okuda, T., Tanaka, H., and Asai, K. (1978). A formulation of fuzzy decision problems with fuzzy information, using probability measures of fuzzy events. *Inf. Control*, **38** (2), 135–147.

Ross, T., Booker, J., and Parkinson, W. (2003). *Fuzzy Logic and Probability Applications: Bridging the Gap* (Philadelphia, PA: Society for Industrial and Applied Mathematics).

Sakawa, M. (1993). *Fuzzy Sets and Interactive Multiobjective Optimization* (New York: Plenum Press).

Shimura, M. (1973). Fuzzy sets concept in rank-ordering objects. *J. Math. Anal. Appl.*, **43**, 717–733.

Tanaka, H., Okuda, T., and Asai, K. (1976). A formulation of fuzzy decision problems and its application to an investment problem. *Kybernetes*, **5**, 25–30.

Terano, T., Asai, K., and Sugeno, M. (1992). *Fuzzy System Theory and Its Applications* (San Diego, CA: Academic Press).

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Von Neumann, J., and Morgenstern, O. (1944). *Theory of Games and Economical Behavior* (Princeton, NJ: Princeton University Press).
Watson, S., Weiss, J., and Donnell, M. (1979). Fuzzy decision analysis. *IEEE Trans. Syst., Man, Cybern.*, **9** (1), 1–9.
Yager, R. (1981). A new methodology for ordinal multiobjective decisions based on fuzzy sets. *Decis. Sci.*, **12**, 589–600.

Problems

Ordering and Synthetic Evaluation

9.1 Suppose we have three fuzzy sets as described by Zadeh's notation:

$$\underset{\sim}{I}_1 = \left\{ \frac{0.7}{3} + \frac{0.9}{5} + \frac{0.8}{7} \right\}$$

$$\underset{\sim}{I}_2 = \left\{ \frac{0.9}{2} + \frac{1.0}{7} \right\}$$

$$\underset{\sim}{I}_3 = \left\{ \frac{0.6}{6} + \frac{0.7}{8} \right\}.$$

Calculate the truth quantities as done in Example 9.2

9.2 A construction company is looking to buy steel in large amounts for its project. The four factors to be considered are strength, durability, economy, and corrosion resistance. Comparisons were made by members of several steel companies and the following data were determined for three quality criteria: Superior (sup), equivalence (eq) and deficient (def)

$$\underset{\sim}{R} = \begin{array}{r} \\ \textit{strength} \\ \textit{durability} \\ \textit{economy} \\ \textit{corrosion resistance} \end{array} \begin{array}{c} \begin{array}{ccc} \textit{sup} & \textit{eq} & \textit{def} \end{array} \\ \begin{bmatrix} 0.6 & 0.0 & 0.4 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.6 & 0.1 \end{bmatrix} \end{array}$$

Using the weight factors for the steel of a = {0.4, 0.2, 0.3, 0.1} evaluate the quality of the steels.

9.3 A power supply needs to be chosen to go along with an embedded system. Four categories of evaluation criteria are important. The first is the physical size of the power supply. The second is the efficiency of the power supply. The third is the "ripple" voltage of the output of the power supply. This is a measure of how clean the power provided is. The fourth criterion is the peak current provided by the power supply. The following matrix defines the type of power supply required for the embedded system application (very good to very bad):

$$\underset{\sim}{R} = \begin{array}{r} \\ physical\ size \\ efficiency \\ ripple\ voltage \\ peak\ current \end{array} \begin{array}{c} \begin{array}{ccccc} VG & G & F & B & VB \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 & 0 & 0 \\ 0.1 & 0.2 & 0.5 & 0.1 & 0.1 \\ 0.3 & 0.4 & 0.2 & 0.1 & 0 \\ 0 & 0.2 & 0.6 & 0.1 & 0.1 \end{bmatrix} \end{array}$$

From this matrix, one can see that for the embedded system in mind, the power supply's physical size is important as is its ripple voltage. Of lesser importance is its efficiency, and lesser yet, is its peak current. So, a small power supply with clean output voltage is needed. It needs to be somewhat efficient and is not required to provide very much peak current. Evaluate a power supply with the following weight vector: a = {0.6, 0.2, 0.1, 0.1}

Nontransitive Ranking

9.4 An aircraft control system is a totally *nonlinear system* when the final approach and landing of an aircraft are considered. It involves maneuvering flight in an appropriate course to the airport and then along the optimum glide path trajectory to the runway. We know that this path is usually provided by an instrument landing system that transmits two radio signals to the aircraft as a navigational aid. These orthogonal radio beams are known as the *localizer and the glide slope* and are transmitted from the ends of the runway to provide the approaching aircraft with the correct trajectory for landing. The pilot executing such a landing must monitor cockpit instruments that display the position of the aircraft relative to the desired flight path and make appropriate corrections to the controls. Presume that four positions are available to the pilot and that four corrections P_1, P_2, P_3, and P_4 from the actual position P are required to put the aircraft on the correct course. The pairwise comparisons for the four positions are as follows:

$$fP_1(P_1) = 1, fP_1(P_2) = 0.6, fP_1(P_3) = 0.8, fP_1(P_4) = 0.3.$$
$$fP_2(P_1) = 0.4, fP_2(P_2) = 1, fP_2(P_3) = 0.9, fP_2(P_4) = 0.6.$$
$$fP_3(P_1) = 0.3, fP_3(P_2) = 0.7, fP_3(P_3) = 1, fP_3(P_4) = 0.7.$$
$$fP_4(P_1) = 0.9, fP_4(P_2) = 0.8, fP_4(P_3) = 0.2, fP_4(P_4) = 1.$$

Now, from these values, compute the comparison matrix, and determine the overall ranking.

9.5 When designing a radar system for imaging purposes, we frequently need to set priorities in accomplishing certain features. Some features that need to be traded off against each other are as follows:

1. The ability to penetrate foliage and even the ground to some depth;
2. The resolution of the resulting radar image;
3. The size of the antenna required for the radar system;
4. The amount of power required to operate at a given frequency.

It is useful to determine the order of importance of these features in selecting an operating frequency for the radar. Let x_1 represent penetration; x_2, resolution; x_3, antenna size; and x_4, power. A crisp ordering will have trouble resolving the importance of penetration compared to resolution, resolution compared to antenna size, and antenna size compared to penetration. These are entities that can be compared only in a subjective manner, ideal for fuzzy techniques and difficult for crisp techniques.

Let $fx_i\,(x_j)$ be the relative importance of feature x_j with respect to x_i. The comparisons $fx_i\,(x_j)$ are subjectively assigned as follows:

		x_j			
x_i	x_1	1	0.6	0.5	0.9
	x_2	0.5	1	0.7	0.8
	x_3	0.9	0.8	1	0.5
	x_4	0.3	0.2	0.3	1

Develop a comparison matrix and determine the overall ranking of the importance of each feature.

9.6 For a dam, engineers are comparing three turbines for the most efficient electricity generator. The three turbines are labeled as T_1, T_2, and T_3. Note $f_{Tj}(Ti)$ means how beneficial turbine x_i is for the specific design with respect to turbine x_j. Find the ranking of the turbines based on the pairwise values given here.

$$f_{T1}(T_1) = 1 \quad f_{T1}(T_2) = 0.7 \quad f_{T1}(T_3) = 0.9$$
$$f_{T2}(T_1) = 0.4 \quad f_{T2}(T_2) = 1 \quad f_{T2}(T_3) = 0.6$$
$$f_{T3}(T_1) = 0.8 \quad f_{T3}(T_2) = 0.8 \quad f_{T3}(T_3) = 1$$

Fuzzy Preferences and Consensus

9.7 The Environmental Protection Agency (EPA) is faced with the challenge of cleaning up contaminated groundwater at many sites around the country. To ensure an efficient cleanup process, it is crucial to select a firm that offers the best remediation technology at a reasonable cost. The EPA is deciding among four environmental firms. The professional engineers at the EPA compared the four firms and created a consensus matrix, shown here:

$$\underset{\sim}{R} = \begin{bmatrix} 0 & 0.3 & 0.7 & 0.4 \\ 0.7 & 0 & 0.2 & 0.6 \\ 0.3 & 0.8 & 0 & 0.3 \\ 0.6 & 0.4 & 0.7 & 0 \end{bmatrix}$$

What is the distance to M_1^* consensus?

9.8 A firm is going to put proposal for a project to build a parking garage and retail center on the basis of the proposals from different contractors concerning the cost and construction time of the buildings. The following relation was developed:

$$\underset{\sim}{R} = \begin{bmatrix} 0 & 0.4 & 0.6 & 0.5 \\ 0.6 & 0 & 0.7 & 0.3 \\ 0.4 & 0.3 & 0 & 0.4 \\ 0.5 & 0.7 & 0.6 & 0 \end{bmatrix}$$

Calculate the average fuzziness, degree of preference measure and the distance to Type I and Type II. Explain the differences between the distances to the two consensuses.

9.9 An engineering firm is assessing four different framing options for a two-story office building and has come up with a reciprocal relation for the four options based on cost and aesthetic appeal. The four options are S_1, timber and masonry; S_2, cast-in-place concrete; S_3, precast concrete; and S_4, steel.

$$\underset{\sim}{R} = \begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \end{array} \begin{array}{c} \begin{array}{cccc} S_1 & S_2 & S_3 & S_4 \end{array} \\ \begin{bmatrix} 0 & 0.4 & 0.6 & 0.5 \\ 0.6 & 0 & 0.9 & 0.2 \\ 0.4 & 0.1 & 0 & 0.6 \\ 0.5 & 0.8 & 0.4 & 0 \end{bmatrix} \end{array}$$

Calculate the average certainty, degree of preference measure, and the distance to Type I and Type II consensuses.

Multiobjective Decision Making

9.10 A carcinogen, trichloroethylene (TCE), has been detected in soil and groundwater at levels higher than the EPA maximum contaminant levels (MCLs). There is an immediate need to remediate soil and groundwater. Three remediation alternatives—(1) pump and treat with air stripping (PTA), (2) pump and treat with photo-oxidation (PTP), and (3) bioremediation of soil with pump and treat and air stripping (BPTA)—are investigated. The objectives are as follows: cost (O_1), effectiveness (O_2, capacity to reduce the contaminant concentration), duration (O_3), and speed of implementation (O_4). The ranking of the alternatives on each objective is given as follows:

$$\underset{\sim}{O}_1 = \left\{ \frac{0.4}{PTA} + \frac{0.7}{PTP} + \frac{0.6}{BPTA} \right\}$$

$$\underset{\sim}{O}_2 = \left\{ \frac{0.9}{PTA} + \frac{0.4}{PTP} + \frac{1.0}{BPTA} \right\}$$

$$\underset{\sim}{O}_3 = \left\{ \frac{0.8}{PTA} + \frac{0.6}{PTP} + \frac{0.7}{BPTA} \right\}$$

$$\underset{\sim}{O}_4 = \left\{ \frac{0.3}{PTA} + \frac{0.2}{PTP} + \frac{1.0}{BPTA} \right\}$$

The preferences for each objective are P = {0.4, 0.6, 0.3, 0.5}. Determine the optimum choice of a remediation alternative.

9.11 For an underground public water network system the city engineer is trying to decide on the material to be used for the pipes. Among many alternatives available, the engineer reduces the list of available materials to three alternatives: PVC (P), HDPE (H), and Steel (S) pipes. These pipes are defined with four main objectives: Cost, Durability, Corrosion, and Maintenance. So, the engineer defines the following fuzzy quantities:

alternatives A = {P, H, S} = $\{a_1, a_2, a_3\}$
objectives, $\underset{\sim}{O}$ = {Cost, Durability, Corrosion, Maintenance} = $\{\underset{\sim}{O}_1, \underset{\sim}{O}_2, \underset{\sim}{O}_3, \underset{\sim}{O}_4\}$.

The municipality gives its preference for these objectives as P = {0.7, 0.8, 0.6, 0.5}. The relationships between the objectives and the choices are as follows:

$$\underset{\sim}{O}_1 = \left\{\frac{0.8}{P} + \frac{0.7}{H} + \frac{0.3}{S}\right\}$$

$$\underset{\sim}{O}_2 = \left\{\frac{0.7}{P} + \frac{0.8}{H} + \frac{0.8}{S}\right\}$$

$$\underset{\sim}{O}_3 = \left\{\frac{0.9}{P} + \frac{0.8}{H} + \frac{0.1}{S}\right\}$$

$$\underset{\sim}{O}_4 = \left\{\frac{0.8}{P} + \frac{0.6}{H} + \frac{0.2}{S}\right\}$$

What is the best choice of material for the pipes?

9.12 In the tertiary treatment process for wastewater, the disinfection process is an important procedure that focuses on the destruction of disease-causing organisms. There are a lot of disinfection technologies available; of these, three popular methods (alternatives) for disinfecting are to use: chlorine (Cl), ozone (Oz), or UV radiation (UV). A new wastewater treatment plant is to be built and the designers are having difficulty selecting a disinfecting method, and thus elect to use a multiobjective decision approach. It is concluded that the selection of a disinfection method should be based on efficiency and performance (EP), availability of large quantities of the disinfectants and reasonable prices (Av), maintenance and operation (MO), and environmental impact (EV). The sets of alternatives (A), objectives (O), and preferences (P) are shown. Using the ratings given for each objective and the preference specified by the facility owner, make a decision on which disinfection technology to use.

Alternatives = {Cl, Oz, UV} = $\{a_1, a_2, a_3\}$.
Objectives = {EP, Av, MO, EV} = $\{\underset{\sim}{O}_1, \underset{\sim}{O}_2, \underset{\sim}{O}_3, \underset{\sim}{O}_4\}$
Preferences = $\{b_1, b_2, b_3, b_4\}$ = {0.4, 0.6, 0.4, 0.2}.

$$\underset{\sim}{O}_1 = \left\{\frac{0.2}{a1} + \frac{0.9}{a2} + \frac{0.4}{a3}\right\} \quad \underset{\sim}{O}_3 = \left\{\frac{1}{a1} + \frac{0.6}{a2} + \frac{0.8}{a3}\right\}$$

$$\underset{\sim}{O}_2 = \left\{\frac{0.6}{a1} + \frac{1}{a2} + \frac{0.2}{a3}\right\} \quad \underset{\sim}{O}_4 = \left\{\frac{0.7}{a1} + \frac{0.7}{a2} + \frac{0.2}{a3}\right\}$$

9.13 For construction of a steel bridge, a request for proposal (RFP) was posted publically. Among all of the bidders, the government decided on three of them to evaluate more closely. They are Bidder1 (B_1), Bidder2 (B_2) and Bidder3 (B_3). There were four main objectives defined: (1) Lowest bidder (LB) (2) Duration (D) (3) Reputation (R) and (4) Source availability (SA). But there was some disagreement as to the importance of each objective, so the government engineers decided to define two sets of preferences P_1 and P_2

Alternatives = A = $\{B_1, B_2, B_3\}$.
Objectives = $\underset{\sim}{O}$ = {Lowest Bidder, Duration, Reputation, Source Availability} = {LB, D, R, SA}
Preferences, $P_1 = \{b_1, b_2, b_3, b_4\} = \{0.6, 0.8, 0.8, 0.9\}$
$P_2 = \{b_1, b_2, b_3, b_4\} = \{0.4, 0.6, 0.7, 0.5\}$.

The degree of membership of each alternative in objectives is as follows:

$$\underset{\sim}{O}_1 = \left\{\frac{0.3}{b1} + \frac{0.8}{b2} + \frac{0.7}{b3}\right\}, \quad \underset{\sim}{O}_3 = \left\{\frac{0.9}{b1} + \frac{0.6}{b2} + \frac{0.4}{b3}\right\},$$
$$\underset{\sim}{O}_2 = \left\{\frac{0.6}{b1} + \frac{0.5}{b2} + \frac{0.7}{b3}\right\}, \quad \underset{\sim}{O}_4 = \left\{\frac{0.8}{b1} + \frac{1.0}{b2} + \frac{0.1}{b3}\right\}.$$

Find the decision for each set of preferences.

9.14 In the city of Calgary, Alberta, subdivisions constructed before 1970 were not required to retain overland stormwater flow on a site during major storm events to the level that has been accepted under current design criteria. To properly mitigate flooding and property damage in older subdivisions prone to flooding, they are being upgraded based on technical feasibility and public acceptance of the work. Presently, a subdivision is being considered for an upgrade of its stormwater sewer. It has been determined that there are two different methods to achieve the mitigation, either larger storm sewers have to be installed through the affected neighborhoods (pipe network) or stormwater retention facilities (pond) have to be built close enough to the neighborhood to reduce the flood threat. The mitigation alternatives (A) and the considered impacts or objectives (O) are described as follows:

Alternatives: A = {pipe, pond}.
Objectives: Additional land required (O_1), cost (O_2), flood damage (O_3), public acceptance (O_4), and environmental constraints (O_5).

On the basis of previous experience with other subdivisions, the city design engineer has determined the following ratings for this subdivision:

$$\underset{\sim}{O}_1 = \left\{\frac{0.8}{pipe} + \frac{0.5}{pond}\right\}, \quad \underset{\sim}{O}_2 = \left\{\frac{0.9}{pipe} + \frac{0.4}{pond}\right\},$$
$$\underset{\sim}{O}_3 = \left\{\frac{0.6}{pipe} + \frac{0.8}{pond}\right\}, \quad \underset{\sim}{O}_4 = \left\{\frac{0.4}{pipe} + \frac{0.9}{pond}\right\}, \quad \underset{\sim}{O}_5 = \left\{\frac{0.7}{pipe} + \frac{0.4}{pond}\right\}.$$

The city council has given the administration the following preference values for each objective. Using the above objectives and preferences determine which system to use for this subdivision:

$$P = \{b_1, b_2, b_3, b_4, b_5\} = \{0.6, 0.5, 0.6, 0.8, 0.6\}.$$

Bayesian Decision Making

9.15 A company produces PCBs as a subcomponent for a system that is integrated (with other subcomponents) by another company. The system integration company cannot give precise information on how many PC boards it needs other than "approximately 10,000." It may require more or less than this number. The PC board manufacturer has three courses of action from which to choose: (1) build somewhat less than 10,000 PC boards, $\underset{\sim}{A}_1$; (2) build approximately 10,000 PC boards, $\underset{\sim}{A}_2$; and (3) build somewhat more than 10,000 PC boards, $\underset{\sim}{A}_3$.

The systems integration company will need the PC boards to meet the demand for its final product. The following are the three fuzzy states of nature:

1. low demand, $\underset{\sim}{D}_1$
2. medium demand, $\underset{\sim}{D}_2$
3. high demand, $\underset{\sim}{D}_3$.

The utility function is given in this table:

	$\underset{\sim}{D}_1$	$\underset{\sim}{D}_2$	$\underset{\sim}{D}_3$
$\underset{\sim}{A}_1$	4	2	-1
$\underset{\sim}{A}_2$	-1	5	2
$\underset{\sim}{A}_3$	-5	2	4

There are six discrete states of nature, s_1–s_6, on which the fuzzy states are defined. The membership functions for the fuzzy states and the prior probabilities $p(s_i)$ of the discrete states are shown in the following table:

	s_1	s_2	s_3	s_4	s_5	s_6
$\mu_{\underset{\sim}{D}1}$	1.0	0.7	0.1	0.0	0.0	0.0
$\mu_{\underset{\sim}{D}2}$	0.0	0.3	0.9	1.0	0.3	0.0
$\mu_{\underset{\sim}{D}3}$	0.0	0	0.0	0.1	0.7	1.0
$p(s_i)$	0.2	0.1	0.4	0.1	0.1	0.1

The demand for the system integrator's product is related to the growth of refineries because the final product is used in refineries. The new samples of refinery growth information are x, and $\underset{\sim}{M}_i$ are the fuzzy sets on this information, defined as

1. low growth, $\underset{\sim}{M}_1$
2. medium growth, $\underset{\sim}{M}_2$
3. high growth, $\underset{\sim}{M}_3$.

	x_1	x_2	x_3	x_4	x_5	x_6
$\mu_{\underset{\sim}{M}_1}$	1.0	0.6	0.2	0.0	0.0	0.0
$\mu_{\underset{\sim}{M}_2}$	0.0	0.4	0.8	0.5	0.3	0.0
$\mu_{\underset{\sim}{M}_3}$	0.0	0.0	0.0	0.5	0.7	1.0

The likelihood values for the probabilistic uncertain information for the data samples are shown here:

	x_1	x_2	x_3	x_4	x_5	x_6
s_1	0.1	0.1	0.5	0.1	0.1	0.1
s_2	0.0	0.0	0.1	0.4	0.4	0.1
s_3	0.1	0.2	0.4	0.2	0.1	0.0
s_4	0.5	0.1	0.0	0.0	0.2	0.2
s_5	0.0	0.0	0.0	0.1	0.3	0.6
s_6	0.1	0.7	0.2	0.0	0.0	0.0

The likelihood values for the probabilistic perfect information for the data samples are shown next:

	x_1	x_2	x_3	x_4	x_5	x_6
s_1	0.0	0.0	1.0	0.0	0.0	0.0
s_2	0.0	0.0	0.0	1.0	0.0	0.0
s_3	0.0	0.0	0.0	0.0	1.0	0.0
s_4	0.1	0.0	0.0	0.0	0.0	0.0
s_5	0.0	0.0	0.0	0.0	0.0	1.0
s_6	0.0	1.0	0.0	0.0	0.0	0.0

For the information just presented, compare the following for perfect and imperfect information:

a. Posterior probabilities of fuzzy state 2 ($\underset{\sim}{D}_2$) given the fuzzy information 3 ($\underset{\sim}{M}_3$).
b. Conditional expected utility for action 1 ($\underset{\sim}{A}_1$) and fuzzy information 2 ($\underset{\sim}{M}_2$).

9.16 In a particular region, a water authority must decide whether to build dikes to prevent flooding in case of excess rainfall. Three fuzzy courses of action may be considered:

1. build a permanent dike ($\underset{\sim}{A}_1$)
2. build a temporary dike ($\underset{\sim}{A}_2$)
3. do not build a dike ($\underset{\sim}{A}_3$).

The sets $\underset{\sim}{A}_1$, $\underset{\sim}{A}_2$, and $\underset{\sim}{A}_3$ are fuzzy sets depending on the type and size of the dike to be built. The utility from each of these investments depends on the rainfall in the region. The crisp states of nature, S = $\{s_1, s_2, s_3, s_4, s_5\}$, are the amount of total rainfall in millimeters in the region. The utility for each of the alternatives has been developed for three levels of rainfall, (1) low ($\underset{\sim}{F}_1$), (2) medium ($\underset{\sim}{F}_2$), and (3) heavy ($\underset{\sim}{F}_3$), which are defined by fuzzy sets on S. The utility matrix may be given as follows:

u_{ij}	$\underset{\sim}{F}_1$	$\underset{\sim}{F}_2$	$\underset{\sim}{F}_3$
$\underset{\sim}{A}_1$	–2	4	12
$\underset{\sim}{A}_2$	1	7	–10
$\underset{\sim}{A}_3$	13	–5	–20

The membership functions of $\underset{\sim}{F}_1$, $\underset{\sim}{F}_2$, and $\underset{\sim}{F}_3$, and the prior probabilities are given here:

	s_1	s_2	s_3	s_4	s_5
$\mu_{\underset{\sim}{F}_1}(s_i)$	1	0.3	0.1	0	0
$\mu_{\underset{\sim}{F}_2}(s_i)$	0	0.7	0.80	0.3	0
$\mu_{\underset{\sim}{F}_3}(s_i)$	0	0	0.1	0.7	1
$P(s_i)$	0.1	0.2	0.2	0.35	0.15

Let X = $\{x_1, x_2, x_3, x_4\}$ be the set of amount of rainfall in the next year. This represents the new information. The conditional probabilities $p(x_j|s_i)$ for probabilistic uncertain information are as given in the following table:

	x_1	x_2	x_3	x_4
s_1	0.7	0.2	0.1	0.0
s_2	0.1	0.7	0.2	0.0
s_3	0.1	0.2	0.7	0.0
s_4	0.0	0.1	0.2	0.7
s_5	0.0	0.0	0.3	0.7

Consider a fuzzy information system,

$$\underset{\sim}{M} = \{\underset{\sim}{M}_1, \underset{\sim}{M}_2, \underset{\sim}{M}_3\},$$

where

$\underset{\sim}{M}_1$ = rainfall is less than approximately 35 mm
$\underset{\sim}{M}_2$ = rainfall is equal to approximately 35 mm
$\underset{\sim}{M}_3$ = rainfall is greater than approximately 35 mm.

The membership functions for the new fuzzy information that satisfy the orthogonality condition are given here:

	x_1	x_2	x_3	x_4
$\mu_{\underset{\sim}{M}_1}(x_i)$	1.0	0.3	0.1	0.0
$\mu_{\underset{\sim}{M}_2}(x_i)$	0.0	0.7	0.8	0.1
$\mu_{\underset{\sim}{M}_3}(x_i)$	0.0	0.0	0.1	0.9

Determine the following:

a. Posterior probabilities for fuzzy state $\underset{\sim}{F}_2$ and fuzzy information $\underset{\sim}{M}_1$, and for fuzzy state $\underset{\sim}{F}_3$ and fuzzy information $\underset{\sim}{M}_3$.
b. Conditional expected utility of building a permanent dike ($\underset{\sim}{A}_1$) when fuzzy information $\underset{\sim}{M}_3$ is given.

9.17 Your design team needs to determine what level of technology to incorporate in a new product. As is usually the case, current technology is least expensive, whereas the most advanced or leading-edge technology is the most expensive. A given technology usually comes down in price with time. The decision cycle of your project is several years. The team must decide what level of technology to incorporate in the product based on the future expected cost. If the technology is still expensive by the time the product goes to the market, the product will not sell. If you do not incorporate the latest affordable technology, your product may not be so advanced as that of the competition and therefore sales may be poor. Consider the following:
Actual discrete states of nature:

s_1: cost is low
s_2: cost is moderate
s_3: cost is high.

Fuzzy actions:

$\underset{\sim}{A}_1$: use current/well-established technology
$\underset{\sim}{A}_2$: use newer/leading-edge/advanced technology.

Fuzzy states on fuzzy information system, μ:

$\underset{\sim}{M}_1$: cost is approximately the cost of implementing with current technology
$\underset{\sim}{M}_2$: cost is approximately two times the cost of the current technology
$\underset{\sim}{M}_3$: cost is approximately 10 times the cost of current technology.

Let $X = \{x_1, x_2, x_3, x_4, x_5\}$ be the set of rates of increase in usage of advanced technology in the next term. Then, we have the following:

Fuzzy states of nature:

$\underset{\sim}{F}_1$: low cost
$\underset{\sim}{F}_2$: medium cost
$\underset{\sim}{F}_3$: high cost.

Prior probabilities:

$$p(s_i) = \begin{array}{c} \begin{bmatrix} 0.30 \\ 0.5 \\ 0.20 \end{bmatrix} \end{array} \begin{array}{l} s_1 \\ s_2 \\ s_3 \end{array}.$$

Utility matrix:

$$u = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{bmatrix} -9 & -5 & 0 \\ 8 & -3 & -10 \end{bmatrix} \end{array} \begin{array}{l} \\ \underset{\sim}{A}_1 \\ \underset{\sim}{A}_2 \end{array}.$$

Membership values for each orthogonal fuzzy state on the actual state system:

$$\mu_{\underset{\sim}{F}} = \begin{array}{c} \begin{array}{ccc} s_1 & s_2 & s_3 \end{array} \\ \begin{bmatrix} 0.7 & 0.1 & 0 \\ 0.3 & 0.8 & 0.2 \\ 0 & 0.1 & 0.8 \end{bmatrix} \end{array} \begin{array}{l} \\ \underset{\sim}{F}_1 \\ \underset{\sim}{F}_2 \\ \underset{\sim}{F}_3 \end{array}.$$

Membership values for each orthogonal fuzzy set on the fuzzy information system:

$$\mu_{\underset{\sim}{M}} = \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{bmatrix} 1 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 1 \end{bmatrix} \end{array} \begin{array}{l} \\ \underset{\sim}{M}_1 \\ \underset{\sim}{M}_2 \\ \underset{\sim}{M}_3 \end{array}.$$

Utility matrix for fuzzy information:

$$u = \begin{array}{c} \begin{array}{ccc} \underset{\sim}{F}_1 & \underset{\sim}{F}_2 & \underset{\sim}{F}_3 \end{array} \\ \begin{bmatrix} -5 & 0 & 5 \\ 10 & 2 & -10 \end{bmatrix} \end{array} \begin{array}{l} \\ \underset{\sim}{A}_1 \\ \underset{\sim}{A}_2 \end{array}.$$

Likelihood values for probabilistic (uncertain) information for the data samples:

$$p(x_i|s_k) = \begin{array}{c} \\ \end{array}\begin{array}{cc} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} & \\ \begin{bmatrix} 0.1 & 0.25 & 0.15 & 0.35 & 0.15 \\ 0.3 & 0.05 & 0.1 & 0.1 & 0.45 \\ 0.2 & 0.4 & 0.35 & 0 & 0.05 \end{bmatrix} & \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \end{array}.$$

Likelihood values for probabilistic perfect information for the data samples:

$$p(x_i|s_k) = \begin{array}{cc} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} & \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0.3 & 0 & 0 & 0 & 0.7 \\ 0 & 0.6 & 0.4 & 0 & 0 \end{bmatrix} & \begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \end{array}.$$

a. Determine the value of information for the fuzzy states and fuzzy actions for uncertain probabilistic information.
b. Determine the value of information for the fuzzy states and fuzzy actions for perfect probabilistic information.

10

Fuzzy Classification and Pattern Recognition

From causes which appear similar, we expect similar effects. This is the sum total of all our experimental conclusions.

David Hume, Scottish philosopher,
Enquiry Concerning Human Understanding, 1748

There is structure in nature. Much of this structure is known to us and is quite beautiful. Consider the natural sphericity of rain drops and bubbles; why do balloons take this shape? How about the elegant beauty of crystals, rhombic solids with rectangular, pentagonal, or hexagonal cross sections? Why do these naturally beautiful, geometric shapes exist? What causes the natural repetition of the mounds of sand dunes? Some phenomena we cannot see directly; for example, the elliptical shape of the magnetic field around the earth, or we can see only when certain atmospheric conditions exist, such as the beautiful and circular appearance of a rainbow or the repetitive patterns of the aurora borealis in the night sky near the North Pole. Some patterns, such as the helical appearance of DNA or the cylindrical shape of some bacteria, have only appeared to us since the advent of extremely powerful electron microscopes. Consider the geometry and colorful patterns of a butterfly's wings; why do these patterns exist in our physical world? The answers to some of these questions are still unknown; many others have been discovered through increased understanding of physics, chemistry, and biology.

Just as there is structure in nature, we believe there is an underlying structure in most of the phenomena we wish to understand. Examples abound in image recognition, molecular biology applications such as protein folding and three-dimensional molecular structure, oil exploration, cancer detection, and many others. For fields dealing with diagnosis, we often seek to find structure in the data obtained from observation. Our observations can be visual, audio, or any of a variety of sensor-based electronic or optical signals. Finding the structure in data is the essence of classification. As the quotation at the beginning of this chapter suggests, our experimental observations lead us to develop relationships between the inputs and outputs of an experiment. As we are able to conduct more experiments, we see the relationships forming some recognizable, or classifiable, structure. By finding structure, we are classifying the data according to similar patterns, attributes, features, and other characteristics. The general area is known as *classification*. The process of classification precedes the process of pattern recognition because the classification process develops the patterns through which the pattern recognition process unveils itself as a powerful automated tool. This chapter is organized into the two sections: classification and pattern recognition.

Fuzzy Classification

In classification, also termed *clustering*, the most important issue is deciding what criteria to classify against. For example, suppose we want to classify people. In describing people we will look at their height, weight, gender, religion, education, appearance, and so on. Many of these features are numerical quantities such as height and weight; other features are simply linguistic descriptors. We can easily classify people according to gender, or one feature. For this classification the criterion is simple: female or male. We might want to classify people into three size categories: small, medium, and large. For this classification we might need only two of the features describing people: height and weight. Here, the classification criterion might be some algebraic combination of height and weight. Suppose we want to classify people according to whether we would want them as *neighbors*. Here, the number of features to be used in the classification is not at all clear, and we might also have trouble developing a criterion for this classification. Nevertheless, a criterion for classification must be prepared before we can segregate the data into definable classes. As is often the case in classification studies, the number and kind of features and the type of classification criteria are choices that are continually changed as the data are manipulated; this iteration continues until we think we have a grouping of the data, which seems plausible from a structural and physical perspective.

This chapter summarizes only two popular methods of classification. The first is classification using equivalence relations (Zadeh, 1971; Bezdek and Harris, 1978). This approach makes use of certain special properties of equivalence relations and the concept of defuzzification known as *lambda-cuts* (λ -cuts) on the relations. The second method of classification is a popular method known as *fuzzy c-means* (FCM), so named because of its close analog in the crisp world, *hard c-means* (HCM) (Bezdek, 1981). This method uses concepts in *n*-dimensional Euclidean space to determine the geometric *closeness* of data points by assigning them to various clusters or classes and then determining the distance between the clusters.

Classification by Equivalence Relations

Crisp Relations

Define a set, $[x_i] = \{x_j | (x_i, x_j) \in \mathrm{R}\}$, as the equivalent class of x_i on a universe of data points, X. This class is contained in a special relation, R, known as an *equivalence relation* (Chapter 3). This class is a set of all elements related to x_i that have the following properties (Bezdek, 1974):

1. $x_i \in [x_i]$ therefore $(x_i, x_i) \in \mathrm{R}$
2. $[x_i] \neq [x_j] \Rightarrow [x_i] \cap [x_j] = \emptyset$
3. $\cup_{x \in \mathrm{X}} [x] = \mathrm{X}$.

The first property is that of reflexivity (Chapter 3), the second property indicates that equivalent classes do not overlap, and the third property simply expresses that the union of all equivalent classes exhausts the universe. Hence, the equivalence relation R can divide the universe X into mutually exclusive equivalent classes, that is,

$$X|R = \{[x]|x \in X\}, \tag{10.1}$$

where X | R is called the *quotient set*. The quotient set of X relative to R, denoted X | R, is the set whose elements are the equivalence classes of X under the equivalence relation R. The cardinality of X | R (i.e., the number of distinct equivalence classes of X under R) is called the rank of the matrix R.

Example 10.1 (Ross, 1995).
Define a universe of integers X = {1, 2, 3,…, 10} and define R as the crisp relation for "the identical remainder after dividing each element of the universe by 3." We have

$$R = \begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10 \end{array} \begin{array}{c} \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

We note that this relation is reflexive, it is symmetric, and as can be determined by inspection (Chapter 3), it is also transitive; hence, the matrix is an equivalence relation. We can group the elements of the universe into classes as follows:

$$\begin{aligned} \{X_1\} &= [1] = [4] = [7] = [10] = \{1, 4, 7, 10\}, && \text{with remainder} = 1; \\ \{X_2\} &= [2] = [5] = [8] = \{2, 5, 8\}, && \text{with remainder} = 2; \\ \{X_3\} &= [3] = [6] = [9] = \{3, 6, 9\}, && \text{with remainder} = 0. \end{aligned}$$

Then, we can show that the classes do not overlap, that is, they are mutually exclusive:

$$\{X_1\} \cap \{X_2\} = \emptyset \text{ and } \{X_2\} \cap \{X_3\} = \emptyset,$$

and that the union of all the classes exhausts (comprises) the universe:

$$\cup[x] = X.$$

The quotient set is then determined to have three classes,

$$X|R = \{(1, 4, 7, 10), (2, 5, 8), (3, 6, 9)\}.$$

Not all relations are equivalent, but if a relation is at least a tolerance relation (i.e., it exhibits properties of reflexivity and symmetry), then it can be converted to an equivalence relation through max–min compositions with itself.

Example 10.2

Suppose you have a collection (universe) of five data points,

$$X = \{x_1, x_2, x_3, x_4, x_5\},$$

and these data points show similarity to one another according to the following tolerance relation, which is reflexive and symmetric:

$$R_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

We see that this tolerance relation is not transitive from the expression

$$(x_1, x_2) \in R_1, \quad (x_2, x_5) \in R_1, \quad \text{but} \quad (x_1, x_5) \notin R_1.$$

As indicated in Chapter 3, any tolerance relation can be reformed into an equivalence relation through at most $n-1$ compositions with itself. In this case, one composition of R_1 with itself results in an equivalence relation,

$$R_1 \circ R_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix} = R.$$

As one can see in the relation, R, there are three classes. The first, second, and fifth columns are identical and the third and fourth columns are each unique. The data points can then be classified into three groups or classes, as delineated:

$$[x_1] = [x_2] = [x_5] = \{x_1, x_2, x_5\} \quad [x_3] = \{x_3\} \quad [x_4] = \{x_4\}.$$

As already illustrated, crisp equivalence relations can be used to divide the universe X into mutually exclusive classes. In the case of fuzzy relations, for all fuzzy equivalence relations, their λ-cuts are equivalent ordinary relations. Hence, to classify data points in the universe using fuzzy relations, we need to find the associated fuzzy equivalence relation.

Example 10.3

Example 3.11 had a tolerance relation, say $\underset{\sim}{R}_t$, describing five data points, that was formed into a fuzzy equivalence relation, $\underset{\sim}{R}$, by composition; this process is repeated here for this classification example.

$$\underset{\sim}{R}_t = \begin{bmatrix} 1 & 0.8 & 0 & 0.1 & 0.2 \\ 0.8 & 1 & 0.4 & 0 & 0.9 \\ 0 & 0.4 & 1 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0.5 \\ 0.2 & 0.9 & 0 & 0.5 & 1 \end{bmatrix} \rightarrow \underset{\sim}{R} = \begin{bmatrix} 1 & 0.8 & 0.4 & 0.5 & 0.8 \\ 0.8 & 1 & 0.4 & 0.5 & 0.9 \\ 0.4 & 0.4 & 1 & 0.4 & 0.4 \\ 0.5 & 0.5 & 0.4 & 1 & 0.5 \\ 0.8 & 0.9 & 0.4 & 0.5 & 1 \end{bmatrix}.$$

By taking λ-cuts of fuzzy equivalent relation R at values of $\lambda = 1, 0.9, 0.8, 0.5$, and 0.4, we get the following:

$$R_1 = \begin{bmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ 0 & & & & 1 \end{bmatrix}, \quad R_{0.9} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad R_{0.8} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{bmatrix},$$

$$R_{0.5} = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix}, \quad R_{0.4} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

where we can see that the clustering of the five data points according to the λ-cut level is as shown in Table 10.1.

We can express the classification scenario described in Table 10.1 with a systematic classification diagram, as shown in Figure 10.1. In the figure, it can be seen that the higher the value of λ, the finer is the classification. That is, as λ gets larger, the tendency of classification tends to approach the trivial case where each data point is assigned to its own class.

Table 10.1 Classification of five data points according to λ-cut level.

λ-cut level	Classification
1.0	$\{x_1\}\{x_2\}\{x_3\}\{x_4\}\{x_5\}$
0.9	$\{x_1\}\{x_2.x_5\}\{x_3\}\{x_4\}$
0.8	$\{x_1.x_2.x_5\}\{x_3\}\{x_4\}$
0.5	$\{x_1.x_2.x_4.x_5\}\{x_3\}$
0.4	$\{x_1.x_2.x_3.x_4.x_5\}$

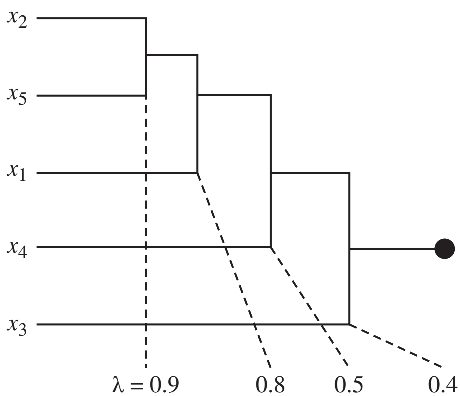

Figure 10.1 Classification diagram for Example 10.3.

Another example in fuzzy classification considers grouping photographs of family members together according to visual similarity in attempting to determine genetics of the family tree when considering only facial image.

Example 10.4 (Tamura, Higuchi, and Tanaka, 1971).
Three families exist, which have a total of 16 people, all of whom are related by blood. Each person has their photo taken, and the 16 photos are mixed. A person not familiar with the members of the three families is asked to view the photographs to grade their resemblance to one another. In conducting this study, the person assigns the similarity relation matrix r_{ij} as shown in Table 10.2. The matrix developed by the person is a tolerance fuzzy relation, but it does not have properties of equivalence,
that is,

$$r_{ij} = 1, \quad \text{for } i = j.$$

$$r_{ij} = r_{ji}.$$

$$r_{ij} \geq \lambda_1 \quad \text{and} \quad r_{jk} \geq \lambda_2, \quad \text{but} \quad r_{ik} < \min(\lambda_1, \lambda_2), \text{ i.e., transitivity does not hold.}$$

For example,

$$r_{16} = 0.5, \quad r_{68} = 0.8, \quad \text{but} \quad r_{18} = 0.4 < 0.5.$$

Table 10.2 Similarity relation matrix, r_{ij}.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.0															
2	0.0	1.0														
3	0.0	0.0	1.0													
4	0.0	0.0	0.4	1.0												
5	0.0	0.8	0.0	0.0	1.0											
6	0.5	0.0	0.2	0.2	0.0	1.0										
7	0.0	0.8	0.0	0.0	0.4	0.0	1.0									
8	0.4	0.2	0.2	0.5	0.0	0.8	0.0	1.0								
9	0.0	0.4	0.0	0.8	0.4	0.2	0.4	0.0	1.0							
10	0.0	0.0	0.2	0.2	0.0	0.0	0.2	0.0	0.2	1.0						
11	0.0	0.5	0.2	0.2	0.0	0.0	0.8	0.0	0.4	0.2	1.0					
12	0.0	0.0	0.2	0.8	0.0	0.0	0.0	0.0	0.4	0.8	0.0	1.0				
13	0.8	0.0	0.2	0.4	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.0	1.0			
14	0.0	0.8	0.0	0.2	0.4	0.0	0.8	0.0	0.2	0.2	0.6	0.0	0.0	1.0		
15	0.0	0.0	0.4	0.8	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.2	0.2	0.0	1.0	
16	0.6	0.0	0.0	0.2	0.2	0.8	0.0	0.4	0.0	0.0	0.0	0.0	0.4	0.2	0.4	1.0

Table 10.3 Equivalence relation.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.0															
2	0.4	1.0														
3	0.4	0.4	1.0													
4	0.5	0.4	0.4	1.0												
5	0.4	0.8	0.4	0.4	1.0											
6	0.6	0.4	0.4	0.5	0.4	1.0										
7	0.4	0.8	0.4	0.4	0.8	0.4	1.0									
8	0.6	0.4	0.4	0.5	0.4	0.8	0.4	1.0								
9	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	1.0							
10	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	1.0						
11	0.4	0.8	0.4	0.4	0.8	0.4	0.8	0.4	0.4	0.4	1.0					
12	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	0.8	0.4	1.0				
13	0.8	0.4	0.4	0.5	0.4	0.6	0.4	0.6	0.5	0.5	0.4	0.5	1.0			
14	0.4	0.8	0.4	0.4	0.8	0.4	0.8	0.4	0.4	0.4	0.8	0.4	0.4	1.0		
15	0.5	0.4	0.4	0.8	0.4	0.5	0.4	0.5	0.8	0.8	0.4	0.8	0.5	0.4	1.0	
16	0.6	0.4	0.4	0.5	0.4	0.8	0.4	0.8	0.5	0.5	0.4	0.5	0.6	0.4	0.5	1.0

By composition, the equivalence relation shown in Table 10.3 is obtained.

When we take a λ-cut of this fuzzy equivalent relation at $\lambda = 0.6$, we get the defuzzified relation shown in Table 10.4.

Four distinct classes are identified:

$$\{1, 6, 8, 13, 16\}, \ \{2, 5, 7, 11, 14\}, \ \{3\}, \ \{4, 9, 10, 12, 15\}.$$

Table 10.4 Defuzzified relation.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1															
2	0	1														
3	0	0	1													
4	0	0	0	1												
5	0	1	0	0	1											
6	1	0	0	0	0	1										
7	0	1	0	0	1	0	1									
8	1	0	0	0	0	1	0	1								
9	0	0	0	1	0	0	0	0	1							
10	0	0	0	1	0	0	0	0	1	1						
11	0	1	0	0	1	0	1	0	0	0	1					
12	0	0	0	1	0	0	0	0	1	1	0	1				
13	1	0	0	0	0	1	0	1	0	0	0	0	1			
14	0	1	0	0	1	0	1	0	0	0	1	0	0	1		
15	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	
16	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1

From this clustering it seems that only photograph number 3 cannot be identified with any of the three families. Perhaps a lower value of λ might assign photograph 3 to one of the other three classes. The other three clusters are all correct in that the members identified in each class are, in fact, the members of the correct families as described in Tamura and colleagues (1971).

Classification using equivalence relations can also be employed to segregate data that are originally developed as a similarity relation using some of the similarity methods developed at the end of Chapter 3. The following problem is an example of this, involving earthquake damage assessment. It was first introduced in Chapter 3 as Example 3.12.

Example 10.5
Five regions have suffered damage from a recent earthquake (Example 3.12). The buildings in each region are characterized according to three damage levels: no damage, medium damage, and serious damage. The percentage of buildings for a given region in each of the damage levels is given in Table 10.5.

Using the cosine amplitude approach, described in Chapter 3, we obtain the following tolerance relation, $\underset{\sim}{R}_1$:

$$\underset{\sim}{R}_1 = \begin{bmatrix} 1 & & & & \\ 0.836 & 1 & & \text{sym} & \\ 0.914 & 0.934 & 1 & & \\ 0.682 & 0.6 & 0.441 & 1 & \\ 0.982 & 0.74 & 0.818 & 0.774 & 1 \end{bmatrix}.$$

Three max–min compositions produce a fuzzy equivalence relation,

Table 10.5 Proportion of buildings damaged in three levels by region.

	Regions				
	x_1	x_2	x_3	x_4	x_5
x_{i1}, ratio with no damage	0.3	0.2	0.1	0.7	0.4
x_{i2}, ratio with medium damage	0.6	0.4	0.6	0.2	0.6
x_{i3}, ratio with serious damage	0.1	0.4	0.3	0.1	0

$$\underset{\sim}{R} = \underset{\sim}{R}_1^3 = \begin{bmatrix} 1 & & & \text{sym} & \\ 0.914 & 1 & & & \\ 0.914 & 0.934 & 1 & & \\ 0.774 & 0.774 & 0.774 & 1 & \\ 0.982 & 0.914 & 0.914 & 0.774 & 1 \end{bmatrix}.$$

Now, if we take λ-cuts at two different values of λ, say λ = 0.914 and λ = 0.934, the following defuzzified crisp equivalence relations and their associated classes are derived:

$$\lambda = 0.914: \quad R_\lambda = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

$$\{x_1, x_2, x_3, x_5\}, \{x_4\}$$

$$\lambda = 0.934: \quad R_\lambda = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\{x_1, x_5\}, \{x_2, x_3\}, \{x_4\}$$

Hence, if we wanted to classify the earthquake damage for purposes of insurance payout into, say, two intensities on the modified Mercalli scale (the Mercalli scale is a measure of an earthquake's strength in terms of average damage the earthquake causes in structures in a given region), then regions 1, 2, 3, and 5 belong to a larger Mercalli intensity and region 4 belongs to a smaller Mercalli intensity (see λ = 0.914). But, if we wanted to have a finer division for, say, three Mercalli scales, we could assign the regions shown in Table 10.6.

Table 10.6 Classification of earthquake damage by region for $\lambda = 0.934$.

Regions	Mercalli intensity
$\{x_4\}$	VII
$\{x_1, x_5\}$	VIII
$\{x_2, x_3\}$	IX

Cluster Analysis

Clustering refers to identifying the number of subclasses of c clusters in a data universe X comprising n data samples, and partitioning X into c clusters ($2 \le c < n$). Note that $c = 1$ denotes rejection of the hypothesis that there are clusters in the data, whereas $c = n$ constitutes the trivial case where each sample is in a "cluster" by itself. There are two kinds of c-partitions of data: hard (or crisp) and soft (or fuzzy). For numerical data, one assumes that the members of each cluster bear more mathematical similarity to each other than to members of other clusters. Two important issues to consider in this regard are how to measure the similarity between pairs of observations and how to evaluate the partitions once they are formed.

One of the simplest similarity measures is distance between pairs of feature vectors in the feature space. If one can determine a suitable distance measure and compute the distance between all pairs of observations, then one may expect that the distance between points in the same cluster will be considerably less than the distance between points in different clusters. Several circumstances, however, mitigate the general utility of this approach, such as the combination of values of incompatible features, as would be the case, for example, when different features have significantly different scales. The clustering method described in this chapter defines "optimum" partitions through a global criterion function that measures the extent to which candidate partitions optimize a weighted sum of squared errors between data points and *cluster centers* in feature space. Many other clustering algorithms have been proposed for distinguishing substructure in high-dimensional data (Bezdek, Grimball, Carson, and Ross, 1986). It is emphasized here that the method of clustering must be closely matched with the particular data under study for successful interpretation of substructure in the data.

Cluster Validity

In many cases, the number c of clusters in the data is known. In other cases, however, it may be reasonable to expect cluster substructure at more than one value of c. In this situation it is necessary to identify the value of c that gives the most plausible number of clusters in the data for the analysis at hand. This problem is known as *cluster validity* (see Duda and Hart, 1973; Bezdek, 1981). If the data used are labeled, there is a unique and absolute measure of cluster validity: the c that is given. For unlabeled data, no absolute measure of clustering validity exists. Although the importance of these differences is not known, it is clear that the features nominated should be sensitive to the phenomena of interest and not to other variations that might not matter to the applications at hand.

c-Means Clustering

Bezdek (1981) developed an extremely powerful classification method to accommodate fuzzy data. It is an extension of a method known as *c-means*, or *hard c-means*, when employed in a crisp classification sense. To introduce this method, we define a sample set of n data samples that we wish to classify:

$$X = \{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3, \ldots, \mathbf{x}_n\}. \tag{10.2}$$

Each data sample, $\mathbf{x}_i$, is defined by m features, that is,

$$\mathbf{x}_i = \{x_{i1},\ x_{i2},\ x_{i3}, \ldots, \mathbf{x}_{im}\}, \tag{10.3}$$

where each $\mathbf{x}_i$ in the universe X is an m-dimensional vector of m elements or m features. Because the m features all can have different units, in general, we have to normalize each of the features to a unified scale before classification. In a geometric sense, each $\mathbf{x}_i$ is a *point* in m-dimensional feature space, and the universe of the data sample, X, is a *point set* with n elements in the sample space.

Bezdek (1981) suggested using an objective function approach for clustering the data into hyperspherical clusters. This idea for hard clustering is shown in three-dimensional feature space in Figure 10.2. In this figure, each cluster of data is shown as a hyperspherical shape with a hypothetical geometric cluster center. The objective function is developed so as to do two things simultaneously: first, minimize the Euclidean distance between each data point in a cluster and its cluster center (a calculated point). Second, maximize the Euclidean distance between cluster centers.

Hard c-Means (HCM)

HCM is used to classify data in a crisp sense. By this we mean that each data point will be assigned to one, and only one, data cluster. In this sense these clusters are also called *partitions*, that is, partitions of the data. Define a family of sets $\{A_i, i = 1, 2, \ldots, c\}$ as a hard c-partition of X, where the following set-theoretic forms apply to the partition:

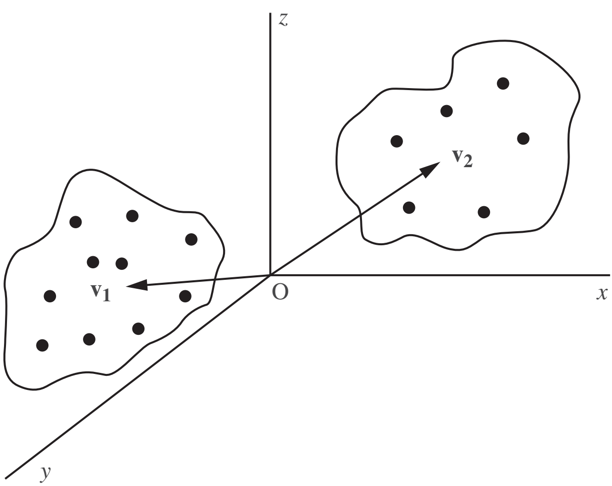

Figure 10.2 Cluster idea with hard c-means.

$$\bigcup_{i=1}^{c} \mathrm{A}_i = \mathrm{X}. \tag{10.4}$$

$$\mathrm{A}_i \cap \mathrm{A}_j = \emptyset \quad \text{all}\, i \neq j. \tag{10.5}$$

$$\emptyset \subset \mathrm{A}_i \subset \mathrm{X} \quad \text{all}\, i. \tag{10.6}$$

Again, where $\mathrm{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ is a finite set space comprising the universe of data samples, and c is the number of classes, or partitions, or clusters, into which we want to classify the data. We note the obvious

$$2 \leq c < n, \tag{10.7}$$

where $c = n$ classes just places each data sample into its own class, and $c = 1$ places all data samples into the same class; neither case requires any effort in classification, and both are intrinsically uninteresting. Equation (10.4) expresses the fact that the set of all classes exhausts the universe of data samples. Equation (10.5) indicates that none of the classes overlap in the sense that a data sample can belong to more than one class. Equation (10.6) simply expresses that a class cannot be empty and it cannot contain all the data samples.

Suppose we have the case where $c = 2$. Equations (10.4) and (10.5) are then manifested in the following set expressions:

$$\mathrm{A}_2 = \overline{\mathrm{A}}_1, \quad \mathrm{A}_1 \cup \overline{\mathrm{A}}_1 = \mathrm{X}, \quad \text{and} \quad \mathrm{A}_1 \cap \overline{\mathrm{A}}_1 = \emptyset.$$

These set expressions are equivalent to the excluded middle axioms (Equation (2.12)).

The function-theoretic expressions associated with Equations (10.4) to (10.6) are as follows:

$$\sum_{i=1}^{c} \chi_{\mathrm{A}_i}(\mathbf{x}_k) = 1, \quad \text{for all } k, \tag{10.8}$$

$$\chi_{\mathrm{A}_i}(x_k) \wedge \chi_{\mathrm{A}_j}(\mathbf{x}_k) = 0, \quad \text{for all } k, \tag{10.9}$$

$$0 < \sum_{k=1}^{n} \chi_{\mathrm{A}_i}(\mathbf{x}_k) < n, \quad \text{for all } i, \tag{10.10}$$

where the characteristic function $\chi_{\mathrm{A}_i}(\mathbf{x}_k)$ is defined once again as

$$\chi_{\mathrm{A}_i}(x_k) = \begin{cases} 1, & \mathbf{x}_k \in \mathrm{A}_i; \\ 0, & \mathbf{x}_k \notin \mathrm{A}_i. \end{cases} \tag{10.11}$$

Equations (10.8) and (10.9) explain that any sample $\mathbf{x}_k$ can only and definitely belong to one of the c classes. Equation (10.10) implies that no class is empty and no class is a whole set X (i.e., the universe).

For simplicity in notation, our membership assignment of the jth data point in the ith cluster, or class, is defined to be $\chi_{ij}\chi_{\mathrm{A}_i}(\mathbf{x}_j)$. Now define a matrix U comprising elements χ_{ij} ($i = 1, 2, \ldots, c$; $j = 1, 2, \ldots, n$); hence, U is a matrix with c rows and n columns. Then, we define a hard c-partition space for X as the following matrix set:

$$\mathrm{M}_c = \left\{ \mathrm{U} \middle| \chi_{ij} \in \{0,\ 1\},\ \sum_{i=1}^{c} \chi_{ik} = 1,\ 0 < \sum_{k=1}^{n} \chi_{ik} < n \right\}. \tag{10.12}$$

Any matrix $U \in M_c$ is a hard c-partition. The cardinality of any hard c-partition, M_c, is

$$\eta_{M_c} = \left(\frac{1}{c!}\right)\left[\sum_{i=1}^{c}\binom{c}{i}(-1)^{c-i}\cdot i^n\right], \tag{10.13}$$

where the expression $\binom{c}{i}$ is the binomial coefficient of c things taken i at a time.

Example 10.6

Suppose we have five data points in a universe, $X = \{x_1, x_2, x_3, x_4, x_5\}$. Also, suppose we want to cluster these five points into two classes. For this case we have $n = 5$ and $c = 2$. The cardinality, using Equation (10.13), of this hard c-partition is given as

$$\eta_{M_c} = \frac{1}{2}\left[2(-1) + 2^5\right] = 15.$$

Some of the 15 possible hard two partitions are listed here:

$$\begin{bmatrix}1&1&1&1&0\\0&0&0&0&1\end{bmatrix}\begin{bmatrix}1&1&1&0&0\\0&0&0&1&1\end{bmatrix}\begin{bmatrix}1&1&0&0&0\\0&0&1&1&1\end{bmatrix}\begin{bmatrix}1&0&0&0&0\\0&1&1&1&1\end{bmatrix},$$

$$\begin{bmatrix}1&0&1&0&0\\0&1&0&1&1\end{bmatrix}\begin{bmatrix}1&0&0&1&0\\0&1&1&0&1\end{bmatrix}\qquad\begin{bmatrix}1&0&0&0&1\\0&1&1&1&0\end{bmatrix}$$

and so on.

Notice that the two matrices

$$\begin{bmatrix}1&1&1&1&0\\0&0&0&0&1\end{bmatrix} \text{ and } \begin{bmatrix}0&0&0&0&1\\1&1&1&1&0\end{bmatrix}$$

are not different-clustering two partitions. In fact, they are the same two partitions irrespective of an arbitrary row-swap. If we label the first row of the first U matrix class c_1 and we label the second row class c_2, we would get the same classification for the second U matrix by simply relabeling each row: the first row is c_2 and the second row is c_1. The cardinality measure given in Equation (10.13) gives the number of *unique* c-partitions for n data points.

An interesting question now arises: Of all the possible c-partitions for n data samples, how can we select the most reasonable c-partition for the partition space M_c? For instance, in the example just provided, which of the 15 possible hard two partitions for five data points and two classes is the best? The answer to this question is provided by the objective function (or classification criteria) to be used to classify or cluster the data. The one proposed for the HCM algorithm is known as a *within-class sum of squared errors approach* using a Euclidean norm to characterize distance. This algorithm is denoted $J(U, \mathbf{v})$, where U is the partition matrix, and the parameter $\mathbf{v}$ is a vector of cluster centers. This objective function is given as

$$J(U, \mathbf{v}) = \sum_{k=1}^{n}\sum_{i=1}^{c}\chi_{ik}(d_{ik})^2, \tag{10.14}$$

where d_{ik} is a Euclidean distance measure (in m-dimensional feature space, R^m) between the kth data sample $\mathbf{x}_k$ and ith cluster center $\mathbf{v}_i$, which is given as follows

$$d_{ik} = d(\mathbf{x}_k - \mathbf{v}_i) = ||\mathbf{x}_k - \mathbf{v}_i|| = \left[\sum_{j=1}^{m} \left(x_{kj} - v_{ij}\right)^2\right]^{1/2} \tag{10.15}$$

Because each data sample requires m coordinates to describe its location in R^m-space, each cluster center also requires m coordinates to describe its location in this same space. Therefore, the ith cluster center is a vector of length m,

$$\mathbf{v}_i = \{v_{i1}, v_{i2}, \ldots, v_{im}\},$$

where the jth coordinate is calculated by

$$v_{ij} = \frac{\sum_{k=1}^{n} \chi_{ik} \cdot x_{kj}}{\sum_{k=1}^{n} \chi_{ik}}. \tag{10.16}$$

We seek the optimum partition, U^*, to be the partition that produces the minimum value for the function, J. That is,

$$J(U^*, \mathbf{v}^*) = \min_{U \in M_c} J(U, \mathbf{v}). \tag{10.17}$$

Finding the optimum partition matrix, U^*, is exceedingly difficult for practical problems because $M_c \to \infty$ for even modest-sized problems. For example, for the case where $n = 25$ and $c = 10$, the cardinality approaches an extremely large number, that is, $M_c \to 10^{18}$, a very large number. Obviously, a search for optimality by exhaustion is *not* computationally feasible for problems of reasonable interest. Fortunately, useful and effective alternative search algorithms have been devised (Bezdek, 1981).

One such search algorithm is known as *iterative optimization*. Basically, this method is like many other iterative methods in that we start with an initial guess at the U matrix. From this assumed matrix (input values for the number of classes) and iteration tolerance (the accuracy we demand in the solution), we calculate the centers of the clusters (classes). From these cluster, or class, centers, we recalculate the membership values that each data point has in the cluster. We compare these values with the assumed values and continue this process until the changes from cycle to cycle are within our prescribed tolerance level.

The step-by-step procedures in this iterative optimization method are provided as follows (Bezdek, 1981):

1. Fix c $(2 \le c < n)$ and initialize the U matrix:

$$U^{(0)} \in M_c.$$

Then do $r = 0, 1, 2, \ldots$.

2. Calculate the c center vectors:

$$\left\{ \mathbf{v}_{\mathrm{i}}^{(r)} \text{ with } \mathrm{U}^{(r)} \right\}.$$

3. Update $\mathrm{U}^{(r)}$; calculate the updated characteristic functions (for all i, k):

$$\chi_{ik}^{(r+1)} = \begin{cases} 1, & d_{ik}^{(r)} = \min\left\{ d_{jk}^{(r)} \right\} \text{ for all } j \in c; \\ 0, & \text{otherwise.} \end{cases} \tag{10.18}$$

4. If

$$\left\| \mathrm{U}^{(r+1)} - \mathrm{U}^{(r)} \right\| \leq \varepsilon \text{ (tolerance level)} \tag{10.19}$$

stop; otherwise set $r = r + 1$ and return to step 2.

In step 4, the notation $\|$ is any matrix norm such as the Euclidean norm.

Example 10.7 (Bezdek, 1981).

A good illustration of the iterative optimization method is provided with the "butterfly problem" shown in Figure 10.3. In this problem, we have 15 data points and one of them is on a vertical line of symmetry (the point in the middle of the data cluster). Suppose we want to cluster our data into two classes. We can see that the points to the left of the line of symmetry should be in one class and the points to the right of the line of symmetry should be in the other class. The problem lies in assigning the point on the line of symmetry to a class. To which class should this point belong? Whichever class the algorithm assigns this point to, there will be a good argument that it should be a member of the other class. Alternatively, the argument may revolve

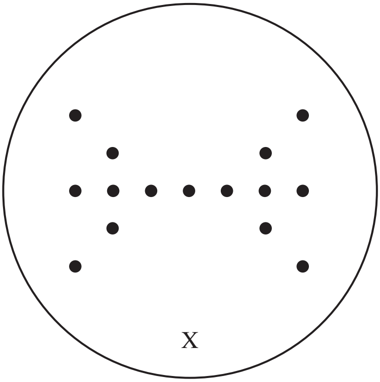

Figure 10.3 Butterfly classification problem. Adapted from Bezdek, 1981.

around the fact that the choice of two classes is a poor one for this problem. Three classes might be the best choice, but the physics underlying the data might be binary and two classes may be the only option.

In conducting the iterative optimization approach, we have to assume an initial U matrix. This matrix will have two rows (two classes, $c = 2$) and 15 columns (15 data points, $n = 15$). It is important to understand that the classes may be unlabeled in this process. That is, we can look at the structure of the data without the need for the assignment of labels to the classes. This is often the case when one is first looking at a group of data. After several iterations with the data, and as we become more and more knowledgeable about the data, we can then assign labels to the classes. We start the solution with the assumption that the point in the middle (i.e., the eighth column) is assigned to the class represented by the bottom row of the initial U matrix, $U^{(0)}$:

$$U^{(0)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

After four iterations (Bezdek, 1981), this method converges to within a tolerance level of $\varepsilon = 0.01$, as

$$U^{(4)} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

We note that the point on the line of symmetry (i.e., the eighth column) is still assigned to the class represented by the second row of the U matrix. The elements in the U matrix indicate membership of that data point in the first or second class. For example, the point on the line of symmetry has full membership in the second class and no membership in the first class, yet it is plain to see from Figure 10.3 that physically it should probably share membership with each class. This is not possible with crisp classification; membership is binary; that is, a point is either a member of a class or not.

The following example illustrates again the crisp classification method. The process will be instructive because of its similarity to the subsequent algorithm to be developed for the fuzzy classification method.

Example 10.8

In a chemical engineering process involving an automobile's catalytic converter (which converts carbon monoxide to carbon dioxide), we have a relationship between the conversion efficiency of the catalytic converter and the *inverse of the temperature* of the catalyst. Two classes of data are known from the reaction efficiency. Points of high conversion efficiency and high temperature are indicators of a nonpolluting system (class c_1) and points of low conversion efficiency and low temperature are indicative of a polluting system (class c_2). Suppose you measure the conversion efficiency and temperature (T) of four different catalytic converters and attempt to characterize them as polluting or nonpolluting. The four data points ($n = 4$) are shown in Figure 10.4, where the y axis is conversion efficiency and the x axis is the inverse of the temperature [in a conversion process like this the exact solution takes the form of $\ln(1/T)$].

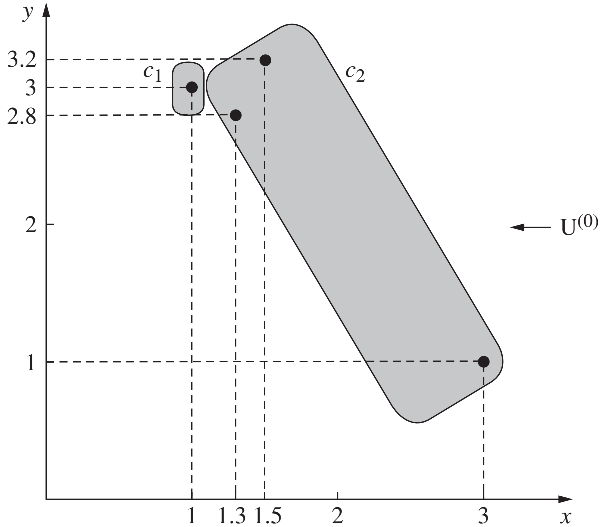

Figure 10.4 Four data points in two-dimensional feature space.

The data are described by two features ($m = 2$), and have the following coordinates in two-dimensional space:

$$\mathbf{x}_1 = \{1,\ 3\}$$
$$\mathbf{x}_2 = \{1.5,\ 3.2\}$$
$$\mathbf{x}_3 = \{1.3,\ 2.8\}$$
$$\mathbf{x}_4 = \{3,\ 1\}.$$

We desire to classify these data points into two classes ($c = 2$). It is sometimes useful to calculate the cardinality of the possible number of crisp partitions for this system, that is, to find η_{M_c} using Equation (10.13); thus,

$$\eta_{M_c} = \left(\frac{1}{c!}\right)\left[\sum \binom{c}{i}(-1)^{c-i} i^n\right] = \frac{1}{2!}\left[\binom{2}{1}(-1)^{(1)}(1)^4 + \binom{2}{2}(-1)^0(2)^4\right]$$
$$= \frac{1}{2}[-2 + 16] = 7,$$

which says that there are seven unique ways (irrespective of row-swaps) to classify the four points into two clusters. Let us begin the iterative optimization algorithm with an initial guess of the crisp partition, U, by assuming $\mathbf{x}_1$ to be in class 1 and $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ to be in class 2, as shown in Figure 10.4, that is,

$$U^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

Now, from the initial $U^{(0)}$ (which is one of the seven possible crisp partitions), we seek the optimum partition U^*, that is,

$$U^{(0)} \rightarrow U^{(1)} \rightarrow U^{(2)} \rightarrow \cdots \rightarrow U^*.$$

Of course, optimality is defined in terms of the desired tolerance or convergence level, ε. In general, for class 1 we calculate the coordinates of the cluster center,

$$\begin{aligned} v_{1j} &= \frac{\chi_{11}x_{1j} + \chi_{12}x_{2j} + \chi_{13}x_{3j} + \chi_{14}x_{4j}}{\chi_{11} + \chi_{12} + \chi_{13} + \chi_{14}} \\ &= \frac{(1)x_{1j} + (0)x_{2j} + (0)x_{3j} + (0)x_{4j}}{1+0+0+0}, \end{aligned}$$

and

$$\mathbf{v}_i = \{v_{i1},\ v_{i2}, \ldots, v_{im}\}.$$

In this case $m = 2$, which means we deal with two coordinates for each data point. Therefore,

$$\mathbf{v}_i = \{v_{i1},\ v_{i2}\}.$$

where

for $c = 1$ (which is class 1), $\mathbf{v}_1 = \{v_{11}, v_{12}\}$
for $c = 2$ (which is class 2), $\mathbf{v}_2 = \{v_{21}, v_{22}\}$.

Therefore, using the expression for v_{ij} for $c = 1$, and $j = 1$ and 2, respectively,

$$\left.\begin{aligned} v_{11} &= \frac{1(1)}{1} = 1 \rightarrow x \text{ coordinate} \\ v_{12} &= \frac{1(3)}{1} = 3 \rightarrow y \text{ coordinate} \end{aligned}\right\} \Rightarrow \mathbf{v}_1 = \{1,\ 3\},$$

which just happens to be the coordinates of point x_1 because this is the only point in the class for the assumed initial partition, $U^{(0)}$. For $c = 2$ or class 2, we get cluster center coordinates

$$v_{2j} = \frac{(0)x_{1j} + (1)x_{2j} + (1)x_{3j} + (1)x_{4j}}{0+1+1+1} = \frac{x_{2j} + x_{3j} + x_{4j}}{3}.$$

Hence, for $c = 2$ and $j = 1$ and 2, respectively,

$$\left.\begin{aligned} v_{21} &= \frac{1(1.5) + 1(1.3) + 1(3)}{3} = 1.93 \rightarrow x \text{ coordinate} \\ v_{22} &= \frac{1(3.2) + 1(2.8) + 1(1)}{3} = 2.33 \rightarrow y \text{ coordinate} \end{aligned}\right\} \Rightarrow \mathbf{v}_2 = \{1.93,\ 2.33\}.$$

Now, we compute the values for d_{ik} or the distances from the sample $\mathbf{x}_k$ (a data set) to the center $\mathbf{v}_i$ of the ith class. Using Equation (10.15),

$$d_{ik} = \left[\sum_{j=1}^{m} \left(x_{kj} - v_{ij} \right)^2 \right]^{1/2},$$

we get, for example, for $c = 1$, $d_{1k} = [(x_{k1} - v_{11})^2 + (x_{k2} - v_{12})^2]^{1/2}$. Therefore, for each data set $k = 1$ to 4, we compute the values of d_{ik} as follows: for cluster 1,

$$\begin{aligned}
d_{11} &= \sqrt{(1-1)^2 + (3-3)^2} = 0.0, \\
d_{12} &= \sqrt{(1.5-1)^2 + (3.2-3)^2} = 0.54, \\
d_{13} &= \sqrt{(1.3-1)^2 + (2.8-3)^2} = 0.36, \\
d_{14} &= \sqrt{(3-1)^2 + (1-3)^2} = 2.83,
\end{aligned}$$

and for cluster 2,

$$\begin{aligned}
d_{21} &= \sqrt{(1-1.93)^2 + (3-2.33)^2} = 1.14, \\
d_{22} &= \sqrt{(1.5-1.93)^2 + (3.2-2.33)^2} = 0.97, \\
d_{23} &= \sqrt{(1.3-1.93)^2 + (2.8-2.33)^2} = 0.78, \\
d_{24} &= \sqrt{(3-1.93)^2 + (1-2.33)^2} = 1.70,
\end{aligned}$$

Now, we update the partition to $\mathrm{U}^{(1)}$ for each data point (for $(c-1)$ clusters) using Equation (10.18). Hence, for class 1 we compare d_{ik} against the minimum of $\{d_{1k}, d_{2k}\}$:

For $k = 1$,

$$d_{11} = 0.0 \quad \min(d_{11}, d_{21}) = \min(0, 1.14) = 0.0; \text{ thus } \chi_{11} = 1.$$

For $k = 2$,

$$d_{12} = 0.54 \quad \min(d_{12}, d_{22}) = \min(0.54, 0.97) = 0.54; \text{ thus } \chi_{12} = 1.$$

For $k = 3$,

$$d_{13} = 0.36 \quad \min(d_{13}, d_{23}) = \min(0.36, 0.78) = 0.36; \text{ thus } \chi_{13} = 1.$$

For $k = 4$,

$$d_{14} = 2.83 \quad \min(d_{14}, d_{24}) = \min(2.83, 1.70) = 1.70; \text{ thus } \chi_{14} = 0.$$

Therefore, the updated partition is

$$\mathrm{U}^{(1)} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Because the partitions $\mathrm{U}^{(0)}$ and $\mathrm{U}^{(1)}$ are different, we repeat the same procedure based on the new setup of two classes. For $c=1$, the center coordinates are

$$\mathbf{v}_{1j} \text{ or } \mathbf{v}_j = \frac{x_{1j}+x_{2j}+x_{3j}}{1+1+1+0}, \quad \text{since } \chi_{14}=0,$$

$$\left.\begin{aligned} \mathbf{v}_{11} &= \frac{x_{11}+x_{21}+x_{31}}{3} = \frac{1+1.5+1.3}{3} = 1.26 \\ \mathbf{v}_{12} &= \frac{x_{12}+x_{22}+x_{32}}{3} = \frac{3+3.2+2.8}{3} = 3.0 \end{aligned}\right\} \mathbf{v}_1 = \{1.26,\ 3.0\},$$

and for $c=2$, the center coordinates are

$$\mathbf{v}_{2j} \text{ or } \mathbf{v}_j = \frac{x_{4j}}{0+0+0+1}, \quad \text{since } \chi_{21}=\chi_{22}=\chi_{23}=0,$$

$$\left.\begin{aligned} \mathbf{v}_{21} &= \frac{3}{1} = 3 \\ \mathbf{v}_{22} &= \frac{1}{1} = 1 \end{aligned}\right\} \mathbf{v}_2 = \{3,\ 1\}.$$

Now, we calculate the distance measures again:

$$\begin{aligned} d_{11} &= \sqrt{(1-1.26)^2+(3-3)^2} = 0.26 & d_{21} &= \sqrt{(1-3)^2+(3-1)^2} = 2.83 \\ d_{12} &= \sqrt{(1.5-1.26)^2+(3.2-3)^2} = 0.31 & d_{22} &= \sqrt{(1.5-3)^2+(3.2-1)^2} = 2.66 \\ d_{13} &= \sqrt{(1.3-1.26)^2+(2.8-3)^2} = 0.20 & d_{23} &= \sqrt{(1.3-3)^2+(2.8-1)^2} = 2.47 \\ d_{14} &= \sqrt{(3-1.26)^2+(1-3)^2} = 2.65 & d_{24} &= \sqrt{(3-3)^2+(1-1)^2} = 0.0 \end{aligned}$$

and again update the partition $\mathrm{U}^{(1)}$ to $\mathrm{U}^{(2)}$:

For $k=1$,

$$d_{11} = 0.26 \quad \min(d_{11},\ d_{21}) = \min(0.26,\ 2.83) = 0.26; \text{ thus } \chi_{11} = 1.$$

For $k=2$,

$$d_{12} = 0.31 \quad \min(d_{12},\ d_{22}) = \min(0.31,\ 2.66) = 0.31; \text{ thus } \chi_{12} = 1.$$

For $k=3$,

$$d_{13} = 0.20 \quad \min(d_{13},\ d_{23}) = \min(0.20,\ 2.47) = 0.20; \text{ thus } \chi_{13} = 1.$$

For $k=4$,

$$d_{14}=2.65 \quad \min(d_{14},\ d_{24})=\min(2.65,\ 0.0)=0.0; \quad \text{thus } \chi_{14}=0.$$

Because the partitions $U^{(1)}$ and $U^{(2)}$ are identical, we could say the iterative process has converged; therefore, the optimum hard partition (crisp) is

$$U^{(*)}=\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This optimum partition tells us that for this catalytic converter example, the data points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are more similar in the two-dimensional feature space, and different from data point $\mathbf{x}_4$. We could say that points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are more indicative of a nonpolluting converter than is data point $\mathbf{x}_4$.

Fuzzy *c*-Means (FCM)

Let us consider whether the butterfly example in Figure 10.3 could be improved with the use of fuzzy set methods. To develop these methods in classification, we define a family of fuzzy sets $\{\underset{\sim}{A}, i=1, 2, \ldots, c\}$, as a fuzzy c-partition on a universe of data points, X. Because fuzzy sets allow for degrees of membership, we can extend the crisp classification idea into a fuzzy classification notion. Then, we can assign membership to the various data points in each fuzzy set (fuzzy class, fuzzy cluster). Hence, a single point can have partial membership in more than one class. It will be useful to describe the membership value that the kth data point has in the ith class with the following notation:

$$\mu_{ik}=\mu_{\underset{\sim}{A}i}(x_k)\in[0,\ 1],$$

with the restriction (as with crisp classification) that the sum of all membership values for a single data point in all of the classes has to be unity:

$$\sum_{i=1}^{c}\mu_{ik}=1, \quad \text{for all } k=1,\ 2, \ldots,\ n. \tag{10.20}$$

As before in crisp classification, there can be no empty classes and there can be no class that contains all the data points. This qualification is manifested in the following expression:

$$0<\sum_{k=1}^{n}\mu_{ik}<n. \tag{10.21}$$

Because each data point can have partial membership in more than one class, the restriction of Equation (10.9) is not present in the fuzzy classification case, that is,

$$\mu_{ik}\wedge\mu_{jk}\neq 0. \tag{10.22}$$

The provisions of Equations (10.8) and (10.10) still hold for the fuzzy case, however,

$$\sum_{i=1}^{c} \mu_{\underset{\sim}{A}_i}(x_k) = 1, \quad \text{for all } k. \tag{10.23}$$

$$0 < \sum_{k=1}^{n} \mu_{\underset{\sim}{A}_i}(x_k) < n, \quad \text{for all } i. \tag{10.24}$$

Before, in the case of $c = 2$, the classification problem reduced to that of the excluded middle axioms for crisp classification. Because we now allow partial membership, the case of $c = 2$ does not follow the restrictions of the excluded middle axioms, that is, for two classes $\underset{\sim}{A}_i$ and $\underset{\sim}{A}_j$,

$$\underset{\sim}{A}_i \cap \underset{\sim}{A}_j \neq \emptyset. \tag{10.25}$$

$$\emptyset \subset \underset{\sim}{A}_j \subset X. \tag{10.26}$$

We can now define a family of fuzzy partition matrices, M_{fc}, for the classification involving c classes and n data points:

$$M_{fc} = \left\{ \underset{\sim}{U} | \mu_{ik} \in [0,\ 1];\ \sum_{i=1}^{c} \mu_{ik} = 1;\ 0 < \sum_{k=1}^{n} \mu_{ik} < n \right\}, \tag{10.27}$$

where $i = 1, 2, \ldots, c$ and $k = 1, 2, \ldots, n$.

Any $\underset{\sim}{U} \in M_{fc}$ is a fuzzy c-partition, and it follows from the overlapping character of the classes and the infinite number of membership values possible for describing class membership that the cardinality of M_{fc} is also infinity, that is, $\eta_{M_{fc}} = \infty$.

Example 10.9 (Similar to Bezdek, 1981).
Suppose you are a fruit geneticist interested in genetic relationships among fruits. In particular, you know that a tangelo is a cross between a grapefruit and a tangerine. You describe the fruit with features such as color, weight, sphericity, sugar content, skin, and texture. Hence, your feature space could be highly dimensional. Suppose you have three fruits (three data points):

$$X = \{x_1 = \text{grapefruit},\ x_2 = \text{tangelo},\ x_3 = \text{tangerine}\}.$$

These data points are described by m features, as discussed. You want to class the three fruits into two classes to determine the genetic assignment for the three fruits. In the crisp case, the classification matrix can take one of the three forms, that is, the cardinality for this case where $n = 3$ and $c = 2$ is $\eta_{M_c} = 3$ (Equation (10.13)). Suppose you arrange your $\underset{\sim}{U}$ matrix as follows:

$$\underset{\sim}{U} = \begin{array}{c} \\ C_1 \\ C_2 \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{array}.$$

The three possible partitions of the matrix are

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Notice that in the first partition, we are left with the uncomfortable segregation of the grapefruit in one class and the tangelo and the tangerine in the other; the tangelo shares nothing in common with the grapefruit! In the second partition, the grapefruit and the tangelo are in a class, suggesting that they share nothing in common with the tangerine! Finally, the third partition is the most genetically discomforting of all, because here the tangelo is in a class by itself, sharing nothing in common with its progenitors! One of these three partitions will be the final partition when any algorithm is used. The question is, which one is best? Intuitively, the answer is none, but in crisp classification, we have to use one of these.

In the fuzzy case, this segregation and genetic absurdity are not a problem. We can have the most intuitive situation where the tangelo shares membership with both classes with the parents. For example, the following partition might be a typical outcome for the fruit genetics problem:

$$\underset{\sim}{U} = \begin{array}{c} \\ 1 \\ 2 \end{array} \!\!\begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} 0.91 & 0.58 & 0.13 \\ 0.09 & 0.42 & 0.87 \end{bmatrix} \end{array}.$$

In this case, Equation (10.24) shows that the sum of each row is a number between 0 and n, or

$$0 < \sum_k \mu_{1k} = 1.62 < 3,$$
$$0 < \sum_k \mu_{2k} = 1.38 < 3,$$

and for Equation (10.22) there is overlap among the classes for each data point,

$$\mu_{11} \wedge \mu_{21} = \min(0.91,\ 0.09) = 0.09 \neq 0,$$
$$\mu_{12} \wedge \mu_{22} = \min(0.58,\ 0.42) = 0.42 \neq 0,$$
$$\mu_{13} \wedge \mu_{23} = \min(0.13,\ 0.87) = 0.13 \neq 0.$$

Fuzzy c*-Means Algorithm*

To describe a method to determine the fuzzy c-partition matrix $\underset{\sim}{U}$ for grouping a collection of n data sets into c classes, we define an objective function J_m for a fuzzy c-partition:

$$J_m(\underset{\sim}{U}, \mathrm{v}) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^{m'} (d_{ik})^2, \qquad (10.28)$$

where

$$d_{ik} = d(\mathbf{x}_k - \mathbf{v}_i) = \left[\sum_{j=1}^{m} \left(x_{kj} - v_{ij} \right)^2 \right]^{1/2} \tag{10.29}$$

and where μ_{ik} is the membership of the kth data point in the ith class.

As with crisp classification, the function J_m can have a large number of values, the smallest one associated with the *best* clustering. Because of the large number of possible values (now infinite because of the infinite cardinality of fuzzy sets) we seek to find the best possible, or optimum, solution without resorting to an exhaustive, or expensive, search. The distance measure, d_{ik} in Equation (10.29), is again a Euclidean distance between the ith cluster center and the kth data set (data point in m space). A new parameter is introduced in Equation (10.28) called a *weighting parameter*, m' (Bezdek, 1981). This value has a range $m' \in [1, \infty)$. This parameter controls the amount of fuzziness in the classification process and is discussed shortly. Also, as before, $\mathbf{v}_i$ is the ith cluster center, which is described by m features (m coordinates) and can be arranged in vector form as before, $\mathbf{v}_i = \{v_{i1}, v_{i2}, \ldots, v_{im}\}$.

Each of the cluster coordinates for each class can be calculated in a manner similar to the calculation in the crisp case (Equation (10.16)):

$$v_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} \cdot x_{ki}}{\sum_{k=1}^{n} \mu_{ik}^{m'}}, \tag{10.30}$$

where j is a variable on the feature space, that is, $j = 1, 2, \ldots, m$.

As in the crisp case, the optimum fuzzy c-partition will be the smallest of the partitions described in Equation (10.28), that is,

$$J_m^*(\underset{\sim}{\mathrm{U}}^*, \mathbf{v}^*) = \min_{\mathrm{M_{fc}}} J(\underset{\sim}{\mathrm{U}}, \mathbf{v}). \tag{10.31}$$

As with many optimization processes (Chapter 12), the solution to Equation (10.31) cannot be guaranteed to be a global optimum, that is, the best of the best. What we seek is the best solution available within a prespecified level of accuracy. An effective algorithm for fuzzy classification, called *iterative optimization*, was proposed by Bezdek (1981). The steps in this algorithm are as follows:

1. Fix c ($2 \le c < n$) and select a value for parameter m'. Initialize the partition matrix, $\underset{\sim}{\mathrm{U}}^{(0)}$. Each step in this algorithm will be labeled r, where $r = 0, 1, 2, \ldots$
2. Calculate the c centers $\left\{ \mathbf{v}_i^{(r)} \right\}$ for each step.
3. Update the partition matrix for the rth step, $\underset{\sim}{\mathrm{U}}^{(r)}$, as follows:

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{2/(m'-1)} \right]^{-1}, \quad \text{for } \mathrm{I}_k = \varnothing \tag{10.32a}$$

or

$$\mu_{ik}^{(r+1)} = 0, \quad \text{for all classes } i \text{ where } i \in \underset{\sim}{\mathrm{I}}_k, \tag{10.32b}$$

where

$$\mathrm{I}_k = \left\{ i | 2 \le c < n; d_{ik}^{(r)} = 0 \right\} \tag{10.33}$$

and

$$\underset{\sim}{\mathrm{I}}_k = \{1, 2, \ldots, c\} - \mathrm{I}_k, \tag{10.34}$$

and

$$\sum_{i \in \mathrm{I}_k} \mu_{ik}^{(r+1)} = 1. \tag{10.35}$$

4. If $\left\| \underset{\sim}{\mathrm{U}}^{(r+1)} - \underset{\sim}{\mathrm{U}}^{(r)} \right\| \le \varepsilon_L$, stop; otherwise set $r = r + 1$ and return to step 2.

In step 4, we compare a matrix norm ‖ of two successive fuzzy partitions to a prescribed level of accuracy, ε_L, to determine whether the solution is good enough. In step 3, there is a considerable amount of logic involved in Equations (10.32) to (10.35). Equation (10.32a) is straightforward enough, except when the variable d_{jk} is zero. Because this variable is in the denominator of a fraction, the operation is undefined mathematically, and computer calculations are abruptly halted. So the parameters I_k and $\underset{\sim}{\mathrm{I}}_k$ comprise a bookkeeping system to handle situations when some of the distance measures, d_{ij}, are zero, or extremely small in a computational sense. If a zero value is detected, Equation (10.32b) sets the membership for that partition value to be zero. Equations (10.33) and (10.34) describe the bookkeeping parameters I_k and $\underset{\sim}{\mathrm{I}}_k$, respectively, for each of the classes. Equation (10.35) simply says that all the nonzero partition elements in each column of the fuzzy classification partition, $\underset{\sim}{\mathrm{U}}$, sum to unity. The following example serves to illustrate Equations (10.32) to (10.35).

Example 10.10

Suppose we have calculated the following distance measures for one step in our iterative algorithm for a classification problem involving three classes and five data points. The values in Table 10.7 are simple numbers for ease of illustration. The bookkeeping parameters I_k and $\underset{\sim}{\mathrm{I}}_k$, where in this example $k = 1$, 2, 3, 4, 5, are given next, as illustration of the use of Equations (10.33) and (10.34):of the use of Equations (10.33) and (10.34):

Table 10.7 Distance measures for hypothetical example (c = 3, n = 5).

$d_{11} = 1$	$d_{21} = 2$	$d_{31} = 3$
$d_{12} = 0$	$d_{22} = 0.5$	$d_{32} = 1$
$d_{13} = 1$	$d_{23} = 0$	$d_{33} = 0$
$d_{14} = 3$	$d_{24} = 1$	$d_{34} = 1$
$d_{15} = 0$	$d_{25} = 4$	$d_{35} = 0$

$$\mathrm{I}_1 = \emptyset \quad \underset{\sim}{\mathrm{I}}_1 = \{1,2,3\} - \emptyset = \{1,2,3\},$$
$$\mathrm{I}_2 = \{1\} \quad \underset{\sim}{\mathrm{I}}_2 = \{1,2,3\} - \{1\} = \{2,3\},$$
$$\mathrm{I}_3 = \{2,3\} \quad \underset{\sim}{\mathrm{I}}_3 = \{1,2,3\} - \{2,3\} = \{1\},$$
$$\mathrm{I}_4 = \emptyset \quad \underset{\sim}{\mathrm{I}}_4 = \{1,2,3\} - \emptyset = \{1,2,3\},$$
$$\mathrm{I}_5 = \{1,3\} \quad \underset{\sim}{\mathrm{I}}_5 = \{1,2,3\} - \{1,3\} = \{2\}.$$

Now, Equations (10.32) and (10.35) are illustrated:

$$\text{For data point 1: } \mu_{11}, \mu_{21}, \mu_{31} \neq 0 \quad \text{and} \quad \mu_{11} + \mu_{21} + \mu_{31} = 1$$
$$\text{For data point 2: } \mu_{12} = 0, \quad \mu_{22}, \mu_{32} \neq 0 \quad \text{and} \quad \mu_{22} + \mu_{23} = 1$$
$$\text{For data point 3: } \mu_{13} = 1 \quad \text{and} \quad \mu_{23} = \mu_{33} = 0$$
$$\text{For data point 4: } \mu_{14}, \mu_{24}, \mu_{34} \neq 0 \quad \text{and} \quad \mu_{14} + \mu_{24} + \mu_{34} = 1$$
$$\text{For data point 5: } \mu_{25} = 1 \quad \text{and} \quad \mu_{15} = \mu_{35} = 0.$$

The algorithm given in Equation (10.28) is a *least squares* function, where the parameter n is the number of data sets and c is the number of classes (partitions) into which one is trying to classify the data sets. The squared distance, d_{ik}^2, is then weighted by a measure, $(u_{ik})^{m'}$, of the membership of x_k in the ith cluster. The value of J_m is then a measure of the sum of all the *weighted* squared errors; this value is then minimized with respect to two constraint functions. First, J_m is minimized with respect to the squared errors within each cluster, that is, for each specific value of c. Simultaneously, the distance between cluster centers is maximized, that is, max $|\mathbf{v}_i - \mathbf{v}_j|$, $i \neq j$.

As indicated, the range for the membership exponent is $m' \in [1, \infty)$. For the case $m' = 1$, the distance norm is Euclidean and the FCM algorithm approaches a HCM algorithm, that is, only zeros and ones come out of the clustering. Conversely, as $m' \to \infty$, the value of the function $J_m \to 0$. This result seems intuitive because the membership values are numbers less than or equal to 1, and large powers of fractions less than 1 approach 0. In general, the larger m' is, the fuzzier the membership assignments of the clustering are; conversely, as $m' \to 1$, the clustering values become hard, that is, 0 or 1. The exponent m' thus controls the extent of membership sharing between fuzzy clusters. If all other algorithmic parameters are fixed, then increasing m' will result in decreasing J_m. No theoretical optimum choice of m' has emerged in the literature. However, the bulk of the literature seems to report values in the range 1.25–2. Convergence of the algorithm tends to be slower as the value of m' increases.

The algorithm described here can be remarkably accurate and robust in the sense that poor guesses for the initial partition matrix, $\underset{\sim}{\mathrm{U}}^{(0)}$, can be overcome quickly, as illustrated in the next example.

Example 10.11

Continuing with the chemical engineering example on a catalytic converter shown in Figure 10.4, we can see that a visual display of these points in two-dimensional feature space ($m = 2$) makes it easy for humans to cluster the data into two convenient classes based on the proximity of the points to one another. The fuzzy classification method generally converges quite rapidly, even when the initial guess for the fuzzy partition is quite poor, in a classification sense. The fuzzy iterative optimization method for this case would proceed as follows.

Using U* from the previous example as the initial fuzzy partition, $\underset{\sim}{\mathrm{U}}^{(0)}$, and assuming a weighting factor of $m' = 2$ and a criterion for convergence of $\varepsilon_L = 0.01$, that is,

$$\max_{i,k} |\mu_{ik}^{(r+1)} - \mu_{ik}^{(r)}| \le 0.01,$$

we want to determine the optimum fuzzy two-partition $\underset{\sim}{\mathrm{U}}^*$. To begin, the initial fuzzy partition is

$$\underset{\sim}{\mathrm{U}}^{(0)} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Next is the calculation of the initial cluster centers using Equation (10.30), where $m' = 2$:

$$v_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^2 \cdot x_{kj}}{\sum_{k=1}^{n} (\mu_{ik})^2},$$

where for $c = 1$,

$$\begin{aligned} v_{1j} &= \frac{\mu_1^2 x_{1j} + \mu_2^2 x_{2j} + \mu_3^2 x_{3j} + \mu_4^2 x_{4j}}{\mu_1^2 + \mu_2^2 + \mu_3^2 + \mu_4^2} \\ &= \frac{(1)x_{1j} + (1)x_{2j} + (1)x_{3j} + (0)x_{4j}}{1+1+1+0} = \frac{x_{1j} + x_{2j} + x_{3j}}{1^2 + 1^2 + 1^2 + 0}, \end{aligned}$$

$$\left.\begin{aligned} v_{11} &= \frac{1 + 1.5 + 1.3}{3} = 1.26 \\ v_{12} &= \frac{3 + 3.2 + 2.8}{3} = 3.0 \end{aligned}\right\} \mathbf{v}_1 = \{1.26, 3.0\},$$

and for $c = 2$,

$$\mathbf{v}_{2j} \text{ or } \mathbf{v}_j = \frac{x_{4j}}{0+0+0+1}, \quad \text{since } x_{21} = x_{22} = x_{23} = 0,$$

$$\left.\begin{aligned} \mathbf{v}_{21} &= \frac{3}{1} = 3 \\ \mathbf{v}_{22} &= \frac{1}{1} = 1 \end{aligned}\right\} \mathbf{v}_2 = \{3,1\}.$$

Now the distance measures (distances of each data point from each cluster center) are found using Equation (10.29):

$$\begin{aligned} d_{11} &= \sqrt{(1-1.26)^2 + (3-3)^2} = 0.26 & d_{21} &= \sqrt{(1-3)^2 + (3-1)^2} = 2.82 \\ d_{12} &= \sqrt{(1.5-1.26)^2 + (3.2-3)^2} = 0.31 & d_{22} &= \sqrt{(1.5-3)^2 + (3.2-1)^2} = 2.66 \\ d_{13} &= \sqrt{(1.3-1.26)^2 + (2.8-3)^2} = 0.20 & d_{23} &= \sqrt{(1.3-3)^2 + (2.8-1)^2} = 2.47 \\ d_{14} &= \sqrt{(3-1.26)^2 + (1-3)^2} = 2.65 & d_{24} &= \sqrt{(3-3)^2 + (1-1)^2} = 0.0 \end{aligned}$$

With the distance measures, we can now update $\underset{\sim}{U}$ using Equations (10.33) to (10.35) (for $m' = 2$), that is,

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c}\left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}}\right)^2\right]^{-1},$$

and we get

$$\mu_{11} = \left[\sum_{j=1}^{c}\left(\frac{d_{11}}{d_{j1}}\right)^2\right]^{-1} = \left[\left(\frac{d_{11}}{d_{11}}\right)^2 + \left(\frac{d_{11}}{d_{21}}\right)^2\right]^{-1} = \left[\left(\frac{0.26}{0.26}\right)^2 + \left(\frac{0.26}{2.82}\right)^2\right]^{-1} = 0.991,$$

$$\mu_{12} = \left[\left(\frac{d_{12}}{d_{12}}\right)^2 + \left(\frac{d_{12}}{d_{22}}\right)^2\right]^{-1} = \left[1 + \left(\frac{0.31}{2.66}\right)^2\right]^{-1} = 0.986,$$

$$\mu_{13} = \left[\left(\frac{d_{13}}{d_{13}}\right)^2 + \left(\frac{d_{13}}{d_{23}}\right)^2\right]^{-1} = \left[1 + \left(\frac{0.20}{2.47}\right)^2\right]^{-1} = 0.993,$$

$$\mu_{14} = \left[\left(\frac{d_{14}}{d_{14}}\right)^2 + \left(\frac{d_{14}}{d_{24}}\right)^2\right]^{-1} = \left[1 + \left(\frac{2.65}{0}\right)^2\right]^{-1} \rightarrow 0.0, \text{ for } I_4 \neq \emptyset.$$

Using Equation (10.20) for the other partition values, μ_{2j}, for $j = 1, 2, 3, 4$, the new membership functions form an updated fuzzy partition, which is given as

$$\underset{\sim}{U}^{(1)} = \begin{bmatrix} 0.991 & 0.986 & 0.993 & 0 \\ 0.009 & 0.014 & 0.007 & 1 \end{bmatrix}.$$

To determine whether we have achieved convergence, we choose a matrix norm ‖ such as the maximum absolute value of pairwise comparisons of each of the values in $\underset{\sim}{U}^{(0)}$ and $\underset{\sim}{U}^{(1)}$, for example,

$$\max_{i,k}\left|\mu_{ik}^{(1)} - \mu_{ik}^{(0)}\right| = 0.0134 > 0.01.$$

This result suggests that our convergence criteria have not yet been satisfied, so we need another iteration of the method.

For the next iteration, we proceed by again calculating cluster centers, but now using values from the latest fuzzy partition, $\underset{\sim}{U}^{(1)}$; for $c = 1$,

$$v_{1j} = \frac{(0.991)^2 x_{1j} + (0.986)^2 x_{2j} + (0.993)^2 x_{3j} + (0) x_{4j}}{0.991^2 + 0.986^2 + 0.993^2 + 0},$$

$$\left.\begin{aligned} v_{11} &= \frac{0.98(1) + 0.97(1.5) + 0.99(1.3)}{2.94} = \frac{3.719}{2.94} \approx 1.26 \\ v_{12} &= \frac{0.98(3) + 0.97(3.2) + 0.99(2.8)}{2.94} = \frac{8.816}{2.94} \approx 3.0 \end{aligned}\right\} \mathbf{v}_1 = \{1.26, 3.0\},$$

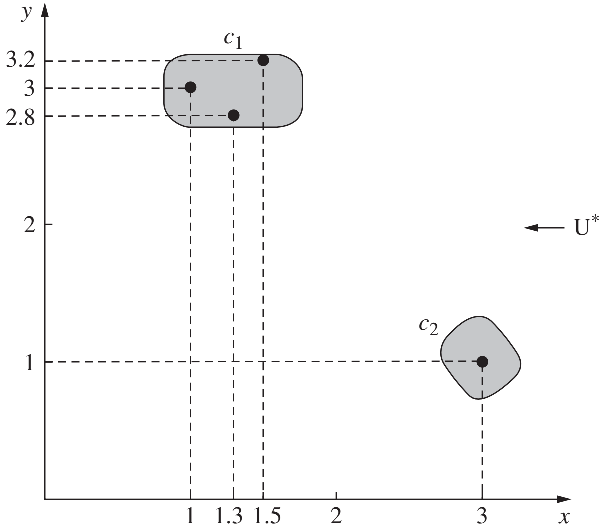

Figure 10.5 Converged fuzzy partition for catalytic converter example.

and for $c = 2$,

$$v_{2j} = \frac{(0.009)^2 x_{1j} + (0.014)^2 x_{2j} + (0.007)^2 x_{3j} + (1) x_{4j}}{0.009^2 + 0.014^2 + 0.007^2 + 1^2},$$

$$\left.\begin{aligned} v_{21} &= \frac{0.009^2(1) + 0.014^2(1.5) + 0.007^2(1.3) + 1(3)}{1.000} \approx 3.0 \\ v_{22} &= \frac{0.009^2(3) + 0.014^2(3.2) + 0.007^2(2.8) + 1(1)}{1.000} \approx 1.0 \end{aligned}\right\} \mathbf{v}_2 = \{3.0, 1.0\}.$$

We see that these two cluster centers are identical to those from the first step, at least to within the stated accuracy of (0.01); hence, the final partition matrix will be unchanged, to an accuracy of two digits, from that obtained in the previous iteration. As suggested previously, convergence is rapid, at least for this example. The final partition, $\underset{\sim}{\mathrm{U}}^{(2)}$, results in a classification shown in Figure 10.5.

Classification Metric

In most studies involving fuzzy pattern classification, the data used in the classification process typically come from electrically active transducer readings (Bezdek et al., 1986). When a fuzzy clustering is accomplished, a question remains concerning the uncertainty of the clustering in terms of the features used. That is, our interest should lie with the extent to which pairs of fuzzy classes of $\underset{\sim}{U}$ overlap; a true classification with no uncertainty would contain classes with no overlap. The question then is as follows: How fuzzy is a fuzzy c-partition? Suppose we compare two memberships for a given data set, x_k, pairwise, using the minimum function, that is,

$$\min\{u_i(x_k), u_j(x_k)\} > 0. \tag{10.36}$$

This comparison would indicate that membership of x_k is *shared* by u_i and u_j, whereas the minimum of these two values reflects that the minimum amount of *unshared* membership x_k can claim in either u_i or u_j. Hence, a fuzziness measure based on functions min $\{u_i(x_k),\ u_j(x_k)\}$ would constitute a point-by-point assessment—not of overlap, but of "anti-overlap" (Bezdek, 1974). A more useful measure of fuzziness in this context, which has values directly dependent on the relative overlap between nonempty fuzzy class intersections, can be found with the algebraic product of u_i and u_j, or the form $u_i\ u_j(x_k)$. An interesting interpretation of the fuzzy clustering results is to compute the fuzzy partition coefficient,

$$F_c(\underset{\sim}{U}) = \frac{\mathrm{tr}(\underset{\sim}{U}^* \underset{\sim}{U}^{\mathrm{T}})}{n}, \tag{10.37}$$

where $\underset{\sim}{U}$ is the fuzzy partition matrix being segregated into c classes (partitions), n is the number of data sets, and the operation * is standard matrix multiplication. The product $\underset{\sim}{U}^* \underset{\sim}{U}^{\mathrm{T}}$ is a matrix of size $c \times c$. The partition coefficient, $F_c(\underset{\sim}{U})$, has some special properties (Bezdek, 1974): $F_c(\underset{\sim}{U}) = 1$ if the partitioning in $\underset{\sim}{U}$ is crisp (comprising zeros and ones); $F_c(\underset{\sim}{U}) \le 1/c$ if all the values $u_i = 1/c$ (complete ambiguity); and in general $1/c \le F_c(\underset{\sim}{U}) \le 1$. The diagonal entries of $\underset{\sim}{U}^* \underset{\sim}{U}^{\mathrm{T}}$ are proportional to the amount of *unshared* membership of the data sets in the fuzzy clusters, whereas the off-diagonal elements of $\underset{\sim}{U}^* \underset{\sim}{U}^{\mathrm{T}}$ represent the amount of membership *shared* between pairs of fuzzy clusters of $\underset{\sim}{U}$. If the off-diagonal elements of $\underset{\sim}{U}^* \underset{\sim}{U}^{\mathrm{T}}$ are zero, then the partitioning (clustering) is crisp. As the partition coefficient approaches a value of unity, the fuzziness in overlap in classes is minimized. Hence, as $F_c(\underset{\sim}{U})$ increases, the decomposition of the data sets into the classes chosen is more successful.

Example 10.12 (Ross, Hasselman, and Chrostowski, 1993).
Forced response dynamics of a simple mechanical two-degrees-of-freedom (2-DOF) oscillating system, as shown in Figure 10.6, are conducted. In the figure, the parameters m_1, k_1, and c_1, and m_2, k_2, and c_2 are the mass, stiffness, and damping coefficients for the two masses, respectively, and the base of the system is assumed fixed to ground. The associated frequency and damping ratios for the two modes of free vibration for this system are summarized in Table 10.8. The 2-DOF system is excited with a force actuator on the larger of the two masses (Figure 10.6), and a displacement sensor on this same mass collects data on displacement versus time.

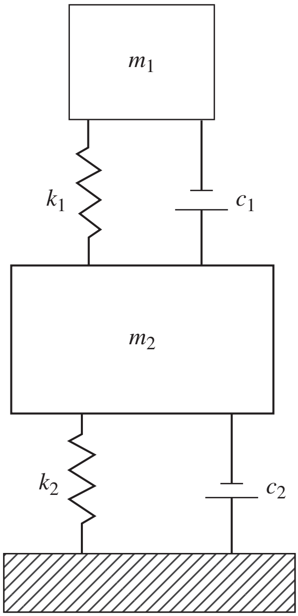

Figure 10.6 Mechanical system with two degrees of freedom.

Table 10.8 Modal response parameters for 2-DOF example.

Mode	Frequency (rads^{-1})	Damping ratio
1	0.98617	0.01
2	1.17214	0.01

Table 10.9 FRF sensitivity-parameter uncertainty data sets.

Dataset	FRF derivative	Variance
x_1	3.5951	0.2370
x_2	1.7842	0.2906
x_3	0.1018	0.3187
x_4	3.3964	0.2763
x_5	1.7620	0.2985
x_6	0.1021	0.4142

The response is computed for the displacement sensor in the form of frequency response functions (FRF). The derivatives of these FRF were computed for a specific exciting frequency of 1.0 rad s^{-1} with respect to the six modal mass and stiffness matrix elements (denoted in Table 10.9 as $x_1, x_2, \ldots, x_6$). Amplitude derivatives and the covariance matrix entries of these parameters are given in Table 10.9.

Table 10.10 Clustering results for a simple 2-DOF problem.

	Data pairs ($c = 2$)					
	x_1	x_2	x_3	x_4	x_5	x_6
Class 1	0.000	0.973	0.998	0.000	0.976	0.998
Class 2	1.000	0.027	0.002	1.000	0.024	0.002
	Data pairs ($c=3$)					
Class 1	0	0	1	0	0	1
Class 2	0	1	0	0	1	0
Class 3	1	0	0	1	0	0

Table 10.11 Partitioning coefficient for two different classes.

c	$F_c(\underset{\sim}{U})$
2	0.982
3	1.000

A simple FCM classification approach was conducted on the feature data ($m=2$) given in Table 10.9 and the fuzzy partitions shown in Table 10.10 resulted for a two-class case ($c=2$) and a three-class case ($c=3$).

The resulting values of $F_c(\underset{\sim}{U})$ from Equation (10.37) for the two clustering cases are listed in Table 10.11. Of course, the result for $c=3$ is intuitively obvious from inspection of Table 10.10, which is crisp. However, such obviousness is quickly lost when one deals with problems characterized by a large database.

Hardening the Fuzzy *c*-Partition

There are two popular methods, among many others, to defuzzify fuzzy partitions, $\underset{\sim}{U}$, that is, for hardening the fuzzy classification matrix. This defuzzification may be required in the ultimate assignment of data to a particular class. These two methods are called the *maximum membership method* and the *nearest center classifier*.

In the max membership method, the largest element in each column of the $\underset{\sim}{U}$ matrix is assigned a membership of unity and all other elements in each column are assigned a membership value of zero. In mathematical terms, if the largest membership in the kth column is μ_{ik}, then x_k belongs to class i, that is, if

$$\mu_{ik} = \max_{j\in c}\{\mu_{ik}\}, \text{then } \mu_{ik} = 1;\;\; \mu_{jk} = 0, \text{for all } j \neq i, \tag{10.38}$$

for $i=2, \ldots, c$ and $k=1, 2, \ldots, n$.

In the nearest center classifier, each of the data points is assigned to the class that it is closest to; that is, the minimum Euclidean distance from a given data point and the c cluster centers dictate the class assignment of that point. In mathematical terms, if

$$d_{ik} = \min_{j \in c} \{d_{jk}\} = \min_{j \in c} \left\| \mathbf{x}_k - \mathbf{v}_j \right\|,$$

then

$$\begin{aligned} \mu_{ik} &= 1; \\ \mu_{jk} &= 0, \quad \text{for all } j \neq i. \end{aligned} \tag{10.39}$$

Example 10.13

If we take the partition matrix, $\underset{\sim}{U}$, developed on the catalytic converter in Example 10.8 as shown in Figure 10.5, and harden it using the methods in Equations (10.38) and (10.39), we get the following:

$$\underset{\sim}{U} = \begin{bmatrix} 0.991 & 0.986 & 0.993 & 0 \\ 0.009 & 0.014 & 0.007 & 1 \end{bmatrix}.$$

Max membership method:

$$U^{\text{Hard}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Nearest center classifier: If we take the distance measures from the catalytic converter problem, that is,

$$\begin{aligned} d_{11} &= 0.26 \; d_{21} = 2.82 \\ d_{12} &= 0.31 \; d_{22} = 2.66 \\ d_{13} &= 0.20 \; d_{23} = 2.47 \\ d_{14} &= 2.65 \; d_{24} = 0 \end{aligned}$$

and arrange these values in a 2×4 matrix, such as

$$d_{ij} = \begin{bmatrix} 0.26 & 0.31 & 0.20 & 2.65 \\ 2.82 & 2.66 & 2.47 & 0 \end{bmatrix},$$

then the minimum value (distance) in each column is set to unity, and all other values (distances) in that column are set to zero. This process results in the following hard c-partition:

$$U^{\text{Hard}} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

which, for this example, happens to be the same partition that is derived using the max membership hardening method.

Similarity Relations from Clustering

The classification idea can be recast in the form of a similarity relation that is also a tolerance relation. This idea represents another way to look at the structure in data, by comparing the data points to one another, pairwise, in a similarity analysis. In classification, we seek to segregate data into clusters where points in each cluster are as "similar" to one another as possible and where clusters are dissimilar to one another. This notion of similarity, then, is central to classification. The use of a fuzzy similarity relation can be useful in the classification process (Bezdek and Harris, 1978).

A fuzzy relation $\underset{\sim}{R}$ can be constructed from the fuzzy partition $\underset{\sim}{U}$ as follows:

$$\underset{\sim}{R} = \left(\underset{\sim}{U}^{T}\left(\sum\wedge\right)\underset{\sim}{U}\right) = \left[r_{kj}\right], \tag{10.40}$$

$$r_{kj} = \sum_{i=1}^{c} \mu_{ik} \wedge \mu_{ij} \tag{10.41}$$

where the symbol $(\sum\wedge)$ denotes "sum of mins."

Example 10.14

We take the fuzzy partition $\underset{\sim}{U}$ from the fruit genetics example (Example 10.9) and perform the mixed algebraic and set operations as provided in Equations (10.40) and (10.41). So, for

$$\underset{\sim}{U}^{T} = \begin{bmatrix} 0.91 & 0.09 \\ 0.58 & 0.42 \\ 0.13 & 0.87 \end{bmatrix} \quad \text{and} \quad \underset{\sim}{U} = \begin{bmatrix} 0.91 & 0.58 & 0.13 \\ 0.09 & 0.42 & 0.87 \end{bmatrix},$$

we get

$$\begin{aligned} r_{11} &= \min(0.91,0.91) + \min(0.09,0.09) = 1, \\ r_{12} &= \min(0.91,0.58) + \min(0.09,0.42) = 0.67, \\ r_{13} &= \min(0.91,0.13) + \min(0.09,0.87) = 0.22, \\ r_{23} &= \min(0.58,0.13) + \min(0.42,0.87) = 0.55, \end{aligned}$$

and so forth, and the following fuzzy similarity relation results:

$$\underset{\sim}{R} = \begin{bmatrix} 1 & 0.67 & 0.22 \\ 0.67 & 1 & 0.55 \\ 0.22 & 0.55 & 1 \end{bmatrix}.$$

The fuzzy similarity relation $\underset{\sim}{R}$ provides similar information about clustering as does the original fuzzy partition, $\underset{\sim}{U}$. The fuzzy classification partition groups the data according to class type; the fuzzy relation shows the pairwise similarity of the data without regard to class type. Data that have strong similarity, or high membership values in $\underset{\sim}{R}$, should tend to have high membership in the same class in $\underset{\sim}{U}$. Although the two measures are based on the same data (i.e., the features describing each data point), their information content is slightly different.

Fuzzy Pattern Recognition

Pattern recognition can be defined as a process of identifying structure in data by comparisons to known structure; the known structure is developed through methods of classification (Bezdek, 1981) as illustrated in this chapter. In the statistical approach to numerical pattern recognition, which is treated thoroughly by Fukunaga (1972), each input observation is represented as a multidimensional data vector (feature vector) where each component is called a *feature*. The purpose of the pattern recognition system is to assign each input to one of c possible pattern classes (or data clusters). Presumably, different input observations should be assigned to the same class if they have similar features and to different classes if they have dissimilar features. Statistical pattern recognition systems rest on mathematical models; it is crucial that the measure of mathematical similarity used to match feature vectors with classes assesses a property shared by physically similar components of the process generating the data.

The data used to design a pattern recognition system are usually divided into two categories: design (or training) data and test data, much like the categorization used in neural networks. Design data are used to establish the algorithmic parameters of the pattern recognition system. The design samples may be labeled (the class to which each observation belongs is known) or unlabeled (the class to which each data sample belongs is unknown). Test data are labeled samples used to test the overall performance of the pattern recognition system.

In the descriptions that follow, the following notation is used:

$\mathrm{X} = \{x_1, x_2, \ldots, x_n\}$ the universe if data samples
n number of data samples in universe
p number of original (nominated) features
$x_k \in \mathrm{R}^p$ kth data sample in X, in p-dimensional space
$x_{kj} \in \mathrm{R}$ jth measured feature of x_k
s number of selected or extracted features
c number of clusters of classes.

There are many similarities between classification and pattern recognition. The information provided in Figure 10.7 summarizes the distinction between the two made in this textbook. Basically, classification establishes (or seeks to determine) the structure in data, whereas pattern recognition attempts to take new data and assign them to one of the classes defined in the classification process. Simply stated, classification *defines* the patterns and pattern recognition *assigns* data to a class; hence, the processes of define and assign are a coupled pair in the process described in Figure 10.7. In both the classification process and the pattern recognition process, there are necessary feedback loops: the first loop in classification is required when one is seeking a better segmentation of the data (i.e., better class distinctions) and the second loop is required when pattern matching fails (i.e., no useful assignment can be made).

Single-Sample Identification

A typical problem in pattern recognition is to collect data from a physical process and classify them into known patterns. The known patterns are represented as typical class structures, where each class structure is described by a number of features. For simplicity in presentation, the

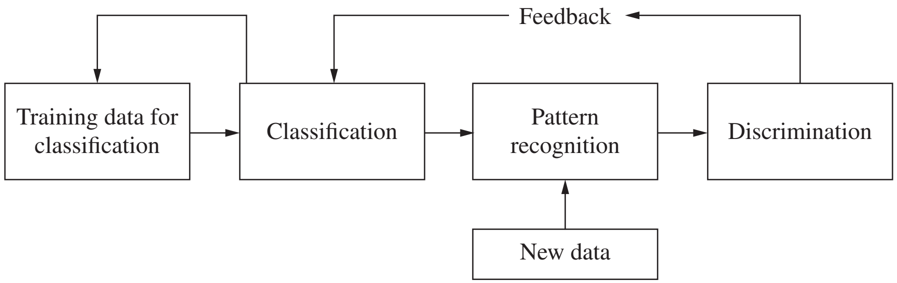

Figure 10.7 Difference between classification and pattern recognition.

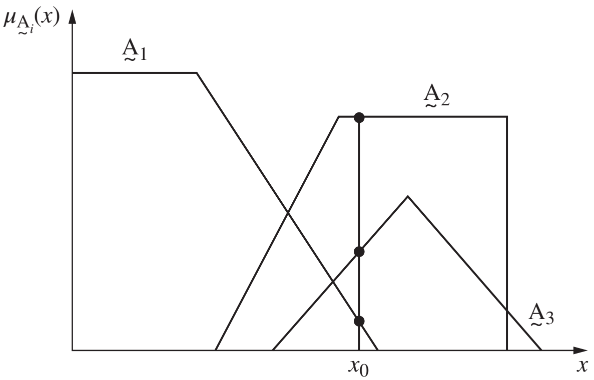

Figure 10.8 Single data sample using max membership criteria.

material that follows represents classes or patterns characterized by one feature; hence, the representation can be considered one dimensional.

Suppose we have several typical patterns stored in our knowledge base (i.e., the computer), and we are given a new data sample that has not yet been classified. We want to determine which pattern the sample most closely resembles. Express the typical patterns as fuzzy sets $\underset{\sim}{A}_1, \underset{\sim}{A}_2, \ldots, \underset{\sim}{A}_m$. Now, suppose we are given a new data sample that is characterized by the crisp singleton, x_0. Using the simple criterion of maximum membership, the typical pattern that the data sample most closely resembles is found by the following expression:

$$\mu_{\underset{\sim}{A}}i(x_0) = \max\{\mu_{\underset{\sim}{A}_1}(x_0),\ \mu_{\underset{\sim}{A}_2}(x_0),\ \ldots, \mu_{\underset{\sim}{A}}m(x_0)\}, \tag{10.42}$$

where x_0 belongs to the fuzzy set $\underset{\sim}{A}_i$ which is the set indication for the set with the highest membership at point x_0. Figure 10.8 shows the idea expressed in Equation (10.42), where clearly the new data sample defined by the singleton expressed by x_0 most closely resembles the pattern described by fuzzy set $\underset{\sim}{A}_2$.

Example 10.15 (Ross, 1995).
We can illustrate the single data sample example using the problem of identifying a triangle, as described in Chapter 6. Suppose the single data sample is described by a data triplet, where the

three coordinates are the angles of a specific triangle, for example, the triangle as shown in Figure 6.2, $x_0 = \{A = 85°, B = 50°, C = 45°\}$. Recall from Chapter 6 that there were five known patterns stored: isosceles, right, right and isosceles, equilateral, and all other triangles. If we take this single triangle and determine its membership in each of the known patterns, we get the following results (as we did before in Chapter 6):

$$\begin{aligned}
\mu_{\underset{\sim}{I}}(85,50,45) &= 1-\frac{5}{60}=0.916,\\
\mu_{\underset{\sim}{R}}(85,50,45) &= 1-\frac{5}{90}=0.94,\\
\mu_{\underset{\sim}{IR}}(85,50,45) &= 1-\max\left[\frac{5}{60},\frac{5}{90}\right]=0.916,\\
\mu_{\underset{\sim}{E}}(85,50,45) &= 1-\frac{1}{180}(40)=0.78,\\
\mu_{\underset{\sim}{T}}(85,50,45) &= \frac{1}{180}\min\left[3.35,\ 3(5),\ \frac{2}{5},\ 40\right]=0.05.
\end{aligned}$$

Using the criterion of maximum membership, we see from these values that x_0 most closely resembles the right-triangle pattern, $\underset{\sim}{R}$.

Now let us extend the paradigm to consider the case where the new data sample is not crisp, but rather a fuzzy set itself. Suppose we have m typical patterns represented as fuzzy sets $\underset{\sim}{A}_i$ on X ($i = 1, 2, \ldots, m$) and a new piece of data, perhaps consisting of a group of observations, is represented by a fuzzy set $\underset{\sim}{A}_i$ on X. The task now is to find which $\underset{\sim}{A}_i$ the sample $\underset{\sim}{B}$ most closely matches. To address this issue, we develop the notion of *fuzzy vectors*.

There are some interesting features and operations on fuzzy vectors, which will become quite useful in the discipline of fuzzy pattern recognition (Dong, 1986). Formally, a vector $\underset{\sim}{\mathbf{a}} = (a_1, a_2, \ldots, a_n)$ is called a *fuzzy vector* if for any element we have $0 \le a_i \le 1$ for $i = 1, 2, \ldots, n$. Similarly, the transpose of the fuzzy vector $\underset{\sim}{\mathbf{a}}$ denoted $\underset{\sim}{\mathbf{a}}^{\mathrm{T}}$, is a column vector if $\underset{\sim}{\mathbf{a}}$ is a row vector, that is,

$$\underset{\sim}{\mathbf{a}}^{\mathrm{T}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Let us define $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$ as fuzzy vectors of length n, and define

$$\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \bigvee_{i=1}^{n} (a_i \wedge b_i) \tag{10.43}$$

as the fuzzy *inner product* of $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$, and

$$\underset{\sim}{\mathbf{a}} \otimes \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \bigwedge_{i=1}^{n} (a_i \vee b_i) \tag{10.44}$$

as the fuzzy *outer product* of $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$.

Example 10.16

We have two fuzzy vectors of length 4 as defined here and want to find the inner product and the outer product for these two fuzzy vectors:

$$\underset{\sim}{\mathbf{a}} = (0.3.0.7.1.0.4) \text{ and } \mathrm{b} = (0.5.0.9.0.3.0.1).$$

$$\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = (0.3.0.7.1.0.0.4)\begin{pmatrix} 0.5 \\ 0.9 \\ 0.3 \\ 0.1 \end{pmatrix}$$

$$= (0.3 \wedge 0.5) \vee (0.7 \wedge 0.9) \vee (1 \wedge 0.3) \vee (0.4 \wedge 0.1) = 0.3 \vee 0.7 \vee 0.3 \vee 0.1 = 0.7.$$

$$\underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = (0.3 \vee 0.5) \wedge (0.7 \vee 0.9) \wedge (1 \vee 0.3) \wedge (0.4 \vee 0.1) = 0.5 \wedge 0.9 \wedge 1 \wedge 0.4 = 0.4.$$

The symbol $\oplus$ has also been used in the literature to describe the Boolean outer product. In this context, we will use this symbol to refer to the outer product of two fuzzy vectors. An interesting feature of these products is found in comparing them to standard algebraic operations on vectors in physics. Whereas the inner and outer products on fuzzy vectors result in scalar quantities, only the algebraic inner product on vectors in physics produces a scalar; the outer product on two vectors in physics produces another vector, whose direction is orthogonal to the plane containing the original two vectors.

We now define the complement of the fuzzy vector, or fuzzy complement vector, as

$$\underset{\sim}{\bar{\mathbf{a}}} = (1 - a_1, 1 - a_2, \ldots, 1 - a_n) = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n). \tag{10.45}$$

It should be obvious that since $\underset{\sim}{\bar{\mathbf{a}}}$ is subject to the constraint $0 \leq \bar{a}_i \leq 1$ for $i = 1, 2, \ldots, n$, the fuzzy complement vector is also another fuzzy vector. Moreover, we define the largest element $\hat{a}$ in the fuzzy vector $\underset{\sim}{\mathbf{a}}$ as its *upper bound*, that is,

$$\hat{a} = \max_i (a_i) \tag{10.46}$$

and the smallest element $\underset{\wedge}{a}$ in the fuzzy vector $\underset{\sim}{\mathbf{a}}$ as its *lower bound*, that is,

$$\underset{\wedge}{a} = \min_i (a_i). \tag{10.47}$$

Some properties of fuzzy vectors that will become quite useful in the area of pattern recognition will be summarized here. For two fuzzy vectors, $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$, both of length n, the following properties hold:

$$\underset{\sim}{\bar{\mathbf{a}}} \bullet \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \underset{\sim}{\bar{\mathbf{a}}} \oplus \underset{\sim}{\bar{\mathbf{b}}}^{\mathrm{T}} \text{ and alternatively } \underset{\sim}{\bar{\mathbf{a}}} \oplus \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \underset{\sim}{\bar{\mathbf{a}}} \bullet \underset{\sim}{\bar{\mathbf{b}}}^{\mathrm{T}}. \tag{10.48}$$

$$\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = (\hat{a} \wedge \hat{b}) \text{ and alternatively } \underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^{\mathrm{T}} \geq (\underset{\wedge}{a} \vee \underset{\wedge}{b}). \tag{10.49}$$

$$\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{a}}^{\mathrm{T}} = \hat{a} \text{ and } \underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{a}}^{\mathrm{T}} = \underset{\wedge}{a}. \tag{10.50}$$

$$\text{For } \underset{\sim}{\mathbf{a}} \subseteq \underset{\sim}{\mathbf{b}} \text{ then } \underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \hat{a} \text{ and for } \underset{\sim}{\mathbf{b}} \subseteq \underset{\sim}{\mathbf{a}} \text{ then } \underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^{\mathrm{T}} = \underset{\wedge}{a}. \tag{10.51}$$

$$\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\bar{\mathbf{a}}} \le \frac{1}{2} \text{ and } \underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\bar{\mathbf{a}}} \ge \frac{1}{2}. \tag{10.52}$$

From the fuzzy vector properties given in Equations (10.48) to (10.52), one can show (Problems 10.9 to 10.11) that when two separate fuzzy vectors are identical, that is, $\underset{\sim}{\mathbf{a}} = \underset{\sim}{\mathbf{b}}$, the inner product $\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^T$ reaches a maximum while the outer product $\underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^T$ reaches a minimum. This result is extremely powerful when used in any problem requiring a metric of similarity between two vectors. If two vectors are identical, the inner product metric will yield a maximum value, and if the two vectors are completely dissimilar the inner product will yield a minimum value. This chapter makes use of the inverse duality between the inner product and the outer product for fuzzy vectors and fuzzy sets in developing an algorithm for pattern recognition. These two norms, the inner product and the outer product, can be used simultaneously in pattern recognition studies because they measure *closeness* or *similarity*.

We can extend fuzzy vectors to the case of fuzzy sets. Whereas vectors are defined on a finite countable universe, sets can be used to address infinite-valued universes (see example that follows using Gaussian membership functions). Let $P^*(X)$ be a group of fuzzy sets with $\underset{\sim}{A}_i \neq \emptyset$ and $\underset{\sim}{A}_i \neq X$. Now we define two fuzzy sets from this family of sets, that is, $\underset{\sim}{A}, \underset{\sim}{B} \in P^*(X)$; then, either of the expressions (Equations (10.53) and (10.54))

$$(\underset{\sim}{A}, \underset{\sim}{B})_1 = (\underset{\sim}{A} \bullet \underset{\sim}{B}) \wedge (\overline{\underset{\sim}{A} \oplus \underset{\sim}{B}}) \tag{10.53}$$

$$(\underset{\sim}{A}, \underset{\sim}{B})_2 = \frac{1}{2}\left[(\underset{\sim}{A} \bullet \underset{\sim}{B}) + (\overline{\underset{\sim}{A} \oplus \underset{\sim}{B}})\right] \tag{10.54}$$

describes two metrics to assess the degree of similarity of the two sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$:

$$(\underset{\sim}{A}, \underset{\sim}{B}) = (\underset{\sim}{A}, \underset{\sim}{B})_1 \text{ or } (\underset{\sim}{A}, \underset{\sim}{B}) = (\underset{\sim}{A}, \underset{\sim}{B})_2. \tag{10.55}$$

In particular, when either of the values of $(\underset{\sim}{A}, \underset{\sim}{B})$ from Equation (10.55) approaches 1, the two fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are "more closely similar"; when either of the values $(\underset{\sim}{A}, \underset{\sim}{B})$ from Equation (10.55) approaches a value of 0, the two fuzzy sets are "more far apart" (dissimilar). The metric in Equation (10.53) uses a minimum property to describe similarity, and the expression in Equation (10.54) uses an arithmetic metric to describe similarity. It can be shown (Problem 10.13) that the first metric (Equation (10.53)) always gives a value that is less than the value obtained from the second metric (Equation (10.54)). And, the second metric (Equation (10.54)) is not useful if the intersection of the two sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$ is null. Both these metrics represent a concept that has been called the *approaching degree* (Wang, 1983).

Example 10.17

Suppose we have a universe of five discrete elements, $X = \{x_1, x_2, x_3, x_4, x_5\}$, and we define two fuzzy sets, $\underset{\sim}{A}$ and $\underset{\sim}{B}$ on this universe. Note that the two fuzzy sets are special: They are actually crisp sets and both are complements of one another:

$$\underset{\sim}{A} = \left\{\frac{1}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{0}{x_4} + \frac{0}{x_5}\right\}$$

$$\underset{\sim}{B} = \underset{\sim}{\bar{A}} = \left\{\frac{0}{x_1} + \frac{0}{x_2} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{1}{x_5}\right\}.$$

If we calculate the quantities expressed in Equations (10.53) to (10.55), we obtain the following values:

$$\underset{\sim}{A} \bullet \underset{\sim}{B} = 0, \ \underset{\sim}{A} \oplus \underset{\sim}{B} = 1, \ (\underset{\sim}{A}, \ \underset{\sim}{B})_1 = (\underset{\sim}{A}, \ \underset{\sim}{B})_2 = 0$$

The conclusion is that a crisp set and its complement are completely dissimilar.

The value of the approaching degree in the previous example should be intuitive. Since each set is the crisp complement of the other, they should be considered distinctly different patterns, that is, there is no overlap. The inner product being zero and the outer product being unity confirm this mathematically. Conversely, if we assume fuzzy set $\underset{\sim}{B}$ to be identical to A, that is, $\underset{\sim}{B} = \underset{\sim}{A}$, then we would find that the inner product equals 1and the outer product equals 0 and the approaching degree (Equations (10.55)) would equal unity. The reader is asked to confirm this (Problem 10.12). This proof simply reinforces the notion that a set is most similar to itself. The proof is valid only for "normal fuzzy sets" (see definition in Chapter 4).

Example 10.18 (Ross, 1995).
Suppose we have a one-dimensional universe on the real line, X = [-∞, ∞], and we define two fuzzy sets having normal, Gaussian membership functions, $\underset{\sim}{A}$, $\underset{\sim}{B}$, which are defined mathematically as

$$\mu_{\underset{\sim}{A}}(x) = \exp\left[\frac{-(x-a)^2}{\sigma_a^2}\right] \text{ and } \mu_{\underset{\sim}{B}}(x) = \exp\left[\frac{-(x-b)^2}{\sigma_b^2}\right]$$

and shown graphically in Figure 10.9. It can be shown that the inner product of the two fuzzy sets is equal to

$$\underset{\sim}{A} \bullet \underset{\sim}{B} = \exp\left[\frac{-(a-b)^2}{(\sigma_a+\sigma_b)^2}\right] = \mu_{\underset{\sim}{A}}(x_0) = \mu_{\underset{\sim}{B}}(x_0),$$

$$\text{where } x_0 = \frac{\sigma_a \cdot b + \sigma_b \cdot a}{\sigma_a + \sigma_b},$$

and that the outer product is calculated to be $\underset{\sim}{A} \oplus \underset{\sim}{B} = 0$. Hence, the values of Equations (10.53) and (10.54) are

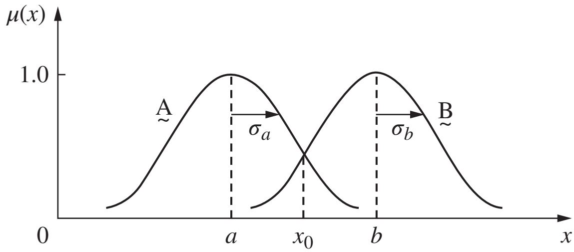

Figure 10.9 Two Gaussian membership functions.

$$(\underset{\sim}{A}, \underset{\sim}{B})_1 = \exp\left[\frac{-(a-b)^2}{(\sigma_a+\sigma_b)^2}\right] \wedge 1 \text{ and } (\underset{\sim}{A}, \underset{\sim}{B})_2 = \frac{1}{2}\left\{\exp\left[\frac{-(a-b)^2}{(\sigma_a+\sigma_b)^2}\right]+1\right\}.$$

The preceding material has presented some examples in which a new data sample is compared to a single known pattern. In the usual pattern recognition problem, we are interested in comparing a data sample to a number of known patterns. Suppose we have a collection of m patterns, each represented by a fuzzy set, $\underset{\sim}{A}_i$, where $i = 1, 2, \ldots, m$, and a sample pattern $\underset{\sim}{B}$, all defined on universe X. Then, the question is as follows: Which known pattern $\underset{\sim}{A}_i$, does data sample $\underset{\sim}{B}_i$ most closely resemble? A useful metric that has appeared in the literature is to compare the data sample to each of the known patterns in a pairwise fashion, determine the approaching degree value for each of these pairwise comparisons, and then select the pair with the largest approaching degree value as the one governing the pattern recognition process. The known pattern that is involved in the maximum approaching degree value is then the pattern the data sample most closely resembles in a maximal sense. This concept has been termed the *maximum approaching degree* (Wang, 1983). Equation (10.56) shows this concept for m known patterns:

$$(\underset{\sim}{B}, \underset{\sim}{A}_i) = \max\{(\underset{\sim}{B}, \underset{\sim}{A}_1),\ (\underset{\sim}{B}, \underset{\sim}{A}_2),\ \ldots,\ (\underset{\sim}{B}, \underset{\sim}{A}_m)\}. \tag{10.56}$$

Example 10.19 (Ross, 1995).

Suppose you are an earthquake engineering consultant hired by the state of California to assess earthquake damage in a region just hit by a large earthquake. Your assessment of damage will be very important to residents of the area because insurance companies will base their claim payouts on your assessment. You must be as impartial as possible. From previous historical records you determine that the six categories of the modified Mercalli intensity (I) scale (VI) to (XI) are most appropriate for the range of damage to the buildings in this region. These damage patterns can all be represented by Gaussian membership functions, $\underset{\sim}{A}_i$, $i = 1, 2, \ldots, 6$, as follows:

$$\mu_{\underset{\sim}{A}}(x) = \exp\left(\frac{-(x-a_i)^2}{\sigma_a^2}\right).$$

where parameters a_i and σ_{a_i} define the shape of each membership function. Your historical database provides the information shown in Table 10.12 for the parameters for the six regions.

You determine via inspection that the pattern of damage to buildings in a given location is represented by a fuzzy set $\underset{\sim}{B}$, with the following characteristics:

$$\mu_{\underset{\sim}{B}}(x) = \exp\left(\frac{-(x-b)^2}{\sigma_b^2}\right); \quad b = 41 \text{ and } \sigma_b = 10.$$

Table 10.12 Parameters for Gaussian membership functions.

	$\underset{\sim}{A}_1$, VI	$\underset{\sim}{A}_2$, VII	$\underset{\sim}{A}_3$, VIII	$\underset{\sim}{A}_4$, IX	$\underset{\sim}{A}_5$, X	$\underset{\sim}{A}_6$, XI
a_i	5	20	35	49	71	92
σ_{a_i}	3	10	13	26	18	4

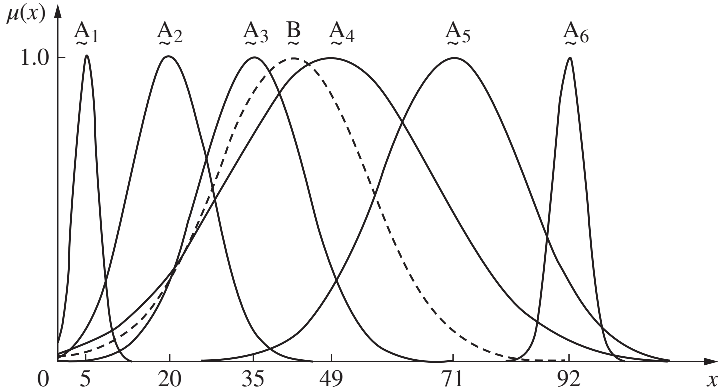

Figure 10.10 Six known patterns and a new fuzzy set data sample.

The system you now have is shown graphically in Figure 10.10. You then conduct the following calculations, using the similarity metric from Equation (10.54), to determine the maximum approaching degree:

$$(\underset{\sim}{B}, \underset{\sim}{A}_1) = \frac{1}{2}(0.004 + 1) \approx 0.5 \qquad (\underset{\sim}{B}, \underset{\sim}{A}_4) = 0.98$$
$$(\underset{\sim}{B}, \underset{\sim}{A}_2) = 0.67 \qquad (\underset{\sim}{B}, \underset{\sim}{A}_5) = 0.65$$
$$(\underset{\sim}{B}, \underset{\sim}{A}_3) = 0.97 \qquad (\underset{\sim}{B}, \underset{\sim}{A}_6) = 0.5.$$

From this list we see that Mercalli intensity IX $(\underset{\sim}{A}_4)$ most closely resembles the damaged area because of the maximum membership value of 0.98.

Suppose you assume the membership function of the damaged region to be a simple singleton with the following characteristics:

$$\mu_{\underset{\sim}{B}}(41) = 1 \text{ and } \mu_{\underset{\sim}{B}}(x \neq 41) = 0, \quad x_0 = 41$$

as shown in Figure 10.10. This example reduces to the single data sample problem posed previously, that is,

$$(\underset{\sim}{B}, \underset{\sim}{A}_i) = \mu_{\underset{\sim}{A}_i}(x_0) \wedge 1 = \mu_{\underset{\sim}{A}_i}(x_0).$$

Your calculations, again using Equation (10.54), produce the following results:

$$\mu_{\underset{\sim}{A}_1}(41) \approx 0 \qquad \mu_{\underset{\sim}{A}_4}(41) = .91$$
$$\mu_{\underset{\sim}{A}_2}(41) = 0.01 \qquad \mu_{\underset{\sim}{A}_5}(41) = 0.06$$
$$\mu_{\underset{\sim}{A}_3}(41) \approx 0.81 \qquad \mu_{\underset{\sim}{A}_6}(41) = 0$$

Again, Mercalli scale IX $(\underset{\sim}{A}_4)$ would be chosen on the basis of maximum membership (0.91). If we were to make the selection without regard to the shapes of the membership

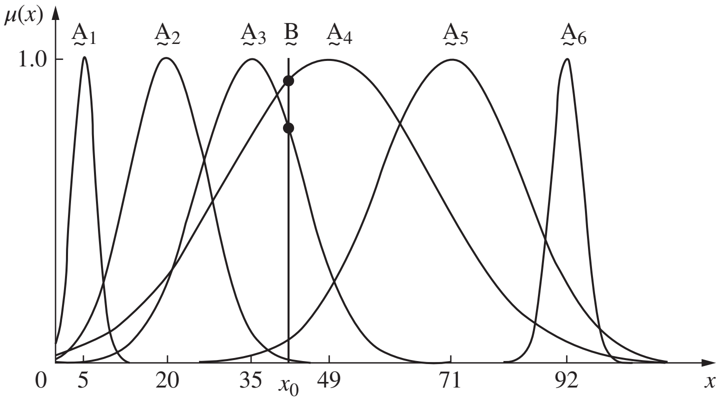

Figure 10.11 Six known patterns and a new singleton data sample.

functions. as shown in Figure 10.11, but instead only considered the mean value of each region, we would be inclined erroneously to select region VIII because its mean value of 35 is closer to the singleton at 41 than it is to the mean value of region IX, that is, to 49.

Multifeature Pattern Recognition

In the material covered so far in this chapter, we have considered only one-dimensional pattern recognition; that is, the patterns here have been constructed based only on a single feature, such as Mercalli earthquake intensity. Suppose the preceding example on earthquake damage also considered, in addition to earthquake intensity, the importance of the particular building (schools vs industrial plants), seismicity of the region, previous history of damaging quakes, and so forth. How could we address the consideration of many features in the pattern recognition process? The literature develops many answers to this question, but this text summarizes three popular and easy approaches: (1) nearest neighbor classifier, (2) nearest center classifier, and (3) weighted approaching degree. The first two methods are restricted to the recognition of crisp singleton data samples.

In the *nearest neighbor classifier*, we can consider m features for each data sample. So, each sample $(\underset{\sim}{x}_i)$ is a vector of features

$$\underset{\sim}{x}_i = \{x_{i1},\ x_{i2},\ x_{i3}, \ldots, x_{i\ m}\}. \tag{10.57}$$

Now, suppose we have n data samples in a universe, or $X = \{\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3,\ \ldots,\ \mathbf{x}_n\}$. Using a conventional fuzzy classification approach, we can cluster the samples into c-fuzzy partitions, then get c-hard partitions from these by using the equivalent relations idea or by “hardening” the soft partition $\underset{\sim}{U}$, both of which are described in this chapter. This would result in hard classes with the following properties:

$$X = \bigcup_{i=1}^{c} A_i; \quad A_i \cap A_j = \emptyset, \quad i \neq j.$$

Now, if we have a new singleton data sample, say $\mathbf{x}$, then the nearest neighbor classifier is given by the following distance measure, d:

$$d(\mathbf{x}, \mathbf{x}_i) = \min_{1 \le k \le n} \{d(\mathbf{x}, \mathbf{x}_k)\} \tag{10.58}$$

for each of the n data samples where $\mathbf{x}_i \in \mathbf{A}_j$. That is, points $\mathbf{x}$ and $\mathbf{x}_i$ are nearest neighbors, and hence both would belong to the same class.

In another method for singleton recognition, the nearest center classifier method works as follows. We again start with n known data samples, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$, and each data sample is m dimensional (characterized by m features). We then cluster these samples into c classes using a fuzzy classification method such as the fuzzy c-means approach described previously in this chapter. These fuzzy classes each have a class center, so

$$\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_c\}$$

is a vector of the c class centers. If we have a new singleton data sample, say $\mathbf{x}$, the nearest center classifier is then given as

$$d(\mathbf{x}, \mathbf{v}_i) = \min_{1 \le k \le c} \{d(\mathbf{x}, \mathbf{v}_k)\}, \tag{10.59}$$

and now the data singleton, $\mathbf{x}$, is classified as belonging to fuzzy partition, $\underset{\sim}{A}_i$.

In the third method for addressing multifeature pattern recognition for a sample with several (m) fuzzy features, we will use the approaching degree concept again to compare the new data pattern with some known data patterns. Define a new data sample characterized by m features as a collection of noninteractive fuzzy sets, $\underset{\sim}{B} = \{\underset{\sim}{B}_1, \underset{\sim}{B}_2, \ldots, \underset{\sim}{B}_m\}$. Because the new data sample is characterized by m features, each of the known patterns, $\underset{\sim}{A}_i$, is also described by m features. Hence, each known pattern in m-dimensional space is a fuzzy class (pattern) given as $\underset{\sim}{A}_i = \{\underset{\sim}{A}_{i1}, \underset{\sim}{A}_{i2} \ldots, \underset{\sim}{A}_{m}\}$, where $i = 1, 2, \ldots, c$ describes c classes (c patterns). Since some of the features may be more important than others in the pattern recognition process, we introduce normalized weighting factors w_j, where

$$\sum_{j=1}^{m} w_j = 1 \tag{10.60}$$

Then, either Equation (10.53) or (10.54) in the approaching degree concept is modified for each of the known c patterns ($i = 1, 2, \ldots, c$) as

$$(\underset{\sim}{B}, \underset{\sim}{A}_i) \sum_{j=1}^{m} w_j (\underset{\sim}{B}_j, \underset{\sim}{A}_{ij}). \tag{10.61}$$

As before in the maximum approaching degree, sample $\underset{\sim}{B}$ is closest to pattern $\underset{\sim}{A}_j$ when

$$(\underset{\sim}{B}, \underset{\sim}{A}_j) = \max_{1 \le i \le c} \{(\underset{\sim}{B}, \underset{\sim}{A}_i)\}. \tag{10.62}$$

Note that when the collection of fuzzy sets $\underset{\sim}{B} = \{\underset{\sim}{B}_1, \underset{\sim}{B}_2, \ldots, \underset{\sim}{B}_m\}$ reduces to a collection of crisp singletons, that is, $\underset{\sim}{B} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$. Equation (10.61) reduces to

$$\mu_{A_i}(x) = \sum_{j=1}^{m} w_j \cdot \mu_{\underset{\sim}{A}_{ij}}(x_j). \tag{10.63}$$

As before in the maximum approaching degree, sample singleton, **x**, is closest to pattern $\underset{\sim}{A}_j$ when Equation (10.62) reduces to

$$\mu_{\underset{\sim}{A}}(\mathbf{x}) = \max_{1 \le i \le c} \{\mu_{\underset{\sim}{A}} i(\mathbf{x})\}. \tag{10.64}$$

Example 10.20

An example of multifeature pattern recognition is given where $m = 2$; the patterns can be illustrated in three-dimensional images (these are illustrated in Figure 2.19, and discussed in Chapter 2). Suppose we have a new pattern, $\underset{\sim}{B}$, which we wish to recognize by comparing it to other known patterns. This new pattern is characterized by two features; hence, it can be represented by a vector of its two noninteractive projections, $\underset{\sim}{B}_1$ and $\underset{\sim}{B}_2$. That is,

$$\underset{\sim}{B} = \{\underset{\sim}{B}_1, \underset{\sim}{B}_2\},$$

where the noninteractive patterns $\underset{\sim}{B}_1$ and $\underset{\sim}{B}_2$ are defined on their respective universes of discourse, X_1 and X_2. The two projections together produce a three-dimensional pattern in the shape of a pyramid, as shown in Figure 10.12.

Further, suppose that we have two ($c = 2$) patterns to which we wish to compare our new pattern; call them patterns $\underset{\sim}{A}_1$ and $\underset{\sim}{A}_2$. Each of these two known patterns could also be represented by their respective noninteractive projections, as shown in Figure 10.13a and 10.13b, where the projections of each known pattern are also defined on X_1 and X_2.

The last step in this process is to assign weights to the various known patterns. Let us assume that $w_1 = 0.3$and $w_2 = 0.7$, since $0.3 + 0.7 = 1$ by Equation (10.60). We compare the new pattern with the two known patterns using Equation (10.61),

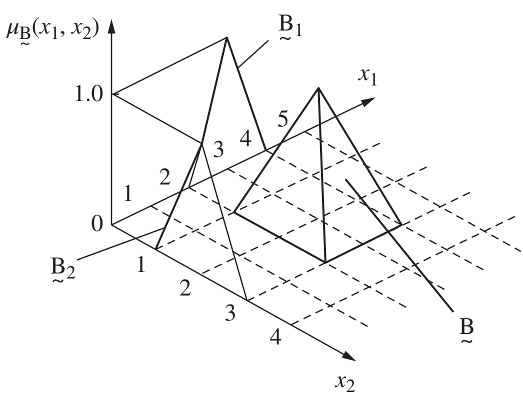

Figure 10.12 New pattern B and its noninteractive projections, $\underset{\sim}{B}_1$ and $\underset{\sim}{B}_2$.

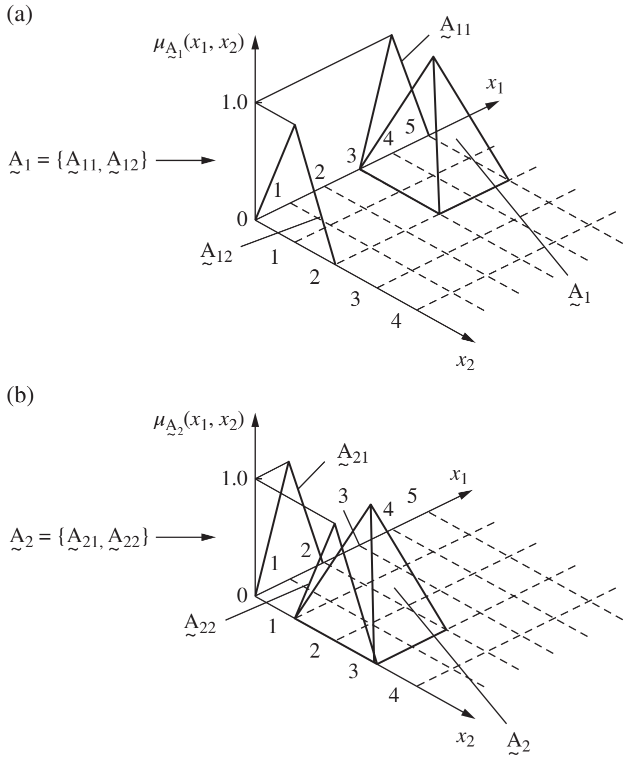

Figure 10.13 Multifeature pattern recognition: (a) known pattern $\underset{\sim}{A}_1$ and its noninteractive projections, $\underset{\sim}{A}_{11}$ and $\underset{\sim}{A}_{12}$; (b) known pattern $\underset{\sim}{A}_2$ and its noninteractive projections, $\underset{\sim}{A}_{21}$ and $\underset{\sim}{A}_{22}$.

$$
\begin{aligned}
(\underset{\sim}{B}, \underset{\sim}{A}_1) &= w_1(\underset{\sim}{B}_1, \underset{\sim}{A}_{11}) + w_2(\underset{\sim}{B}_2, \underset{\sim}{A}_{12}),\\
(\underset{\sim}{B}, \underset{\sim}{A}_2) &= w_1(\underset{\sim}{B}_1, \underset{\sim}{A}_{21}) + w_2(\underset{\sim}{B}_2, \underset{\sim}{A}_{22}),
\end{aligned}
$$

where each of the operations in the preceding expressions is determined using the method of the approaching degree as described in Equation (10.53) or (10.54) for the zth pattern, that is,

$$
(\underset{\sim}{B}, \underset{\sim}{A}_z) = \left[(\underset{\sim}{B} \bullet \underset{\sim}{A}_z) \wedge (\overline{\underset{\sim}{B} \oplus \underset{\sim}{A}_z})\right] \quad \text{or} \quad (\underset{\sim}{B}, \underset{\sim}{A}_z) = \frac{1}{2}\left[(\underset{\sim}{B} \bullet \underset{\sim}{A}_z) + (\overline{\underset{\sim}{B} \oplus \underset{\sim}{A}_z})\right]
$$

Then, we assign the new pattern to the known pattern most closely resembling the new pattern using Equation (10.62), that is,

$$
(\underset{\sim}{B}, \underset{\sim}{A}_z) = \max\left\{(\underset{\sim}{B}, \underset{\sim}{A}_1), (\underset{\sim}{B}, \underset{\sim}{A}_2)\right\}.
$$

The remainder of this example is left as an exercise for the reader.

Although it is not possible to sketch the membership functions for problems dealing with three or more features, the procedures outlined for multifeature pattern recognition work just as they

Table 10.13 Relationships between operation mode and feature values.

Mode (pattern)	Pressure	Temperature	Flow rate
Autoclaving	High	High	Zero
Annealing	High	Low	Zero
Sintering	Low	Zero	Low
Transport	Zero	Zero	High

did with the previous example. The following example in chemical engineering illustrates the multidimensional issues of Equations (10.60) to (10.64).

Example 10.21

A certain industrial production process can be characterized by three features: (1) pressure, (2) temperature, and (3) flow rate. Combinations of these features are used to indicate the current mode (pattern) of operation of the production process. Typical linguistic values for each feature for each mode of operation are defined by the fuzzy sets given in Table 10.13. The pattern recognition task is described as follows: the system *reads* sensor indicators of each feature (pressure, temperature, flow rate), manifested as crisp readout values; it then determines the current mode of operation (i.e., it attempts to recognize a pattern of operation), and then the results are logged.

The four modes (patterns) of operation, and their associated linguistic values for each feature, are as follows:

1. *Autoclaving*: Here the pressure is high, temperature is high, and the flow rate is zero.
2. *Annealing*: Here the pressure is high, temperature is low, and the flow rate is zero.
3. *Sintering*: Here the pressure is low, temperature is zero, and the flow rate is low.
4. *Transport*: Here the pressure is zero, temperature is zero, and the flow rate is high.

This linguistic information is summarized in Table 10.13.

The features of pressure, temperature, and flow rate are expressed in the engineering units of kilopascals (kPa), degrees Celsius (°C), and gallons per hour (gph), respectively. Membership functions for these three features are shown in Figures 10.14–10.16.

Now, suppose the system reads from a group of sensors, a set of crisp readings (pressure = 5kPa, temperature = 150°C, flow = 5gph). We want to assign (recognize) this group of sensor readings to one of our four patterns (modes of operation). To begin, we need to assign weights to each of the features using Equation (10.60). Because there is an explosion hazard associated with the production pressure value (5 kPa), we will weight it more heavily than the other two features:

$$w_{\text{pressure}} = 0.5$$

$$w_{\text{temperature}} = 0.25$$

$$w_{\text{flow}} = 0.25.$$

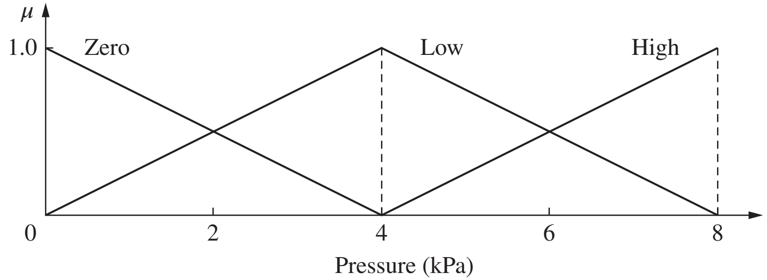

Figure 10.14 Membership functions for pressure.

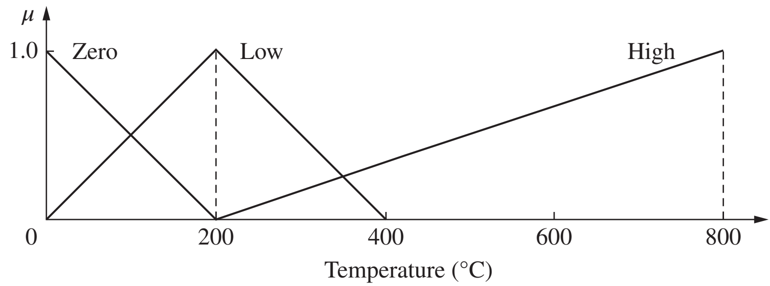

Figure 10.15 Membership functions for temperature.

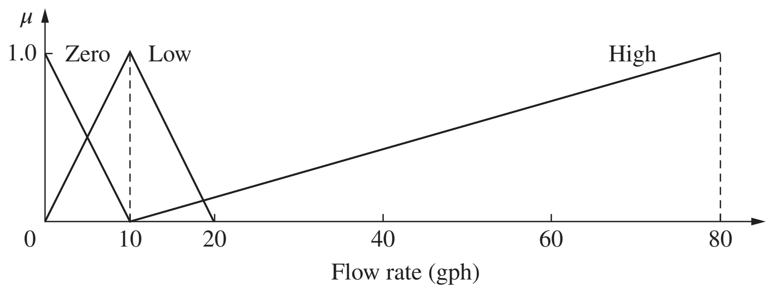

Figure 10.16 Membership functions for flow rate.

Now we will use Equations (10.63) and (10.64) to employ the approaching degree to find which mode of operation is indicated by the above crisp values (5 kPa, 150°C, 5 gph). Using Equation (10.63) and the following two expressions,

$$X = \{5\,\text{kPa},\ 150^\circ\text{C},\ 5\,\text{gph}\}$$

$$W = \{0.5,\ 0.25,\ 0.25\},$$

we find that

$$\mu_{\text{autoclaving}}(x) = (0.5)\cdot(0.25) + (0.25)\cdot(0) + (0.25)\cdot(0.5) = 0.25,$$

$$\mu_{\text{annealing}}(x) = (0.5)\cdot(0.25) + (0.25)\cdot(0.75) + (0.25)\cdot(0.5) = 0.4375,$$

$$\mu_{\text{sintering}}(x) = (0.5)\cdot(0.75) + (0.25)\cdot(0.25) + (0.25)\cdot(0.5) = 0.5625,$$

$$\mu_{\text{transport}}(x) = (0.5)\cdot(0) + (0.25)\cdot(0.25) + (0.25)\cdot(0) = 0.0625,$$

The crisp set, X = {5 kPa, 150°C, 5 gph}, most closely matches the values of pressure, temperature, and flow associated with *sintering*. Therefore, we write the production mode "sintering" in the logbook as the current production mode indicated by crisp readings from our three sensors.

Now, suppose for this industrial process we use the same patterns (autoclaving, annealing, etc.). Suppose now that the readings or information on pressure, temperature, and flow are fuzzy sets rather than crisp singletons, that is, $\underset{\sim}{B} = \{\underset{\sim}{B}_{\text{pressure}}, \underset{\sim}{B}_{\text{temperature}}, \underset{\sim}{B}_{\text{flow}}\}$.

These fuzzy sets are defined in Figures 10.17 to Figure 10.19. Given these fuzzy definitions for our new pattern $\underset{\sim}{B}$, we use Equation (10.62) to find which pattern is best matched by the new values $\underset{\sim}{B}$.

For the approaching degree between our new pattern's pressure feature and the stored autoclaving pattern's pressure feature we get

$$\underset{\sim}{B}_{\text{pressure}} \bullet \text{ autoclaving pressure} = 0(\text{max of mins})$$
$$\underset{\sim}{B}_{\text{pressure}} \oplus \text{ autoclaving pressure} = 0(\text{min of mins})$$

as summarized in Figure 10.20.

For the approaching degree between our new pattern's temperature feature and the stored autoclaving pattern's temperature feature we get

$$\underset{\sim}{B}_{\text{temperature}} \bullet \text{autoclaving temperature} = \max[(0 \wedge 0), (0 \wedge 0.5), (0.166 \wedge 1), (0.33 \wedge 0.5), (0 \wedge 0.5)] = 0.33$$

$$\underset{\sim}{B}_{\text{temperature}} \oplus \text{autoclaving temperature} = \min[(0 \vee 0), (0 \vee 0.5), (0.166 \vee 1), (0.33 \vee 0.5), (0 \vee 0.5)] = 0$$

as summarized in Figure 10.21.

For the approaching degree between our new pattern's flow rate feature and the stored autoclaving pattern's flow rate feature we get

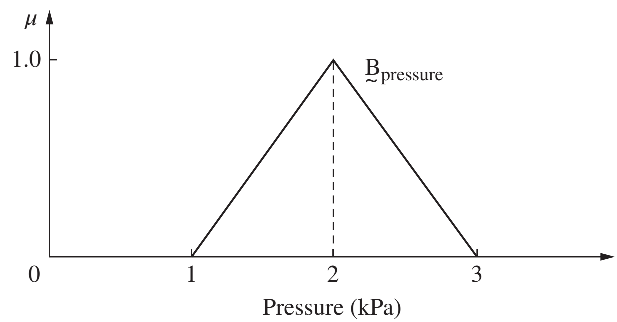

Figure 10.17 Fuzzy sensor reading for pressure.

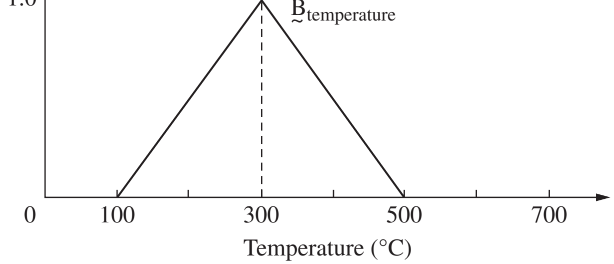

Figure 10.18 Fuzzy sensor reading for temperature.

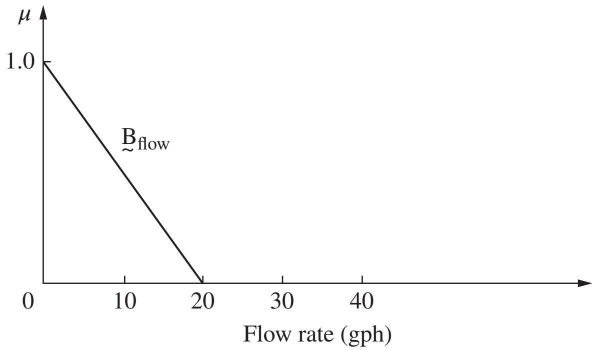

Figure 10.19 Fuzzy sensor reading for flow rate.

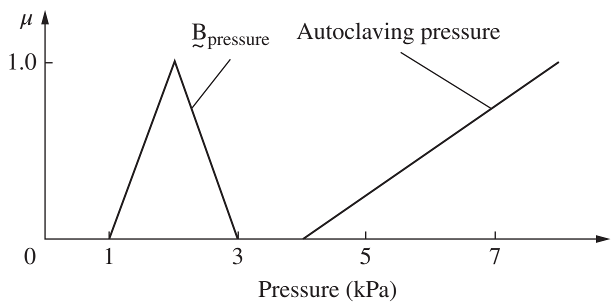

Figure 10.20 Pressure comparisons for autoclaving pattern.

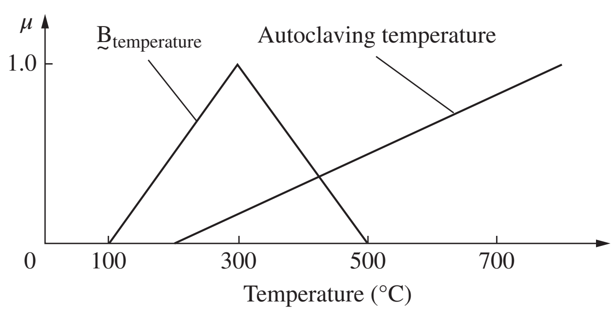

Figure 10.21 Temperature comparisons for autoclaving pattern.

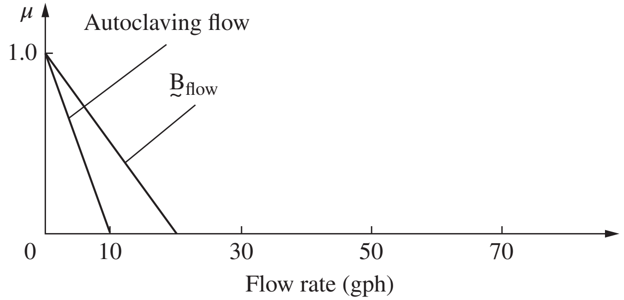

Figure 10.22 Flow rate comparisons for autoclaving pattern.

$$\underset{\sim}{B}_{flow} \bullet \text{autoclaving flow} = \max[(1 \wedge 1),(0 \wedge 0.5),(0 \wedge 0)] = 1.0$$
$$\underset{\sim}{B}_{flow} \oplus \text{autoclaving flow} = \min[(1 \vee 1),(0 \vee 0.5),(0 \vee 0)] = 0$$

as summarized in Figure 10.22.

Now, the use of Equations (10.54) and (10.61) enables us to calculate the approaching degree value between the new sensor pattern and the autoclaving pattern:

$$\begin{aligned}(\underset{\sim}{B}, \text{autoclaving}) &= 0.5\left[\frac{1}{2}(0+1)\right] + 0.25\left[\frac{1}{2}(0.33+1.0)\right] + 0.25\left[\frac{1}{2}(1+1)\right] \\ &= (0.5)(0.5) + (0.25)(0.66) + (0.25)(1) = 0.665.\end{aligned}$$

For the next possible pattern, annealing, we again use Equation (10.62) to determine the approaching degree between the new sensor pressure and the annealing pressure. Because they are disjoint,

$$\underset{\sim}{B}_{pressure} \bullet \text{annealing pressure} = 0$$
$$\underset{\sim}{B}_{pressure} \oplus \text{annealing pressure} = 0$$

as summarized in Figure 10.23.

The approaching degree between the new sensor temperature and the annealing temperature is given as

$$\underset{\sim}{B}_{temperature} \bullet \text{annealing temperature} = 0.75,$$

by inspection of max of mins, and

$$\begin{aligned}&\underset{\sim}{B}_{temperature} \oplus \text{annealing temperature} \\ &= \min[(0 \vee 0),(0 \vee 0.5),(0.5 \vee 1),(1 \vee 0.5),(0.5 \vee 0),(0 \vee 0)] = 0\end{aligned}$$

as summarized in Figure 10.24.

The approaching degree between the new sensor flow rate and the annealing flow rate is given as

$$\underset{\sim}{B}_{flow} \bullet \text{annealing flow} = 1$$
$$\underset{\sim}{B}_{flow} \oplus \text{annealing flow} = 0$$

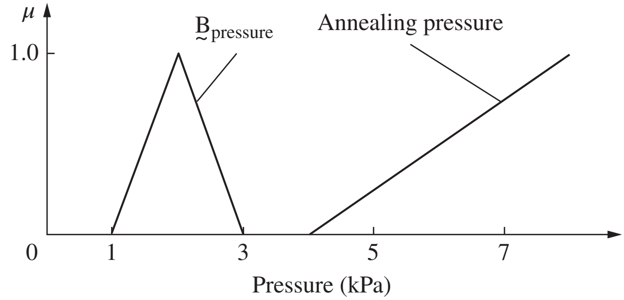

Figure 10.23 Pressure comparisons for annealing pattern.

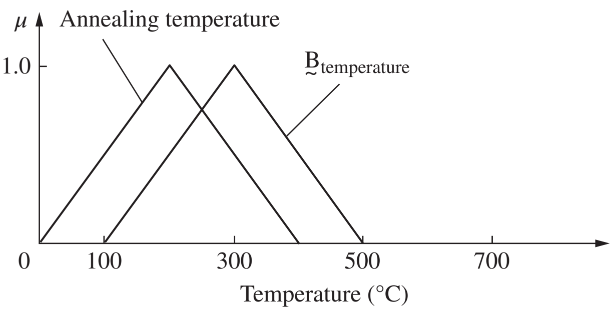

Figure 10.24 Temperature comparisons for annealing pattern.

(identical to $\underset{\sim}{B}_{\text{flow}}$ and autoclaving flow).

Again using Equations (10.54) and (10.61), we get the approaching degree value for the annealing pattern:

$$(\underset{\sim}{B}, \text{annealing}) = 0.5\left[\frac{1}{2}(0+1)\right] + 0.25\left[\frac{1}{2}(0.75+1)\right] + 0.25\left[\frac{1}{2}(1+1)\right]$$

$$= (0.5)(0.5) + (0.25)(0.87) + (0.25)(1) = 0.7175.$$

Now, moving to the next pattern, sintering, we again use Equation (10.62) for each of the features. The first is pressure:

$$\underset{\sim}{B}_{\text{pressure}} \bullet \text{sintering pressure} \approx 0.6,$$

by inspection of max (mins), and

$$\underset{\sim}{B}_{\text{pressure}} \oplus \text{sintering pressure}$$
$$= \min[(0 \vee 0), (0 \vee 0.25), (1 \vee 0.5), (0 \vee 0.75), (0 \vee 1), \ \ldots] = 0$$

as summarized in Figure 10.25.

Next is temperature:

$$\underset{\sim}{B}_{\text{temperature}} \bullet \text{sintering temperature} = 0.25,$$

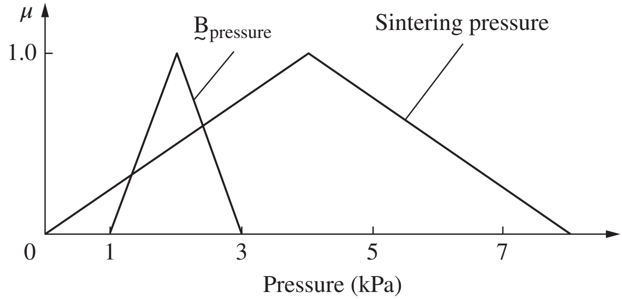

Figure 10.25 Pressure comparisons for sintering pattern.

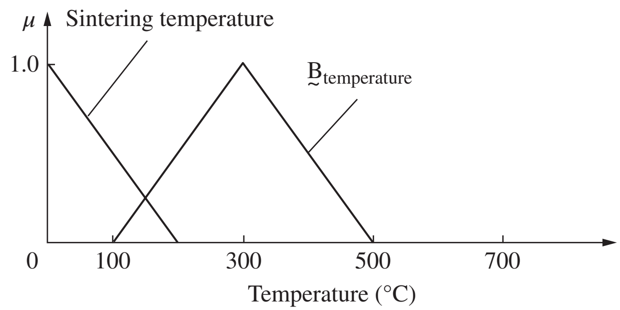

Figure 10.26 Temperature comparisons for sintering pattern.

by inspection of max (mins), and

$$\underset{\sim}{B}_{\text{temperature}} \oplus \text{sintering temperature} = \min[(0 \vee 1), (0 \vee 0.5), (0.5 \vee 0)],$$

as summarized in Figure 10.26.

Finally, we consider the flow rate:

$$\underset{\sim}{B}_{\text{flow}} \bullet \text{sintering flow} = 0.7,$$

by inspection of max (mins), and

$$\underset{\sim}{B}_{\text{flow}} \oplus \text{sintering flow} = \min[(0 \vee 1), (1 \vee 0), (0 \vee 0)] = 0$$

as summarized in Figure 10.27.

Using Equations (10.54) and (10.61) with the metric from Equation (10.54), we get

$$\begin{aligned}(\underset{\sim}{B}, \text{sintering}) &= 0.5\left[\tfrac{1}{2}(0.6+1)\right] + 0.25\left[\tfrac{1}{2}(0.25+1)\right] + 0.25\left[\tfrac{1}{2}(0.7+1)\right] \\ &= (0.5)(0.8) + (0.25)(0.625) + (0.25)(0.85) = 0.7687.\end{aligned}$$

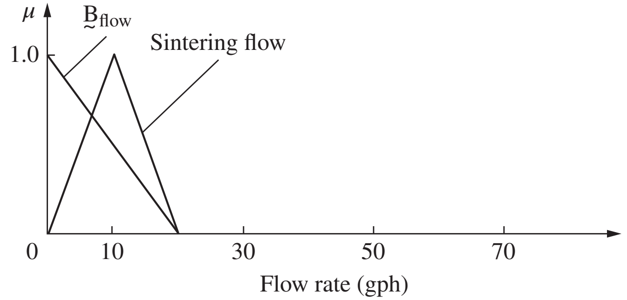

Figure 10.27 Flow rate comparisons for sintering pattern.

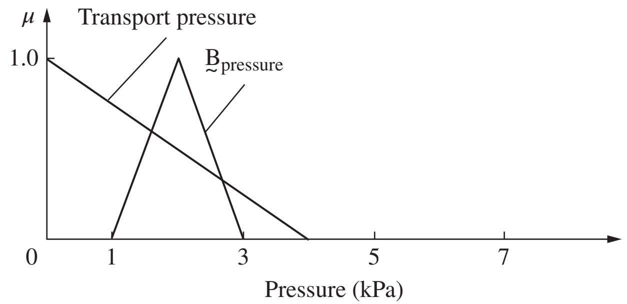

Figure 10.28 Pressure comparisons for transport pattern.

Finally, we consider the last pattern, the transport mode of operation. Using Equation (10.62) for each feature, we begin first with pressure:

$$\underset{\sim}{B}_{pressure} \bullet \text{transport pressure} \approx 0.7,$$

by inspection of max (mins), and

$$\underset{\sim}{B}_{pressure} \oplus \text{transport pressure} = \min[(0 \vee 1),\ (0 \vee 0.75), (1 \vee 0.5), (0.25 \vee 0), (0 \vee 0)] = 0$$

as summarized in Figure 10.28.

Then, moving to temperature:

$$\underset{\sim}{B}_{temperature} \bullet \text{transport temperature} = 0.25$$
$$\underset{\sim}{B}_{temperature} \oplus \text{transport temperature} = 0$$

as summarized in Figure 10.29. And last, moving to flow rate:

$$\underset{\sim}{B}_{flow} \bullet \text{transport flow} = 0.1$$
$$\underset{\sim}{B}_{flow} \oplus \text{transport flow} = 0$$

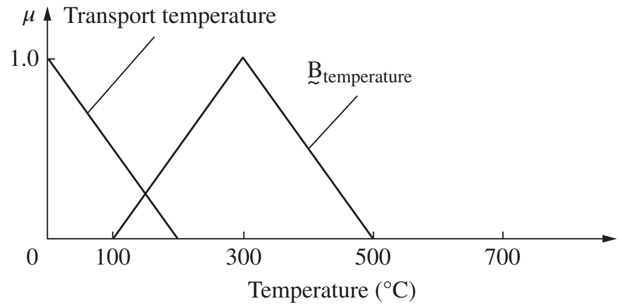

Figure 10.29 Temperature comparisons for transport pattern.

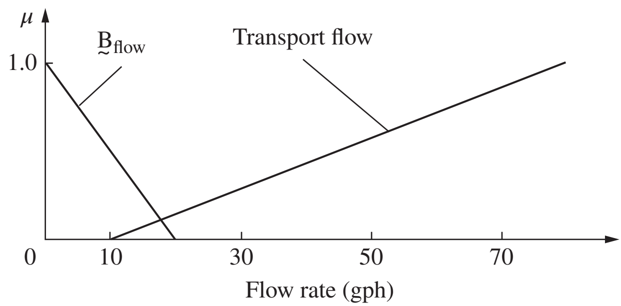

Figure 10.30 Flow rate comparisons for transport pattern.

as summarized in Figure 10.30.

To conclude the calculations using the approaching degree on the last pattern of transport, Equations (10.54) and (10.61) are used to determine

$$\left(\underset{\sim}{B}, \text{transport}\right) = 0.5\left[\tfrac{1}{2}(0.7+1)\right] + 0.25\left[\tfrac{1}{2}(0.25+1)\right] + 0.25\left[\tfrac{1}{2}(0.1+1)\right]$$

$$= (0.5)(0.85) + (0.25)(0.625) + (0.25)(0.55) = 0.7188.$$

Summarizing the results for the four possible patterns, we have

$$(\underset{\sim}{B}, \text{autoclaving}) = 0.665$$
$$(\underset{\sim}{B}, \text{annealing}) = 0.7175$$
$$(\underset{\sim}{B}, \text{sintering}) = 0.7687 (\text{max is here})$$
$$(\underset{\sim}{B}, \text{transport}) = 0.7188$$

The fuzzy readings of pressure, temperature, and flow collectively match most closely, in an approaching degree sense, the *sintering* pattern. We therefore write the production mode “sintering” in our log.

Summary

The concept of a fuzzy set first arose in the study of problems related to pattern classification (Bellman, Kalaba, and Zadeh, 1966). Because the recognition and classification of patterns is integral to human perception, and because these perceptions are fuzzy, this study seems a likely beginning. This chapter has presented a simple idea in the area of classification involving equivalence relations and has dealt in depth with a particular form of classification using a popular clustering method: FCM (fuzzy c-means). The objective in clustering is to partition a given data set into homogeneous clusters; by homogeneous we mean that all points in the same cluster share *similar* attributes and they do not share similar attributes with points in other clusters. However, the separation of clusters and the meaning of *similar* are fuzzy notions and can be described as such. One of the first introductions to the clustering of data was in the area of fuzzy partitions (Ruspini, 1969, 1970, 1973a), where similarity was measured using membership values. In this case, the classification metric was a function involving a distance measure that was minimized. Ruspini (1973b) points out that a definite benefit of fuzzy clustering is that stray points (outliers) or points isolated between clusters (Figure 10.2) may be classified this way; they will have low membership values in the clusters from which they are isolated. In crisp classification methods, these stray points need to belong to at least one of the clusters, and their membership in the cluster to which they are assigned is unity; their distance, or the extent of their isolation, cannot be measured by their membership. The notions of fuzzy classification described in this chapter provides for the point of departure in the recognition of known patterns. A recent example of the use of fuzzy classification to categorize patterns for use in pattern recognition is available in the area of seismic damage to reinforced concrete structures (Elwood and Corotis, 2015). In this work the fuzzy c-means classifier was used to establish patterns of seismically damage buildings based on the features like: building height based on number of stories, building age in years, soil type, and the earthquake intensity.

This chapter has introduced only the most elementary forms of fuzzy pattern recognition. A simple similarity metric called the *approaching degree* (the name is arbitrary; other pseudonyms are possible) is used to assess "closeness" between a known one-dimensional element and an unrecognized one-dimensional element. The idea involved in the approaching degree can be extended to higher-dimensional problems, as illustrated in this chapter, with the use of noninteractive membership functions.

In many recognition problems, structural information plays an important role in describing the patterns. Some examples include image recognition, fingerprint recognition, chromosome analysis, character recognition, scene analysis, and so on. In such cases, when the patterns are complex and the number of possible descriptions is large, it is impractical to regard each description as defining a class; rather, description of the patterns in terms of small sets of simple sub-patterns of primitives and grammatical rules derived from formal language theory becomes necessary (Fu, 1982).

The concept of a grammars was formalized by linguists with a view to finding a means of obtaining structural descriptions of sentences of a language that could not only recognize but also generate the language (Fu, 1982; Hopcroft and Ullman, 1969). Although satisfactory formal grammars have not been obtained to date for describing the English language, the concept of a formal grammar can easily be explained with certain ideas borrowed from English grammar. Some simple illustrations of grammars using fuzzy set theory is available in Ross (2004), and the reader is encouraged to also review Pal and Majumder (1986).

In recent work in pattern recognition, researchers have successfully used the maximum approaching degree for damage pattern recognition. Investigations by El-Kady Reda Taha, and Su (2006) showed the possible use of the maximum approaching degree to quantify submicron damage on the surface of a photonic crystal. Moreover, researchers (Altunok et al., 2006) suggested establishing fuzzy damage states (sets) by solving the inverse problem to satisfy a prescribed approaching degree. Successful damage quantification in structures based on fuzzy damage pattern recognition was also reported (Reda Taha and Lucero, 2005).

References

Altunok, E., Reda Taha, M.M., Epp, D.S., et al. (2006). Damage pattern recognition for structural health monitoring using fuzzy similarity prescription. *J. Comput. Aided Civil Infrastruct. Eng.*, **21**(8), 549– 560.

Bellman, R., Kalaba, R., and Zadeh, L. (1966). Abstraction and pattern classification. *J. Math. Anal. Appl.*, **13**, 1–7.

Bezdek, J. (1974). Numerical taxonomy with fuzzy sets. *J. Math. Biol.*, **1**, 57–71.

Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms* (New York: Plenum).

Bezdek, J., and Harris, J. (1978). Fuzzy partitions and relations: An axiomatic basis for clustering. *Fuzzy Sets Syst.*, **1**, 111–127.

Bezdek, J., Grimball, N., Carson, J., and Ross, T. (1986). Structural failure determination with fuzzy sets. *Civ. Eng. Syst.*, **3**, 82–92.

Dong, W. (1986). Applications of fuzzy sets theory in structural and earthquake engineering. PhD dissertation. Department of Civil Engineering, Stanford University, Stanford, CA.

Duda, R., and Hart, R. (1973). *Pattern Classification and Scene Analysis* (New York: John Wiley & Sons, Inc.).

El-Kady, I., Reda Taha, M.M., and Su, M.F. (2006). Application of photonic crystals in submicron damage detection and quantification. *Appl. Phys. Lett.*, **88**(26), 253109.

Elwood, E., and Corotis, R. (2015). Application of Fuzzy Pattern Recognition of Seismic Damage to Concrete Structures. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A Civil Engineering*, December, **1**(4), pp. 1-12.

Fu, K.S. (1982). *Syntactic Pattern Recognition and Applications* (Englewood Cliffs, NJ: Prentice Hall).

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition* (New York: Academic Press).

Hopcroft, J., and Ullman, J. (1969). *Formal Languages and Their Relation to Automata*, (Reading, MA: Addison-Wesley).

Pal, S., and Majumder, D. (1986). *Fuzzy Mathematical Approach to Pattern Recognition*(New York: John Wiley & Sons, Inc.).

Reda Taha, M.M., and Lucero, J. (2005). Damage identification for structural health monitoring using fuzzy pattern recognition. *Eng. Struct.*, **27**(12), 1774– 1783.

Ross, T. (1995). *Fuzzy Logic with Engineering Applications* (New York: MacGraw-Hill).

Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, 2nd ed. (Chichester, UK: John Wiley & Sons).

Ross, T., Hasselman, T., and Chrostowski, J. (1993). Fuzzy set methods in assessing uncertainty in the modeling and control of space structures. *J. Intell. Fuzzy Syst.*, **1**(2), 135–155.

Ruspini, E. (1969). A new approach to clustering. *Inf. Control*, **15**, 22–32.

Ruspini, E. (1970). Numerical methods for fuzzy clustering. *Inf. Sci.*, **2**, 319–350.

Ruspini, E. (1973a). New experimental results in fuzzy clustering. *Inf. Sci.*, **6**, 273–284.

Ruspini, E. (1973b). *A Fast Method for Probabilistic and Fuzzy Cluster Analysis using Association Measures*. Proceedings of the 6th International Conference on System Sciences, Hawaii, pp. 56–58.

Tamura, S., Higuchi, S., and Tanaka, K. (1971). Pattern classification based on fuzzy relations. *IEEE Trans. Syst., Man, Cybern.*, **1**, 61–66.

Wang, P. (1983). Approaching degree method, in *Fuzzy Sets Theory and Its Applications* [Chinese] (Shanghai, PRC: Science and Technology Press).

Zadeh, L. (1971). Similarity relations and fuzzy orderings. *Inf. Sci.*, **3**, 177–200.

Problems

Exercises for Equivalence Classification

10.1 In a pattern recognition test, four unknown pattern need to be classified according to three known patterns (primitives) a, b, and c. The relationship between primitives and unknown patterns is in the following table:

	x_1	x_2	x_3	x_4
a	0.4	0.1	0.3	0.8
b	0.2	0.7	0.2	0.1
c	0.4	0.2	0.5	0.1

If a λ-cut level is chosen to be 0.5, then into how many classes can these patterns be divided? Hint: Use max-min method (Chapter-3) to first generate a fuzzy similarity relation $\underset{\sim}{R}$.

10.2 As a first step in automatic segmentation of MRI data regarding the head, it is necessary to determine the orientation of a data set to be segmented. The standard radiological orientations are sagittal, coronal, and horizontal. One way to classify the orientation of the new data would be to compare a slice of the new data to slices of known orientation. To do the classification, we will use a simple metric obtained by overlaying slice images and obtaining an area of intersection, then normalizing these based on the largest area of intersection. This metric will be our "degree of resemblance" for the equivalence relation. From the data you have the following fuzzy relation:

$$\begin{array}{r c}
 & \begin{array}{cccc} S & C & H & N \end{array} \\
\begin{array}{r} \textit{Sagittal} \\ \textit{Coronal} \\ \textit{Horizontal} \\ \textit{New Slice} \end{array} &
\begin{bmatrix} 1 & 0.2 & 0.6 & 0.8 \\ 0.2 & 1 & 0.3 & 0.7 \\ 0.6 & 0.3 & 1 & 0.2 \\ 0.8 & 0.7 & 0.2 & 1 \end{bmatrix}
\end{array}$$

a. What kind of relation is this?
b. Determine the equivalence relation and find a classification for λ-cut levels 0.2, 0.5, and 0.7.

Exercises for Fuzzy c-Means

10.3 A problem in construction management is to allocate four different job sites to two different construction teams such that the time wasted in shuttling between the sites is minimized. Let the job sites be designated as x_i where $i = 1, 2, 3, 4$ to give a universe, $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$. If the head office, where the construction teams start every day, has coordinates $\{0, 0\}$, the following vectors give the locations of the four job sites:

$$x_1 = [5, 5]$$
$$x_2 = [6, 8]$$
$$x_3 = [8, 10]$$
$$x_4 = [9, 12]$$

Find fuzzy partition after two cycles, with the initial $\underset{\sim}{U}^0 = \begin{Bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{Bmatrix}$. (Use $m' = 2.0$)

10.4 A radar image of a vehicle is a mapping of the bright (most reflective) parts of it. Suppose we have a radar image that we know contains two vehicles parked close together. The threshold on the instrument has been set such that the image contains seven bright dots. We wish to classify the dots as belonging to one or the other vehicle with a fuzzy membership before we conduct a classification of the vehicle type. The seven bright dots are arranged in a matrix X, and we seek to find an optimum membership matrix U^*. The features defining each of the seven dots are given here:

2	9	9	5	8	5	6
7	3	4	6	8	11	1

Start the calculation with the following initial two partition:

$$\underset{\sim}{U}^0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Find the converged optimal two partition. (Use $m' = 2.0$ and $\varepsilon_L \le 0.01$).

10.5 In a magneto-encephalographic (MEG) experiment, we attempt to partition the space of dipole model order versus reduced chi-square value for the dipole fit. This could be useful to an MEG researcher in determining any trends in his or her data-fitting procedures. Typical ranges for these parameters would be as follows:

$$\text{Dipole model order} = (1, 2, \ldots, 6) = x_{1i},$$
$$\text{Reduced}\,\chi^2 \in (1,3) = x_{2i}.$$

Suppose we have three MEG data points, $\mathbf{x}_i = (x_{1i}, x_{2i})$, i = 1, 2, 3 to classify into two classes. The data are

$$\mathbf{x}_1 = (2, 1.5),\ \mathbf{x}_2 = (3, 2.5) \text{ and } \mathbf{x}_3 = (4, 2).$$

Find the optimum fuzzy 2-partition using the following initial partition:

$$\underset{\sim}{U}(0) = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{Bmatrix} \text{ (Use } m' = 2.0 \text{ and } \varepsilon_L \le 0.01)$$

10.6 A brick manufacturing plant is considering the purchase of a new wire cutting machine. They want to base their decision on cutting time (seconds) and cost ($\$ \times 10^3$). Their decision among five different cutting machines depends on the performance of each machine. The plant has decided to classify the five machines into two classes: good investment and bad investment. The data points in our sample, $X = \{x_1, x_2, x_3, x_4, x_5\}$, are $x_1 = (5,20)$, $x_2 = (3.5,35)$, $x_3 = (4,25)$, $x_4 = (7,10)$ and $x_5 = (8,22)$. Using the initial two partition.

$$\underset{\sim}{U}^{(0)} = \begin{Bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{Bmatrix}$$

Verify that the fuzzy 2-partition after two cycles has not converged, but is

$$\underset{\sim}{U}^{(2)} = \begin{Bmatrix} 0.170 & 0.998 & 0.028 & 0.313 & 0.119 \\ 0.830 & 0.002 & 0.972 & 0.687 & 0.881 \end{Bmatrix} (\text{Use } m' = 2.0 \text{ and } \varepsilon_L \le 0.01)$$

Exercises for Classification Metric and Similarity

10.7 There are many different grades of naphtha, which is a mixture of hydrocarbons characterized by a boiling point range between 80 and 250°C. There are four types of naphtha ($n = 4$) that are characterized based on density, average molecular weight, and hydrogen-to-carbon (H/C) molar ratio ($m = 3$):

	Type I	Type II	Type III	Type IV
Density $\left(\frac{g}{cm^3}\right)$	0.679	0.7056	0.701	0.718
Average mol wt (g)	85.5	93.0	91.0	98.3
$\frac{H}{C}$ *molar ratio*	2.25	2.177	2.253	2.177

There are several studies that predict the products of the naphtha pyrolysis based on light and medium naphtha. It would be useful to classify the above four types into either light or medium classes ($c = 2$). *Note: this problem converges after six iterations, so a computer program for the c-means algorithm is suggested.*

Using $m' = 2$ and $\varepsilon_L = 0.01$ conduct the following:

a. HCM (hard c-means).
b. FCM (fuzzy c-means).
c. Classification metric.
d. Similarity relation that results from the U-partition found in part (b).

10.8 The biomechanical department of a prominent university is conducting research in bone structure. One study involves developing a relationship between the wrist joint angle and the sarcomere length in the lower arm. In this study, the following data were obtained:

Wrist joint angle (degrees)	−70	−40	−25	0	30
Sarcomere length (micro meter)	3	3.25	3.5	2.75	3

a. Classify these data, in one cycle, into two classes using the HCM method.
b. Classify these data into two classes using the FCM method; use $m' = 2$ and $\varepsilon_L = 0.01$ and conduct two cycles. What is the value of the accuracy at the end of two cycles?
c. Find the classification metric.
d. Find the similarity relation for the U-partition that results from part (b).

Exercises for Fuzzy Vectors

10.9 Show that when two separate fuzzy vectors are identical, that is, $\underset{\sim}{\mathbf{a}} = \underset{\sim}{\mathbf{b}}$, the inner product $\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathbf{T}}$ reaches a maximum as the outer product $\underset{\sim}{\mathbf{a}} = \underset{\sim}{\mathbf{b}}$, the inner product $\underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^{\mathbf{T}}$ reaches a minimum.

10.10 For two fuzzy vectors $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$ and the particular case where $\hat{a} = \hat{b} = 1$ and $\underset{\wedge}{a} = \underset{\wedge}{b} = 0$, show that when $\underset{\sim}{\mathbf{a}} = \underset{\sim}{\mathbf{b}}$, then the inner product $\underset{\sim}{\mathbf{a}} \bullet \underset{\sim}{\mathbf{b}}^{\mathbf{T}} = 1$ and the outer product $\underset{\sim}{\mathbf{a}} \oplus \underset{\sim}{\mathbf{b}}^{\mathbf{T}} = 0$.

10.11 For two fuzzy vectors $\underset{\sim}{\mathbf{a}}$ and $\underset{\sim}{\mathbf{b}}$, prove the following expressions (transpose on the second vector in each operation is presumed):

a. $\overline{\underset{\sim}{a} . \underset{\sim}{b}} = \overline{\underset{\sim}{a}} \oplus \overline{\underset{\sim}{b}}$
b. $\underset{\sim}{a} \bullet \underset{\sim}{a} \leq 0.5$

10.12 Prove the following:

a. For any $\underset{\sim}{A} \in P^*(X)$, prove that $(\underset{\sim}{A}, \underset{\sim}{A})_{1\,\mathrm{or}\,2} = 1 - \underset{\wedge}{a}$ for "normal" fuzzy sets.
b. For any $\underset{\sim}{A}$ on X, prove that

$$(\underset{\sim}{A}, \overline{\underset{\sim}{A}})_1 \leq 1/2$$
$$(\underset{\sim}{A}, \overline{\underset{\sim}{A}})_2 \leq 1/2$$

10.13 Show that the metric in Equation (10.53) always gives a value less than or equal to the metric in Equation (10.54) for any pair of fuzzy sets.

Exercises for Maximum Approaching Degree

10.14 Transuranic waste will be store data south eastern New Mexico site known as WIPP (waste isolation pilot plant). The site has underlying strata of rock salt, which is well known for its healing and creeping properties. Healing is the tendency of a material to close fractures or other openings and creep is the capacity of the material to deform under constant load. The radioactive wastes are stored in rooms excavated deep under ground. Because of the creep of the ceiling, these rooms will eventually collapse, thus permanently sealing the wastes in place. The creep properties of salt depend on the depth, moisture content, and clay content of the salt at the location being considered. Rock salt from specified depths was studied through numerous tests conducted at various labs nationwide. These data comprise the known patterns. Hence, each pattern has three features. Now, the possibility of locating a room at a certain depth is being investigated. We wish to determine the creep properties at some depth of salt with a certain clay and moisture content. Membership functions for each of the patterns are shown in Figure P10.14. Find which known pattern the unknown pattern

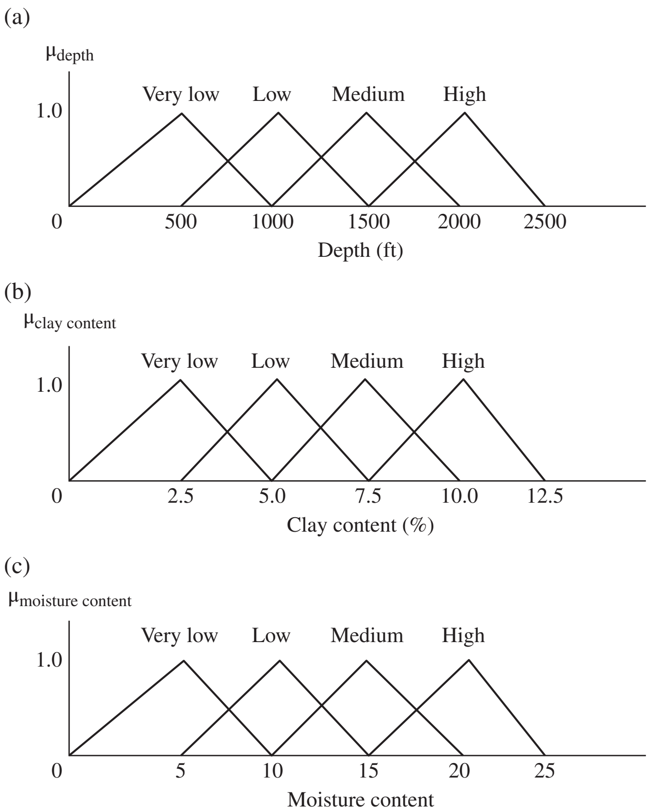

Figure P10.14

matches the best. The features for the unknown pattern are given by the following crisp singleton values:

B = {depth = 1600 ft, clay content = 6.16%, moisture content = 18%}

The weights given to the features are w = {0.4, 0.3, 0.3}

10.15 Using the same known pattern as in Problem 10.14, and using fuzzy features for the new pattern (fuzzy sets B), find which known pattern matches the new pattern most closely.

Features for the new pattern are as follows:

$$\underset{\sim}{B}_{\text{depth}} = \left\{ \frac{0}{1700} + \frac{0.4}{1725} + \frac{1}{1750} + \frac{0.6}{1775} + \frac{0}{1800} \right\},$$

$$\underset{\sim}{B}_{\text{clay content}} = \left\{ \frac{0}{5.5} + \frac{0.7}{5.813} + \frac{1.0}{6.13} + \frac{0.5}{6.44} + \frac{0}{6.75} \right\},$$

$$\underset{\sim}{B}_{\text{moisture content}} = \left\{ \frac{0}{11.0} + \frac{0.5}{11.75} + \frac{1}{12.5} + \frac{0.8}{13.25} + \frac{0}{14.0} \right\}.$$

10.16 A hydrologist is trying to determine how flow velocity and pipe area affect flow rate through pipes. He/she realizes that combinations of these features indicate the current mode (pattern) of flow rate for pipes. Typical linguistic values for each feature for each pattern are defined by Gaussian fuzzy sets. There are three modes of flow rates, and the table below shows how the two features influence the flow rate mode.

	Features	
	Velocity ($m\,s^{-1}$)	Area (cm^2)
Low flow-rate	Low (L-vel)	Small (S-area)
Moderate flow-rate	Medium (M-vel)	Medium (M-area)
High flow-rate	High (H-vel)	Big (B-area)

Presuming each membership function for the two features is a Gaussian, the following table provides the mean, a_i, and spread, σ_{ai}, parameters for each Gaussian feature

$$\mu_{\underset{\sim}{A}} = \exp\left[-\left(\frac{x_j - a_{ij}}{\sigma_{a_{ij}}}\right)^2\right]$$

	(L – vel)	(M – vel)	(H – vel)	(S – vel)	(M – vel)	(B – vel)
a_i	5	10	15	50	100	150
σ_{ai}	3	5	3	26	18	4

Suppose sensors read a set of crisp readings (velocity = 10 m s^{-1}, area = 110 cm^2). Determine the pattern that most closely resembles the crisp input pattern. The weights assigned to the velocity and area features are 0.3 and 0.7, respectively.

10.17 Using the information in Problem 10.16, but selecting a fuzzy input, perform a pattern recognition. The fuzzy input patterns in the form of triangular fuzzy numbers based on the two features are as follows:

$$\underset{\sim}{B}_{\text{Velocity}} = \left\{\frac{0}{7} + \frac{1}{11} + \frac{0}{15}\right\}$$
$$\underset{\sim}{B}_{\text{Area}} = \left\{\frac{0}{90} + \frac{1}{130} + \frac{0}{170}\right\}$$

10.18 Lubricant oils are classified by three features: color, viscosity, and flash point. Depending on the values of these features, the lube oil is classified as 100 neutral (100 N), 150 neutral (150 N), heavy-solvent neutral (HSN), and 500 neutral (500 N). Among the features, color is the most important, followed by viscosity, then flash point. The reason for this ordering is that it is easier to blend lubricant oils to obtain correct viscosity and flash point than it is to blend to obtain proper color. Any material not falling into

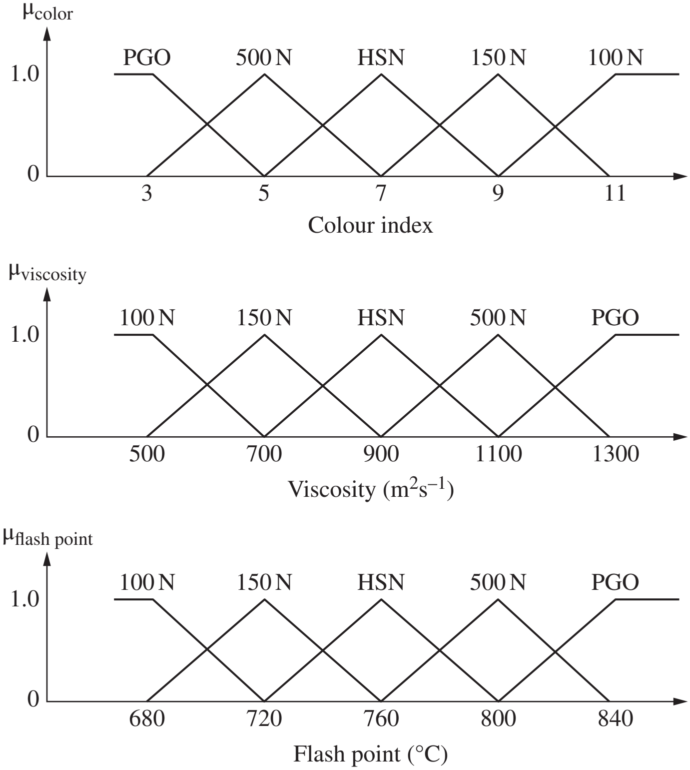

Figure P10.18

one of these lubricant oil categories is downgraded to catalyst cracker feed (petroleum gas oil [PGO]), where it is converted to gasoline.

Fuzzy patterns for each of these features are shown in Figure P10.18. The weights for these features are 0.5 for color, 0.3 for viscosity, and 0.3 for flash point. You receive a lab analysis for a sample described by the crisp singleton,

$$B = \left\{\text{color} = 6.6,\ \text{viscosity} = 840\,\text{m}^2\text{s}^{-1},\ \text{flashpoint} = 780°\text{C}\right\}.$$

Under what category do you classify this sample?

10.19 A traffic engineer is trying to establish the effect of road quality, visibility, and geometry (curves, hills, obstructions) on the traffic capacity of a new road. These three features can be described by Gaussian distributions; the parameters for the distributions are given in the following table. A Gaussian distribution has the form

$$\mu_{\underset{\sim}{A}_{ij}}(X) = \exp\left[-\left(\frac{x_j - a_{ij}}{\sigma_{a_{ij}}}\right)^2\right]$$

The fuzzy patterns are defined on a normalized scale as follows:

	Feature					
	Road quality		Visibility		Road geometry	
	a_i	σ	a_i	σ	a_i	σ
Light capacity	30	15	40	12	20	7
Moderate Capacity	40	20	40	5	35	10
Good capacity	50	15	50	10	60	10
Rush-hour capacity	60	10	60	6	70	15

The weights given to the features are

$$w_{quality} = 0.5$$

$$w_{visibility} = 0.3$$

$$w_{geometry} = 0.2$$

A new road whose features are given on a normalized scale by a crisp singleton,

$$B = \{quality = 45,\ visibility = 55,\ geometry = 35\},$$

is being planned for construction. Determine what capacity of road should be expected.

11

Fuzzy Control Systems

> *The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgment leading to that decision involves the comparison of both paradigms with nature and with each other…the search for assumptions (even for non-existent ones) can be an effective way to weaken the grip of a tradition upon the mind and to suggest the basis for a new one.*
>
> Thomas Kuhn, *The Structure of Scientific Revolutions*, 1962

Control applications are the kinds of problems for which fuzzy logic has had the greatest success and acclaim. Many of the consumer products that we use today involve fuzzy control. And even though fuzzy control is now a standard within industry, the teaching of this subject on academic campuses is still far from being a *standard* offering. But a paradigm shift is being realized in the area of fuzzy control, given its successes for some problems where classical control has not been effective or efficient. In Kuhn's quote, such a paradigm shift can be explained. It was not long ago that fuzzy logic and fuzzy systems were the subject of ridicule and scorn in the scientific communities, but the control community moved quickly in accepting the *new paradigm* and its success is now manifested in the marketplace.

Control systems abound in our everyday life; perhaps we do not see them as such because some of them are larger than what a single individual can deal with, but they are ubiquitous. For example, economic systems are large, global systems that can be controlled; ecosystems are large, amorphous, and long-term systems that can be controlled. Systems that can be controlled have three key features: inputs, outputs, and control parameters (or actions), which are used to perturb the system into some desirable state. The system is monitored in some fashion and left alone if the desired state is realized or perturbed with control actions until the desired state is reached. Usually, the control parameters (actions) are used to perturb the inputs to the system. For example, in the case of economic systems, the inputs might be the balance of trade index, the federal budget deficit, and the consumer price index; outputs might be the inflation rate and the Dow Jones Industrial index; a control parameter might be the federal lending rate that is adjusted occasionally by the U.S. Federal Reserve Board. In the case of ecosystems, the inputs could be the rate of urbanization, automobile traffic, and water use; the outputs could be reductions in green spaces or habitat erosion; a control action could be federal laws and policy on pollution prevention. Other, everyday control situations are evident in our daily lives. Traffic lights are control mechanisms: inputs are arrival rates of cars at an intersection and time of day, outputs are the length of the lines at the lights, and the control parameters are the length of the various light actions (green, yellow, green arrow, etc.). And construction projects involve control scenarios. The inputs on these projects would include the weather, availability of materials, and labor; outputs could be the daily progress toward goals and the dates of key inspections; the control actions could include rewards for finishing on time or early, and penalties for finishing the project late. There are numerous texts that focus just on fuzzy control; a single chapter on this subject could not possibly address all the important topics in this field. References are provided at the end of this chapter for the interested reader. So, in this chapter, we choose to focus on only two types of control: physical system control and industrial process control.

A control system for a physical system is an arrangement of hardware components designed to alter, to regulate, or to command, through a *control action*, another physical system so that it exhibits certain desired characteristics or behavior. Physical control systems are typically of two types: open-loop control systems, in which the control action is independent of the physical system output, and closed-loop control systems (also known as *feedback control systems*), in which the control action depends on the physical system output. Examples of open-loop control systems are a toaster, in which the amount of heat is set by a human, and an automatic washing machine, in which the controls for water temperature, spin-cycle time, and so on are preset by a human. In both these cases, the control actions are not a function of the output of the toaster or the washing machine. Examples of feedback control are a room temperature thermostat, which senses room temperature and activates a heating or cooling unit when a certain threshold temperature is reached, and an autopilot mechanism, which makes automatic course corrections to an aircraft when heading or altitude deviations from certain preset values are sensed by the instruments in the plane's cockpit.

To control any physical variable, we must first measure it. The system for measurement of the *controlled signal* is called a *sensor*. The physical system under control is called a *plant*. In a closed-loop control system, certain forcing signals of the system (the *inputs*) are determined by the responses of the system (the *outputs*). To obtain satisfactory responses and characteristics for the closed-loop control system, it is necessary to connect an additional system, known as a *compensator*, or a *controller*, to the loop. The general form of a closed-loop control system is illustrated in Figure 11.1 (Phillips and Harbor, 1996).

Control systems are sometimes divided into two classes. If the objective of the control system is to maintain a physical variable at some constant value in the presence of disturbances, the system is called a *regulatory* type of control, or a *regulator*. Sometimes this type is also referred to as *disturbance rejection*. The room temperature control and autopilot are examples of regulatory controllers. The second class of control systems is *set-point tracking* controllers. In this scheme of control, a physical variable is required to follow or track some desired time function. An example of this type of system is an automatic aircraft landing system (Example 11.3), in which the aircraft follows a "ramp" to the desired touchdown point.

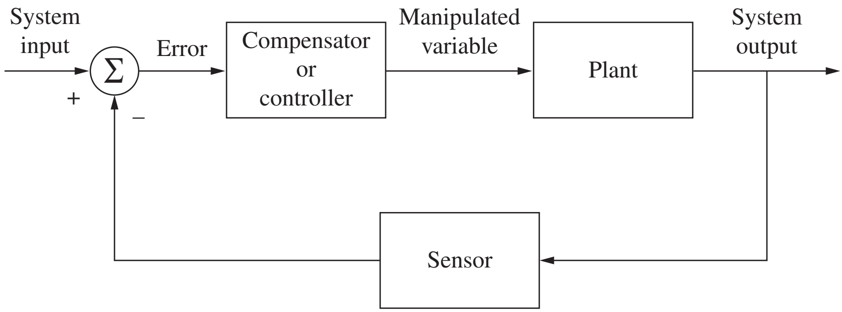

Figure 11.1 A closed-loop control system.

The control problem is stated as follows (Phillips and Harbor, 1996). The output, or response, of the physical system under control (i.e., the plant) is adjusted as required by the *error signal*. The error signal is the difference between the actual response of the plant, as measured by the sensor system, and the desired response, as specified by a *reference input*. In the following section, a typical control system is described (a closed-loop [feedback] control system).

Control System Design Problem

The general problem of feedback control system design is defined as obtaining a generally nonlinear vector-valued function $\mathbf{h}(\cdot)$, defined for some time, t, as follows (Vadiee, 1993):

$$\mathbf{u}(t) = \mathbf{h}[t, \mathbf{x}(\mathrm{t}), \mathbf{r}(\mathrm{t})], \tag{11.1}$$

where $\mathbf{u}(t)$ is the control input to the plant or process, $\mathbf{r}(t)$ is the system reference (desired) input, and $\mathbf{x}(t)$ is the system state vector; the state vector might contain quantities such as the system position, velocity, or acceleration. The feedback control law $\mathbf{h}$ is supposed to stabilize the feedback control system and result in a satisfactory performance.

In the case of a time-invariant system with a regulatory type of controller, where the reference input is a constant set point, the vast majority of controllers is based on one of the general models given in Equations (11.2) and (11.3); that is, either full state feedback or output feedback, as shown in the following:

$$\mathbf{u}(t) = \mathbf{h}[\mathbf{x}(t)], \tag{11.2}$$

$$\mathbf{u}(t) = \mathbf{h}\left[y(t), \dot{y}, \int y\mathrm{d}t\right], \tag{11.3}$$

where $y()$ is the system output or response function. In the case of a simple single-input, single-output (SISO) system and a regulatory type of controller, the function $\mathbf{h}$ takes one of the following forms:

$$\mathbf{u}(t) = K_\mathrm{p} \cdot e(t), \tag{11.4}$$

for a proportional, or P, controller;

$$\mathbf{u}(t) = K_\mathrm{p} \cdot e(t) + K_\mathrm{I} \cdot \int e(t)\mathrm{d}t, \tag{11.5}$$

for a proportional-plus-integral, or PI, controller;

$$\mathbf{u}(t) = K_p \cdot e(t) + K_\mathrm{D} \cdot \dot{e}(t), \tag{11.6}$$

for a proportional-plus-derivative, or PD, controller (Example 11.1 for a PD controller);

$$\mathbf{u}(t) = K_\mathrm{p} \cdot e(t) + K_\mathrm{I} \cdot \int e(t)\mathrm{d}t + K_\mathrm{D} \cdot \dot{e}(t), \tag{11.7}$$

for a proportional-plus-integral-plus-derivative, or PID, controller, where $e(t)$, $\dot{e}(t)$, and $\int e(t)\mathrm{d}t$ are the output error, error derivative, and error integral, respectively, and

$$\mathbf{u}(t) = -[k_1 \cdot x_1(t) + k_2 \cdot x_2(t) + \cdots + k_\mathrm{n} \cdot x_\mathrm{n}(t)], \tag{11.8}$$

for a full state feedback controller.

The problem of control system design is defined as obtaining the generally nonlinear function $\mathbf{h}(\ )$ in the case of nonlinear systems; coefficients K_P, K_I, and K_D in the case of output-feedback systems; and coefficients $k_1, k_2, \ldots, k_n$ in the case of a full state feedback control policy for linear systems. The function $\mathbf{h}(\cdot)$ in Equations (11.2) and (11.3) describes a general nonlinear surface that is known as a control, or decision, surface, discussed in the next section.

Control (Decision) Surface

The concept of a control surface, or decision surface, is central in fuzzy control systems methodology (Ross, 1995). In this section, we define this important concept. The function $\mathbf{h}$ as defined in Equations (11.1) to (11.3) is, in general, defining P nonlinear hypersurfaces in an n-dimensional space. For the case of linear systems with output feedback or state feedback, it is generally a hyperplane in an n-dimensional space. This surface is known as the control, or decision, surface. The control surface describes the dynamics of the controller and is generally a time-varying nonlinear surface. Owing to unmodeled dynamics present in the design of any controller, techniques should exist for adaptively tuning and modifying the control surface shape.

Fuzzy rule-based systems use a collection of fuzzy conditional statements derived from a knowledge base to approximate and construct the control surface (Mamdani and Gaines, 1981; Kiszka, Gupta, and Nikfrouk, 1985; Sugeno, 1985). This paradigm of control system design is based on interpolative and approximate reasoning. Fuzzy rule-based controllers or system identifiers are generally model-free paradigms. Fuzzy rule-based systems are universal nonlinear function approximators, and any nonlinear function (e.g., control surface) of n independent variables and one dependent variable can be approximated to any desired precision.

Alternatively, artificial neural networks are based on analogical learning and trying to learn the nonlinear decision surface through adaptive and converging techniques based on the numerical data available from input–output measurements of the system variables and some performance criteria.

Assumptions in a Fuzzy Control System Design

A number of assumptions are implicit in a fuzzy control system design. Six basic assumptions are commonly made whenever a fuzzy rule-based control policy is selected.

1. The plant is observable and controllable: state, input, and output variables are usually available for observation and measurement or computation.
2. There exists a body of knowledge comprising a set of linguistic rules, engineering common sense, intuition, or a set of input–output measurements data from which rules can be extracted (Chapter 7).
3. A solution exists.
4. The control engineer is looking for a "good enough" solution, not necessarily the optimum one.
5. The controller will be designed within an acceptable range of precision.
6. The problems of stability and optimality are not addressed explicitly; such issues are still open problems in fuzzy controller design.

The following section discusses the procedure for obtaining the control surface, $\mathbf{h}(\cdot)$, from approximations based on a collection of fuzzy IF–THEN rules that describe the dynamics of the controller.

Simple Fuzzy Logic Controllers

First-generation (nonadaptive) simple fuzzy controllers can generally be depicted by a block diagram such as that shown in Figure 11.2.

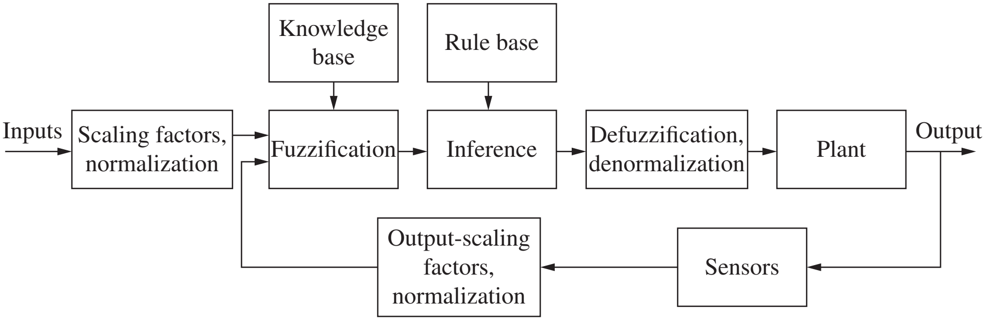

Figure 11.2 A simple fuzzy logic control system block diagram.

The knowledge-base module in Figure 11.2 contains knowledge about all the input and output fuzzy partitions. It will include the term set and the corresponding membership functions defining the input variables to the fuzzy rule-base system and the output variables, or control actions, to the plant under control.

The steps in designing a simple fuzzy control system are as follows:

1. Identify the variables (inputs, states, and outputs) of the plant.
2. Partition the universe of discourse or the interval spanned by each variable into a number of fuzzy subsets, assigning each a linguistic label (subsets include all the elements in the universe).
3. Assign or determine a membership function for each fuzzy subset.
4. Assign the fuzzy relationships between the inputs' or states' fuzzy subsets on the one hand and the outputs' fuzzy subsets on the other hand, thus forming the rule-base.
5. Choose appropriate scaling factors for the input and output variables to normalize the variables to the [0, 1] or the [−1, 1] interval.
6. Fuzzify the inputs to the controller.
7. Use fuzzy approximate reasoning to infer the output contributed from each rule.
8. Aggregate the fuzzy outputs recommended by each rule.
9. Apply defuzzification to form a crisp output.

Examples of Fuzzy Control System Design

Most control situations are more complex than we can deal with mathematically. In this situation, fuzzy control can be developed, provided a body of knowledge about the control process exists, and formed into a number of fuzzy rules. For example, suppose an industrial process output is given in terms of the pressure. We can calculate the difference between the desired pressure and the output pressure, called the *pressure error* (e), and we can calculate the difference between the desired rate of change of the pressure, $\mathrm{d}p/\mathrm{d}t$, and the actual pressure rate, called the *pressure error rate*, ($\dot{e}$). Also, assume that knowledge can be expressed in the form of IF–THEN rules such as

IF pressure error (e) is "positive big (PB)" or "positive medium (PM)" and

IF pressure error rate ($\dot{e}$) is "negative small (NS),"

THEN heat input change is "negative medium (NM)."

The linguistic variables defining the pressure error, "PB" and "PM," and the pressure error rate, "NS" and "NM," are fuzzy, but the measurements of both the pressure and pressure rate as well as the control value for the heat (the control variable) ultimately applied to the system are precise (crisp). The schematic in Figure 11.3 shows this idea. An input to the industrial process (physical system) comes from the controller. The physical system responds with an output, which is sampled and measured by some device. If the measured output is a crisp quantity, it can be fuzzified into a fuzzy set (Chapter 4). This fuzzy output is then considered as the fuzzy input into a fuzzy controller, which consists of linguistic rules. The output of the fuzzy controller is then another series of fuzzy sets. Because most physical systems cannot interpret fuzzy

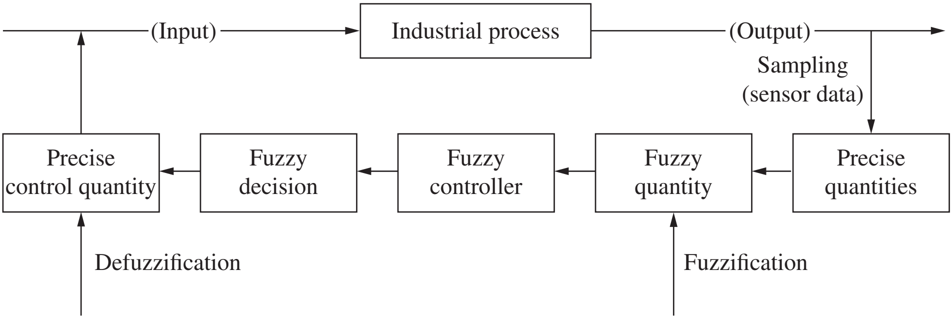

Figure 11.3 Typical closed-loop fuzzy control situation.

commands (fuzzy sets), the fuzzy controller output must be converted into crisp quantities using defuzzification methods (again, see Chapter 4). These crisp (defuzzified) control-output values then become the input values to the physical system and the entire closed-loop cycle is repeated.

Example 11.1

For some industrial plants a human operator is sometimes more efficient than an automatic controller. These intuitive control strategies, which provide a possible method to handle qualitative information, may be modeled by a fuzzy controller. This example looks at a pressure process controlled by a fuzzy controller. The controller is formed by a number of fuzzy rules, such as if pressure error is "positive-big" or "positive-medium," and if the rate of change in the pressure error is "negative-small," then heat input change is "negative medium." This example is illustrated in four steps.

Step 1. Value assignment for the fuzzy input and output variables: We will let the error (e) be defined by eight linguistic variables, labeled $A_1, A_2, \ldots, A_8$, partitioned on the error space of $[-e_m, +e_m]$, and the error rate ($\dot{e}$, or de/dt) be defined by seven variables, labeled $B_1, B_2, \ldots, B_7$, partitioned on the error rate space of $[-\dot{e}_m, \dot{e}_m]$. We will normalize these ranges to the same interval $[-a, +a]$ as

$$e_1 = \left(\frac{a}{e_m}\right) \cdot e,$$

$$\dot{e}_1 = \left(\frac{a}{\dot{e}_m}\right) \cdot \dot{e},$$

For the error, the eight fuzzy variables, $\underset{\sim}{A}_i (i = 1, 2, \ldots, 8)$, will conform to the linguistic variables NB, NM, NS, N0, P0, PS, PM, PB. For the error rate, $\dot{e}$, the seven fuzzy variables, $\underset{\sim}{B}_j (j = 1, 2, \ldots, 7)$, will conform to the linguistic variables NB, NM, NS, 0, PS, PM, PB. The membership functions for these quantities will be on the range $[-a, a]$, where $a = 6$, and are shown in Tables 11.1 and 11.2 (in the tables $x = e$ and $y = \dot{e}$).

The fuzzy output variable, the control quantity (z), will use seven fuzzy variables on the normalized universe, $z = \{-7, -6, -5, \ldots, +7\}$. The control variable will be described by fuzzy linguistic control quantities, $\underset{\sim}{C}_k (k = 1, 2, \ldots, 7)$, which are partitioned on the control universe.

Table 11.1 Membership functions for error (e)[a].

	x	−6	−5	−4	−3	−2	−1	−0	0+	1	2	3	4	5	6
A_i															
A_8	PB	0	0	0	0	0	0	0	0	0	0	0.1	0.4	0.8	1
A_7	PM	0	0	0	0	0	0	0	0	0	0.2	0.7	1	0.7	0.2
A_6	PS	0	0	0	0	0	0	0	0.3	0.8	1	0.5	0.1	0	0
A_5	P0	0	0	0	0	0	0	0	1	0.6	0.1	0	0	0	0
A_4	N0	0	0	0	0	0.1	0.6	1	0	0	0	0	0	0	0
A_3	NS	0	0	0.1	0.5	1	0.8	0.3	0	0	0	0	0	0	0
A_2	NM	0.2	0.7	1	0.7	0.2	0	0	0	0	0	0	0	0	0
A_1	NB	1	0.8	0.4	0.1	0	0	0	0	0	0	0	0	0	0

[a] In the case of crisp control, the membership values in the shaded boxes become unity and all other values become zero.

Table 11.2 Membership functions for error rate ($\mathrm{d}e/\mathrm{d}t$)[a].

	y	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6
B_j														
B_7	PB	0	0	0	0	0	0	0	0	0	0.1	0.4	0.8	1
B_6	PM	0	0	0	0	0	0	0	0	0.2	0.7	1	0.7	0.2
B_5	PS	0	0	0	0	0	0	0	0.9	1	0.7	0.2	0	0
B_4	0	0	0	0	0	0	0.5	1	0.5	0	0	0	0	0
B_3	NS	0	0	0.2	0.7	1	0.9	0	0	0	0	0	0	0
B_2	NM	0.2	0.7	1	0.7	0.2	0	0	0	0	0	0	0	0
B_1	NB	1	0.8	0.4	0.1	0	0	0	0	0	0	0	0	0

[a] In the case of crisp control, the membership values in the shaded boxes become unity and all other values become zero.

Table 11.3 shows the normalized control quantity, z, which is defined by seven linguistic variables.

Step 2. Summary of control rules: According to human operator experience, control rules are of the form

$$\text{If } e \text{ is } \underset{\sim}{A}_1 \text{ and } \dot{e} \text{ is } \underset{\sim}{B}_1, \text{ then } z \text{ is } \underset{\sim}{C}_1.$$
$$\text{If } e \text{ is } \underset{\sim}{A}_1 \text{ and } \dot{e} \text{ is } \underset{\sim}{B}_2, \text{ then } z \text{ is } \underset{\sim}{C}_{12}.$$
$$\text{If } e \text{ is } \underset{\sim}{A}_i \text{ and } \dot{e} \text{ is } \underset{\sim}{B}_j, \text{ then } z \text{ is } \underset{\sim}{C}_k.$$

Each rule can be translated into a fuzzy relation, $\underset{\sim}{R}$. Using such an approach will result in linguistic variables, $\underset{\sim}{C}_k$, shown as control entries in Table 11.4.

Step 3. Conversion between fuzzy variables and precise quantities: From the output of the system we can use an instrument to measure the error (e) and calculate the error rate ($\dot{e}$), both of

Table 11.3 Membership functions for the control quantity $(z)^a$.

	z	−7	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7
C_k																
C_1	PB	0	0	0	0	0	0	0	0	0	0	0	0.1	0.4	0.8	1
C_2	PM	0	0	0	0	0	0	0	0	0	0.2	0.7	1	0.7	0.2	0
C_3	PS	0	0	0	0	0	0	0	0.4	1	0.8	0.4	0.1	0	0	0
C_4	0	0	0	0	0	0	0	0.5	1	0.5	0	0	0	0	0	0
C_5	NS	0	0	0	0.1	0.4	0.8	1	0.4	0	0	0	0	0	0	0
C_6	NM	0	0.2	0.7	1	0.7	0.2	0	0	0	0	0	0	0	0	0
C_7	NB	1	0.8	0.4	0.1	0	0	0	0	0	0	0	0	0	0	0

[a] In the case of crisp control, the membership values in the shaded boxes become unity and all other values become zero.

Table 11.4 Control rules (FAM table).

	A_i							
	NB	NM	NS	N0	P0	PS	PM	PB
B_j								
NB	PB	PM	NB	NB	NB	NB		
NM	PB	PM	NM	NM	NS	NM		
NS	PB	PM	NS	NS	NS	NS	NM	NB
0	PB	PM	PS	0	0	NS	NM	NB
PS	PB	PM	PS	PS	PS	PS	NM	NB
PM			PS	PS	PM	PM	NM	NB
PB			PB	PB	PB	PB	NM	NB

which are precise numbers. A standard defuzzification procedure to develop membership functions, such as the maximum membership principle (Chapter 4), can be used to get the corresponding fuzzy quantities $\left(\underset{\sim}{A}_i, \underset{\sim}{B}_i\right)$. Sending the $\underset{\sim}{A}$ and $\underset{\sim}{B}$ obtained from the output of the system to the fuzzy controller will yield a fuzzy action variable $\underset{\sim}{C}$ (control rules) as discussed in step 2. But before implementing the control, we have to enter the precise control quantity z into the system. We need another conversion from $\underset{\sim}{C}$ to z. This can be done by a maximum membership principle, or by a weighted average method (Chapter 4).

Step 4. Development of control table: When the procedures in step 3 are used for all e and all $\dot{e}$, we obtain a control table as shown in Table 11.5. This table now contains precise numerical quantities for use by the industrial system hardware. If the values in Table 11.5 are plotted, they represent a control surface. Figure 11.4 is the control surface for this example, and Figure 11.5 would be the control surface for this example if it had been conducted using only crisp sets and operations (for the crisp case, the values in Table 11.5 will be different). The volume under a control surface is proportional to the amount of energy expended by the

Table 11.5 Control actions.

	y												
	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6
x													
−6	7	6	7	6	7	7	7	4	4	2	0	0	0
−5	6	6	6	6	6	6	6	4	4	2	0	0	0
−4	7	6	7	6	7	7	7	4	4	2	0	0	0
−3	6	6	6	6	6	6	6	3	2	0	−1	−1	−1
−2	4	4	4	5	4	4	4	1	0	0	−1	−1	−1
−1	4	4	4	5	4	4	1	0	0	0	−3	−2	−1
0+	4	4	4	5	1	1	0	−1	−1	−1	−4	−4	−4
0+	4	4	4	5	1	1	0	−1	−1	−1	−4	−4	−4
1	2	2	2	2	0	0	−1	−4	−4	−3	−4	−4	−4
2	1	1	1	−2	0	−3	−4	−4	−4	−3	−4	−4	−4
3	0	0	0	0	−3	−3	−6	−6	−6	−6	−6	−6	−6
4	0	0	0	−2	−4	−4	−7	−7	−7	−6	−7	−6	−7
5	0	0	0	−2	−4	−4	−6	−6	−6	−6	−6	−6	−6
6	0	0	0	−2	−4	−4	−7	−7	−7	−6	−7	−6	−7

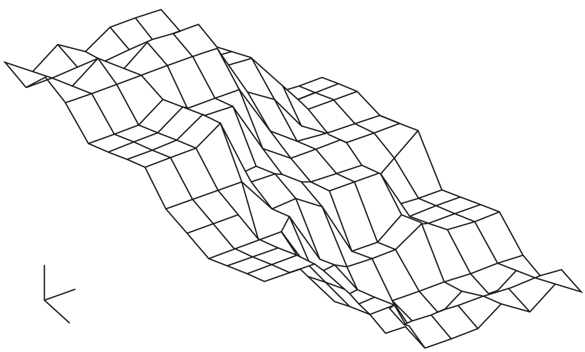

Figure 11.4 Control surface for fuzzy process control in Example 11.1.

controller. It can be shown that the fuzzy control surface (Figure 11.4) will actually fit underneath the crisp control surface (Figure 11.5), which indicates that the fuzzy control expends less energy than the crisp control. Fuzzy control methods, such as this one, have been used for some industrial systems and have achieved significant efficiency (Mamdani, 1974; Pappas and Mamdani, 1976).

In the foregoing example, we did not conduct a simulation of a control process because we do not have a model for the controller. The development of the control surface is derived simply from the control rules and associated membership functions. After the control surface is developed, a simulation can be conducted if a mathematical or linguistic (rule-based) model of the control process is available.

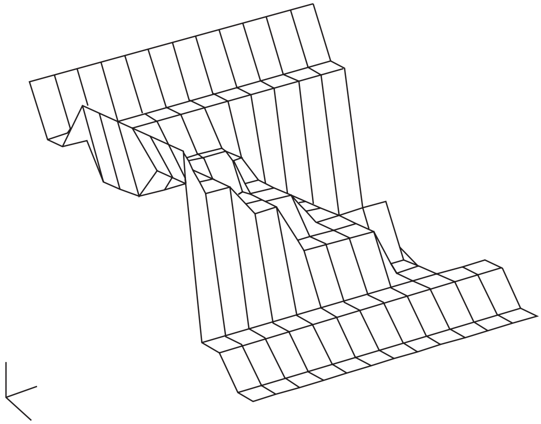

Figure 11.5 Control surface for crisp process control in Example 11.1.

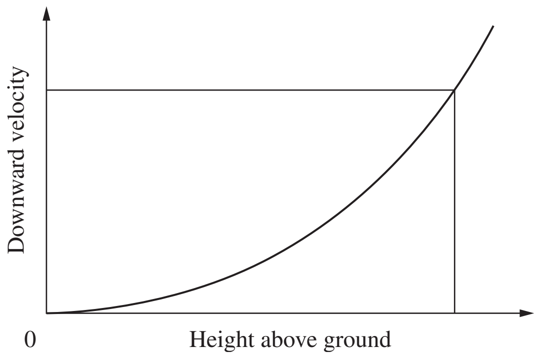

Figure 11.6 The desired profile of downward velocity versus altitude.

Aircraft Landing Control Problem

The following example shows the flexibility and reasonable accuracy of a typical application in fuzzy control.

Example 11.2

We will conduct a simulation of the final descent and landing approach of an aircraft. The desired profile is shown in Figure 11.6. The desired downward velocity is proportional to the square of the height. Thus, at higher altitudes, a large downward velocity is desired. As the height (altitude) diminishes, the desired downward velocity gets smaller and smaller. In the limit, as the height becomes vanishingly small, the downward velocity also goes to zero. In this way, the aircraft will descend from altitude promptly but will touch down very gently to avoid damage.

The two state variables for this simulation will be the height above ground, h, and the vertical velocity of the aircraft, v (Figure 11.7). The control output will be a force that, when applied to

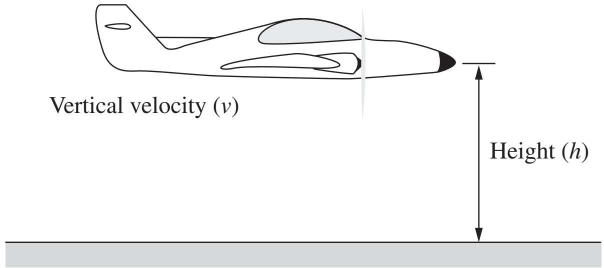

Figure 11.7 Aircraft landing control problem.

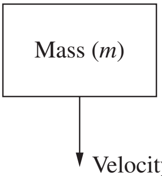

Figure 11.8 Simple momentum model for aircraft landing.

the aircraft, will alter its height, h, and velocity, v. The differential control equations are loosely derived as follows. See Figure 11.8. Mass, m, moving with velocity, v, has momentum, $p = mv$. If no external forces are applied, the mass will continue in the same direction at the same velocity. If a force, f, is applied over a time interval Δt, a change in velocity of $\Delta v = f\Delta t/m$ will result. If we let $\Delta t = 1.0$ (s) and $m = 1.0$ (Ib s^2 ft^{-1}), we obtain $\Delta v = f$ (lb), or the change in velocity is proportional to the applied force.

In difference notation, we get

$$v_{\mathrm{i}+1} = v_i + f_i,$$
$$h_{i+1} = h_i + v_i \cdot \Delta t,$$

where v_{i+1} is the new velocity, v_i is the old velocity, h_{i+1} is the new height, and h_i is the old height. These two "control equations" define the new value of the state variables v and h in response to control input and the previous state variable values. Next, we construct membership functions for the height, h, the vertical velocity, v, and the control force, f:

Step 1. Define membership functions for state variables as shown in Tables 11.6 and 11.7 and Figures 11.9 and 11.10.

Step 2. Define a membership function for the control output, as shown in Table 11.8 and Figure 11.11.

Step 3. Define the rules and summarize them in an FAM table (Table 11.9). The values in the FAM table, of course, are the control outputs.

Step 4. Define the initial conditions, and conduct a simulation for four cycles. Because the task at hand is to control the aircraft's vertical descent during approach and landing, we will

Table 11.6 Membership values for height.

	Height (t)										
	0	100	200	300	400	500	600	700	800	900	1000
Large(L)	0	0	0	0	0	0	0.2	0.4	0.6	0.8	1
Medium(M)	0	0	0	0	0.2	0.4	0.6	0.8	1	0.8	0.6
Small(S)	0.4	0.6	0.8	1	0.8	0.6	0.4	0.2	0	0	0
Nearzero(NZ)	1	0.8	0.6	0.4	0.2	0	0	0	0	0	0

Table 11.7 Membership values for velocity.

	Vertical velocity (ft/s)												
	−30	−25	−20	−15	−10	−5	0	5	10	15	20	25	30
Uplarge(UL)	0	0	0	0	0	0	0	0	0	0.5	1	1	1
Upsmall(US)	0	0	0	0	0	0	0	0.5	1	0.5	0	0	0
Zero(Z)	0	0	0	0	0	0.5	1	0.5	0	0	0	0	0
Downsmall(DS)	0	0	0	0.5	1	0.5	0	0	0	0	0	0	0
Downlarge(DL)	1	1	1	0.5	0	0	0	0	0	0	0	0	0

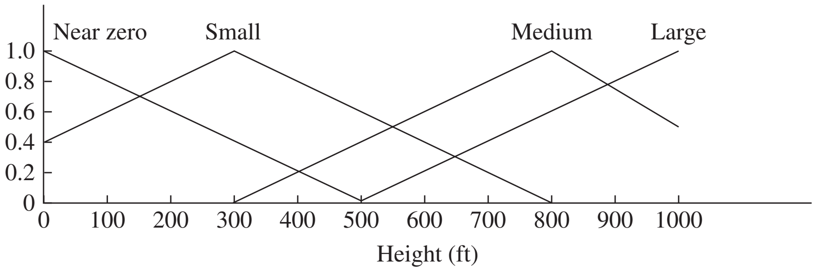

Figure 11.9 Height, h, partitioned.

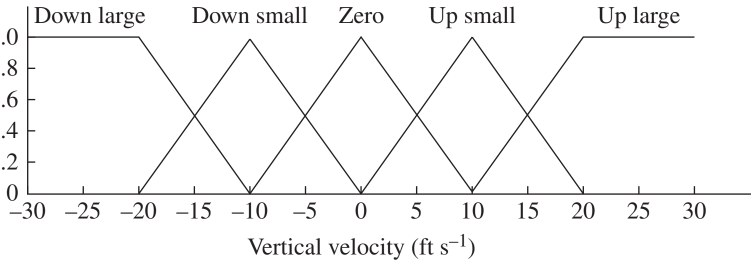

Figure 11.10 Velocity, v, partitioned.

Table 11.8 Membership values for control force.

	Output force (lb)												
	–30	–25	–20	–15	–10	–5	0	5	10	15	20	25	30
Uplarge(UL)	0	0	0	0	0	0	0	0	0	0.5	1	1	1
Upsmall(US)	0	0	0	0	0	0	0	0.5	1	0.5	0	0	0
Zero(Z)	0	0	0	0	0	0.5	1	0.5	0	0	0	0	0
Downsmall(DS)	0	0	0	0.5	1	0.5	0	0	0	0	0	0	0
Downlarge(DL)	1	1	1	0.5	0	0	0	0	0	0	0	0	0

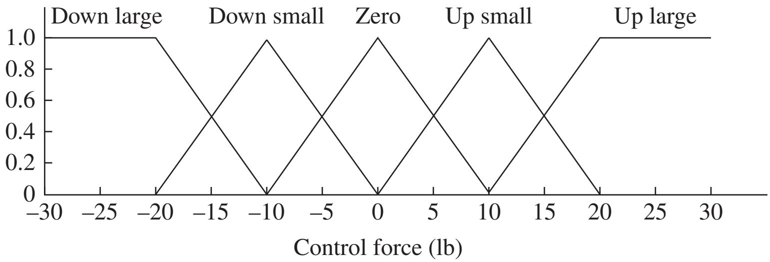

Figure 11.11 Control force, *f*, partitioned.

Table 11.9 FAM table.

	Velocity				
Height	DL	DS	Zero	US	UL
L	Z	DS	DL	DL	DL
M	US	Z	DS	DL	DL
S	UL	US	Z	DS	DL
NZ	UL	UL	Z	DS	DS

start with the aircraft at an altitude of 1000 feet, with a downward velocity of -20 ft s^{-1}. We will use the following equations to update the state variables for each cycle:

$$v_{i+1} = v_i + f_i,$$
$$h_{i+1} = h_i + v_i.$$

Initial height, h_0 : 1000 ft

Initial velocity, v_0 : $-20\,\mathrm{ft\,s^{-1}}$

Control f_0 : to be computed

Height h fires L at 1.0 and M at 0.6

Velocity v fires only DL at 1.0

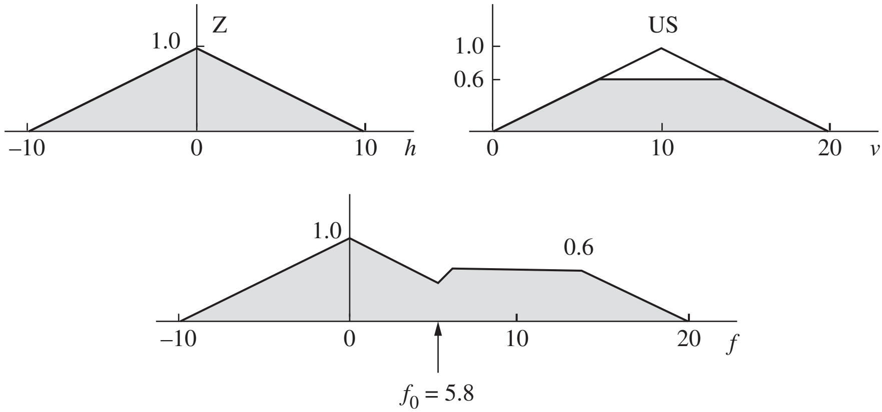

Figure 11.12 Truncated consequents and union of fuzzy consequent for cycle 1.

Height		Velocity		Output
L (1.0)	AND	DL (1.0)	⇒	Z (1.0)
M (0.6)	AND	DL (1.0)	⇒	US (0.6)

We defuzzify using the centroid method and get $f_0 = 5.8$ lb. This is the output force computed from the initial conditions. The results for cycle 1 appear in Figure 11.12.

Now, we compute new values of the state variables and the output for the next cycle:

$$h_1 = h_0 + v_0 = 1000 + (-20) = 980\,\text{ft},$$
$$v_1 = v_0 + f_0 = -20 + 5.8 = -14.2\,\text{ft}\,\text{s}^{-1}.$$

Height $h_1 = 980$ ft fires L at 0.96 and M at 0.64

Velocity $v_1 = -14.2\,\text{ft}\,\text{s}^{-1}$ fires DS at 0.58 and DL at 0.42

Height		Velocity		Output
L (0.96)	AND	DS (0.58)	⇒	DS (0.58)
L (0.96)	AND	DL (0.42)	⇒	Z (0.42)
M (0.64)	AND	DS (0.58)	⇒	Z (0.58)
M (0.64)	AND	DL (0.42)	⇒	US (0.42)

We find the centroid to be $f_1 = -0.5$ lb. Results are shown in Figure 11.13.

We compute new values of the state variables and the output for the next cycle.

$$h_2 = h_1 + v_1 = 980 + (-14.2) = 965.8\,\text{ft},$$
$$v_2 = v_1 + f_1 = -14.2 + (-0.5) = -14.7\,\text{ft}\,\text{s}^{-1}.$$

$h_2 = 965.8$ ft fires L at 0.93 and M at 0.67,

$v_2 = -14.7\,\text{ft}\,\text{s}^{-1}$ fires DL at 0.43 and DS at 0.57.

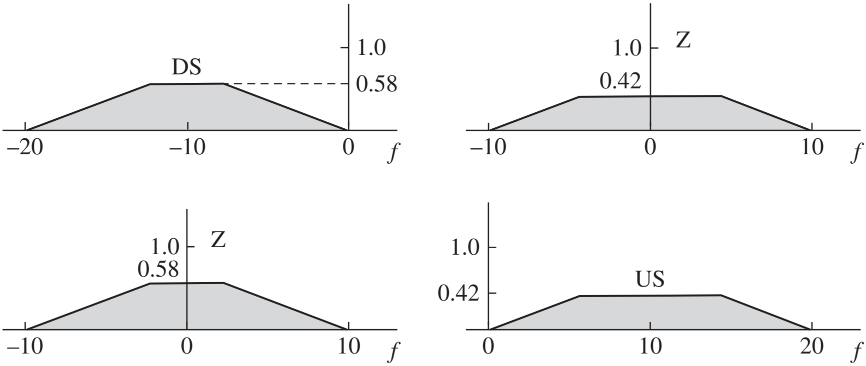

Figure 11.13 Truncated consequents for cycle 2.

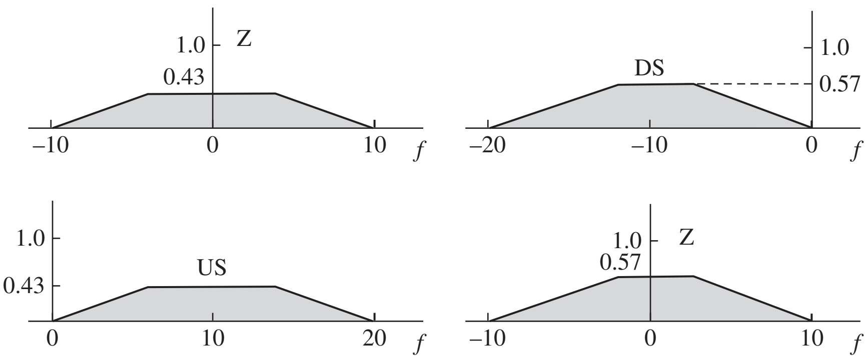

Figure 11.14 Truncated consequents for cycle 3.

Height		Velocity		Output
L (0.93)	AND	DL (0.43)	$\Rightarrow$	Z (0.43)
L (0.93)	AND	DS (0.57)	$\Rightarrow$	DS (0.57)
M (0.67)	AND	DL (0.43)	$\Rightarrow$	US (0.43)
M (0.67)	AND	DS (0.57)	$\Rightarrow$	Z (0.57)

We find the centroid for this cycle to be $f_2 = -0.4$ lb. Results appear in Figure 11.14. Again, we compute new values of state variables and output:

$$h_3 = h_2 + v_2 = 965.8 + (-14.7) = 951.1 \text{ ft},$$
$$v_3 = v_2 + f_2 = -14.7 + (-0.4) = -15.1 \text{ ft s}^{-1},$$

and for one more cycle we get

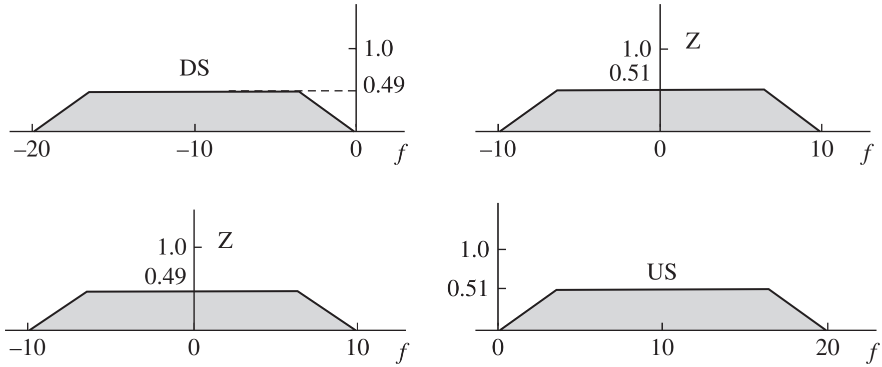

Figure 11.15 Truncated consequents for cycle 4.

$h_3 = 951.1\,\text{ft fires L at } 0.9 \text{ and M at } 0.7,$

$v_3 = -15.1\,\text{ft}\,\text{s}^{-1} \text{ fires DS at } 0.49 \text{ and DL at } 0.51.$

Height		Velocity		Output
L (0.9)	AND	DS (0.49)	⇒	DS (0.49)
L (0.9)	AND	DL (0.51)	⇒	Z (0.51)
M (0.7)	AND	DS (0.49)	⇒	Z (0.49)
M (0.7)	AND	DL (0.51)	⇒	US (0.51)

The results are shown in Figure 11.15, with a defuzzified centroid value of $f_3 = 0.3$ lb.

Now, we compute the final values for the state variables to finish the simulation:

$$h_4 = h_3 + v_3 = 951.1 + (-15.1) = 936.0\,\text{ft},$$
$$v_4 = v_3 + f_3 = -15.1 + 0.3 = -14.8\,\text{ft}\,\text{S}^{-1}.$$

The summary of the four-cycle simulation results is presented in Table 11.10. If we look at the downward velocity versus altitude (height) in Table 11.10, we get a descent profile that appears to be a reasonable start at the desired parabolic curve shown in Figure 11.6 at the beginning of the example.

Fuzzy Engineering Process Control

Engineering process control, or the automatic control of physical processes, is a rather large complex field. We discuss first some simple concepts from classical process control to provide a background for fuzzy process control concepts. Because fuzzy process control systems can be complex and diverse, we present only enough information here to provide an introduction to this interesting topic. We first discuss the classical PID controller (Equation (11.7)) and then some fuzzy logic controllers. Of the two types of control problems, set-point tracking and

Table 11.10 Summary of four-cycle simulation results.

	Cycle 0	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Height (ft)	1000.0	980.0	965.8	951.1	936.0
Velocity (ft s^{-1})	−20	−14.2	−14.7	−15.1	−14.8
Control force	5.8	−0.5	−0.4	0.3	

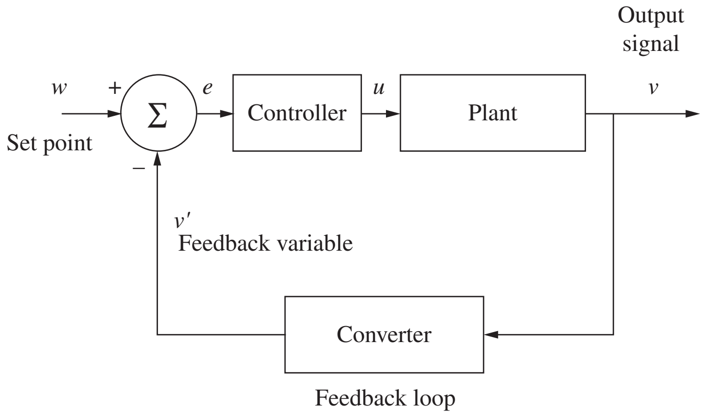

Figure 11.16 Standard block flow diagram for a control system.

disturbance rejection, we will illustrate only the set-point-tracking problem. Most industrial problems are SISO, or at least treated that way because multi-input, multi-output (MIMO) problems are normally significantly more difficult. Fuzzy MIMO problems will be discussed in this chapter because fuzzy controllers usually handle these problems quite well; of the many types, only feedback control systems will be illustrated (Parkinson, 2001).

Classical Feedback Control

The classical feedback control system can be described using a block flow diagram like the one shown in Figure 11.16.

The first rectangular block in the figure represents the controller. The second rectangular block represents the system to be controlled, often called the *plant*. The block in the feedback loop is a converter. The *converter* converts the feedback signal to a signal usable by the *summer*, which is the circle at the far left-hand side of the diagram. The letter w represents the set-point value or the desired control point. This is the desired value of the variable that we are controlling. The letter v represents the output signal or the current value of the variable that we are controlling. The symbol v' represents the feedback variable, essentially the same signal as the output signal but converted to a form that is compatible with the set-point value. The letter e represents the error, or the difference between the set-point value and the feedback variable value. The letter u represents the control action supplied by the *controller* to the *plant*. A short example will clarify this explanation.

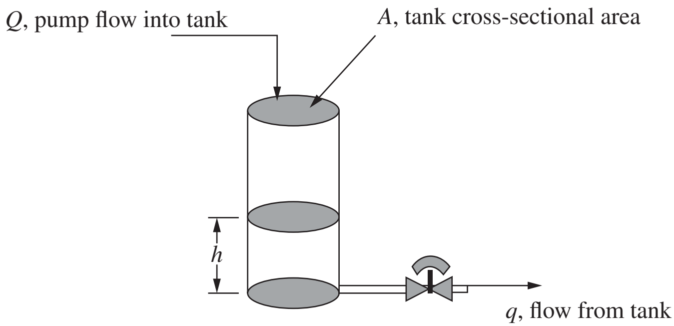

Figure 11.17 Tank with a liquid level that needs to be controlled.

Example 11.3
Liquid-Level Control Consider the tank with liquid in it shown in Figure 11.17. We want to design a controller that will either maintain the liquid level at a desired point, a disturbance rejection problem, or one that can be used to move the level set point from, say, 4 feet to 6 feet, the set-point-tracking problem, or both. We can do either one or both, but for purposes of illustration, it is easier to confine our explanation to the set-point-tracking problem. Suppose that the tank in Figure 11.17 is 10 feet tall and the tank is empty. We want to fill the tank to a level of 5 feet, so we make the current set point, w, equal to 5. The idea is to fill the tank to the desired set point as quickly and smoothly as possible. We want to minimize the amount of overshoot, or the time that the tank has a level greater than the set point value before it finally settles down. The current level at any time, t, is designated as h. Liquid flows out of the tank through an open valve. This flow is designated by the letter q. Liquid flows into the tank by means of a pump. The pump flow, Q, can be regulated by the controller. The tank cross-sectional area is designated by the letter A. Equation (11.9) describes the mass balance for the liquid in the tank as a function of time:

$$A\frac{\mathrm{d}h}{\mathrm{d}t} = Q - q. \tag{11.9}$$

Flow out of the tank, q, through the outlet pipe and the valve is described as

$$q = \Phi A_\mathrm{p}\sqrt{2gh}, \tag{11.10}$$

where Φ is a friction coefficient for flow through both the small exit pipe and the valve. It can be calculated with fair precision, or better, measured. The term A_p represents the cross-sectional area of the small exit pipe. The gravitational constant, g, is equal to 32.2 ft s^{-2} in U.S. engineering units. Figure 11.18 shows the block flow diagram for this example.

Classical PID Control

The PID control algorithm is described as

$$u = K\mathrm{p}e + K_\mathrm{I}\int_0^T e\,\mathrm{d}t + K_\mathrm{D}\frac{\mathrm{d}e}{\mathrm{d}t}, \tag{11.11}$$

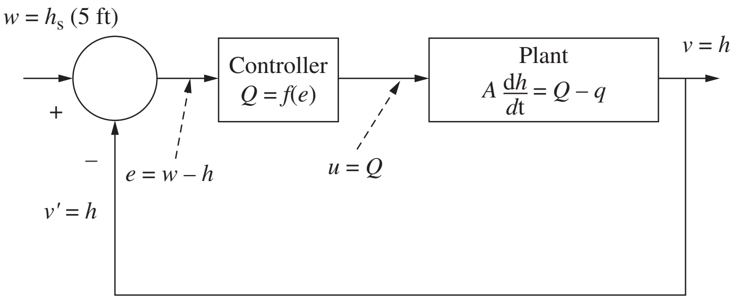

Figure 11.18 Block flow diagram for Example 11.3.

where the symbols K_P, K_I, and K_D are the proportional, integral, and derivative control constants, respectively (as in Equation (11.7)). These constants are specific to the system in question. They are usually picked to optimize the controller performance and ensure that the system remains stable for all possible control actions.

If we use a PID controller, which is linear, or any other linear controller with a linear plant, then the system is called a *linear system*. Linear systems have nice properties. The control engineer can use Laplace transforms to convert the linear equations in the blocks in Figure 11.18 to the Laplace domain. The blocks can then be combined to form a single transfer function for the entire system. Most of the systems studied in the control systems literature are linear systems. In the real world, many systems are at least slightly nonlinear. However, often this fact is ignored or the system is linearized so that linear control systems theory can be used to solve the problem. There are several techniques for linearizing control systems. The most common is to expand the nonlinear function in a truncated Taylor's series. Equation (11.10) shows that our plant in Example 11.3 is not linear. The truncated Taylor's series for linearizing Equation (11.10) about a steady state value, in this case our set point, is given as

$$\sqrt{h} \approx \sqrt{h_s} + \frac{1}{2\sqrt{h_s}}(h - h_s) \tag{11.12}$$

If we choose $h_s = 5$ ft, we can linearize the radical term over some of the control range. Table 11.11 demonstrates how well this works.

The reader can see that Taylor's approximation is not a bad one, at least until one gets near zero. The point of all this is that once the equation for the plant becomes linear, we can take Laplace transforms of the plant equation. We can also take Laplace transforms of the control equation, Equation (11.11), which is already linear. The control engineer typically redefines the variables e and h in Equations (11.9), (11.10), and (11.12) as deviation variables. That is, variables that deviate about some steady state. This causes the constants and boundary conditions to go to zero and the equations become much easier to deal with. The Laplace transforms for the controller and the plant are combined with one for the feedback loop to form an algebraic transfer function, $G(s)$ (Figure 11.19), in the Laplace domain. This transfer function form has some nice properties from a control system point of view.

Table 11.11 Approximate linearization.

h (ft)	$\sqrt{h}$	Equation (11.12)
10	3.162	3.354
9	3.0	3.130
8	2.828	2.907
7	2.646	2.683
6	2.449	2.460
5	2.236	2.236
4	2.0	2.012
3	1.732	1.789
2	1.414	1.565
1	1.0	1.342
0	0	1.118

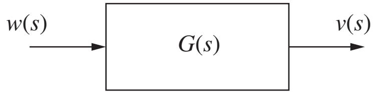

Figure 11.19 Overall system block flow diagram or transfer function for the Laplace domain.

In Figure 11.19 $G(s)$ is the transfer function for the entire block flow diagram including the feedback loop. The variables $w(s)$ and $v(s)$ are the set point and output variables, respectively, converted to the Laplace, s, domain. The control engineer can work with the system transfer function and determine the range in which the control constants, K_P, K_I, and K_D, must fall to keep the system stable. Electrical engineers design controllers for a wide variety of systems. Many of these systems can become unstable as a result of a sudden change in the control action. An example might be an aircraft control system. Chemical engineers, on the other hand, usually design control systems only for chemical processes. Many of these systems are not as likely to become unstable from a sudden change in control action. Our liquid-level controller is an example. A sudden change in the control action, say a response to a leak in the tank, is not likely to make the system become unstable, no matter how abrupt the change. Laplace transforms and the s domain are an important part of classical control theory, but they are not needed to illustrate our example.

The PID controller accounts for the error, the integral of the error, and the derivative of the error to provide an adequate response to the error. Figure 11.20 shows a typical time-domain response curve for a PID controller, for a problem like the tank-filling problem of Example 11.3.

Figure 11.20 represents a typical response curve for a set-point-tracking problem. At time t^*, the tank level $h(t^*)$ is at the point on the response curve that is pointed to by the de/dt arrow. The error, e, is the distance between the set-point level line, h_s, and the current tank level $h(t^*)$. Since in this case

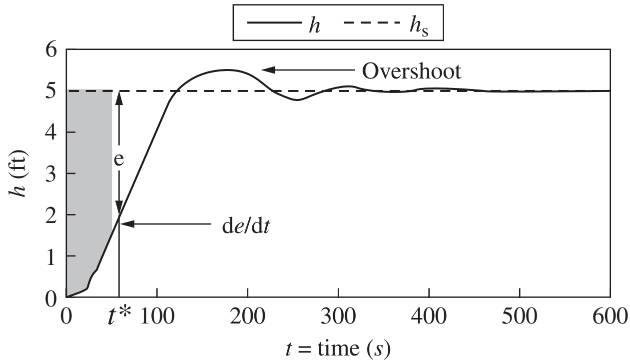

Figure 11.20 Time-level PID response for the tank-filling problem in Example 11.3.

$$de/dt = d(h_s - h(t))/dt = -dh/dt,$$

the derivative term in the PID control equation is shown as the negative derivative of the response curve at time t^*. The shaded area in Figure 11.20 between the h_s line and the h curve between time t_0 and t^* is an approximation to the integral term in Equation (11.11). One criterion for an optimal controller is to find control constants than minimize the error integral. That is, find K_P, K_I, and K_D such that $\int_{t_0}^{t_\infty} e\,dt$ is minimized. The term t_∞ is the time at which the controlled variable, tank level in this case, actually reaches and stays at the set point. The proportional term in Equation (11.11) drives the control action hard when there is a large error and slower when the error is small. The derivative term helps to home in the controller variable to the set point. It also reduces overshoot because of its response to the change in the sign and the rate of change of the error. Without the integral term, however, the controlled variable would never hold at the set point because at the set point both e and de/dt are equal to 0. A pure PD controller would become an on–off controller when operating about the set point. The overshoot shown in Figure 11.20 is something that control engineers normally try to minimize. It can really present a problem for a tank-filling exercise, especially if the level set point is near the top of an open tank.

Fuzzy Control

In the simplest form, a fuzzy control system connects input membership functions, functions representing the input to the controller, e, to output membership functions that represent the control action, u. A good example for the fuzzy control system is a controller that controls the liquid level in the tank shown in Figure 11.17. This time we want to design a controller that will allow us to change the set point either up or down, and one that will correct itself in the case of overshoot. A simple fuzzy control system designed for our tank-level set-point-tracking problem consists of three rules.

1. If the ***Level Error*** is ***Positive*** Then the ***Change in Control Action*** is ***Positive***.
2. If the ***Level Error*** is ***Zero*** Then the ***Change in Control Action*** is ***Zero***.
3. If the ***Level Error*** is ***Negative*** Then the ***Change in Control Action*** is ***Negative***.

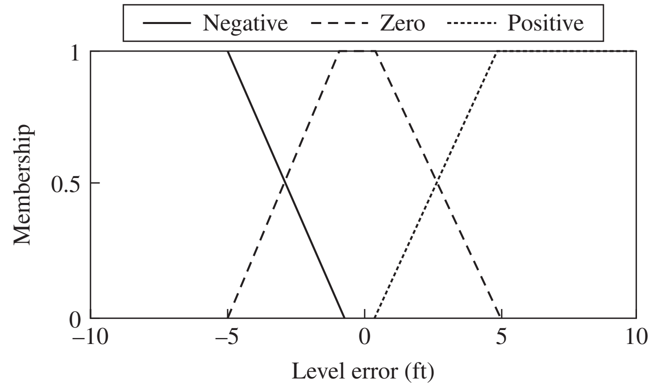

Figure 11.21 Input membership functions for the fuzzy tank-level controller.

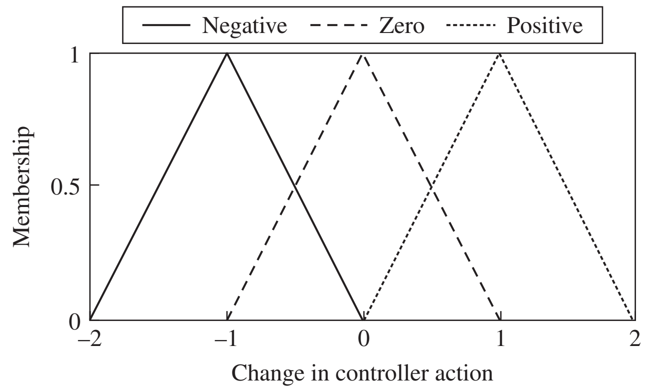

Figure 11.22 Output membership functions for the fuzzy liquid-level controller.

The input membership functions are shown in Figure 11.21. The reader should also notice the "dead band" or "dead zone" in the membership function ***Zero*** between ±3 inches. This is optional and is a feature commonly used with on–off controllers. It is easy to implement with a fuzzy controller and is useful if the control engineer wishes to minimize control response to small transient-level changes. This step can save wear and tear on equipment.

The output membership functions for this controller are shown in Figure 11.22. In this figure, the defuzzified output value from the controller is a fractional value representing the required pump output for the desired level change. It is defined by the following expression:

$$\textbf{\textit{Change in Controller Action}} = \Delta u = \left(Q_i - Q_{\mathrm{sq}}\right)/Range, \tag{11.13}$$

where the term Q_{sp} represents the pump output (gallons per minute) required to maintain the set-point level. The term Q_i is the new pump output requested by the controller. If $\Delta u > 0$, then the *Range* is defined as $Q_{\max} - Q_{\mathrm{sp}}$, where $Q_{\max}$ is the maximum pump output. If $\Delta u < 0$, then the *Range* is defined as Q_{sp}. The term Q_{sp} must be calculated using a steady state mass balance for the tank or it must be estimated in some fashion. The steady state calculation requires only algebra. It requires only a knowledge of the parameter Φ in Equation (11.10). This value can either be measured by experiment, or approximated quite closely from resistance coefficients found in any fluid mechanics text (e.g., Olsen, 1961). This is an engineering calculation that is quite different from, and usually easier to do than, calculations needed to compute K_{P}, K_{I}, and K_{D} for the PID controller.

The ranges of the fuzzy output sets ***Positive*** and ***Negative*** are +2.0 to 0.0 and −2.0 to 0.0, respectively. Because the ***Change in Controller Action*** is a fraction between either 0.0 and 1.0 or 0.0 and −1.0, it is clear that we will never obtain a control action outside the range of −1.0 to 1.0. Our defuzzification technique will require that we include numbers up to 2.0 in the fuzzy set or membership function ***Positive*** and numbers down to −2.0 in the fuzzy set ***Negative***. Even though numbers of this magnitude can never be generated by our fuzzy system, we can still include them in our fuzzy sets. The users can define their fuzzy sets however they wish. The fuzzy mathematics described in previous chapters is capable of handling objects of this type. The user has to define the fuzzy sets so that they make sense for the particular problem. In our case we are going to use the centroid technique for defuzzification. We therefore need to extend our membership functions so that it is possible to obtain centroids of ±1.0. We need this capability for the control system to either turn the pump on "full blast" or turn it completely off.

We can describe our simple fuzzy controller as an approximation to an Integral (I) controller. Our rules are of the form $\Delta u = f(e)$, where Δu is the ***Control Action Change*** for the sample time interval Δt. We can make the approximation that $\Delta u/\Delta t \approx \mathrm{d}u/\mathrm{d}t \approx K_{\mathrm{I}}e$ and $\int \mathrm{d}u = K_{\mathrm{I}} \int e\mathrm{d}t$ and that $u = u_0 + K_{\mathrm{I}} \int_{t_0}^{T} e\ e\ \mathrm{d}t$ or that

$$u = K_{\mathrm{I}} \int_0^T e\mathrm{d}t.$$

Example 11.4 (continuation of Example 11.3).
Suppose that we decide to change our set-point level from 5 feet to 8 feet in the tank described in Example 11.3. The error is defined as the set-point level, 8 feet, minus the current level, 5 feet, or +3 feet. The three rules are fired, producing the following results:

1. ***Positive*** error is 0.5.
2. ***Zero*** error is 0.5.
3. ***Negative*** error is 0.0.

The results are shown graphically in Figure 11.23.

In this example an error of +3 feet intersects the membership function ***Zero*** at approximately 0.5 and the membership function ***Positive*** at approximately 0.5. We say that Rules 1 and 2 were each fired with strength 0.5. The output membership functions corresponding to Rules 1 and 2 are each "clipped" at 0.5. See Figure 11.24.

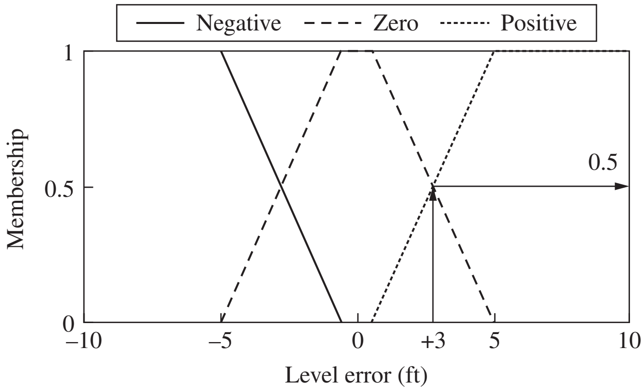

Figure 11.23 Resolution of input for Example 11.4.

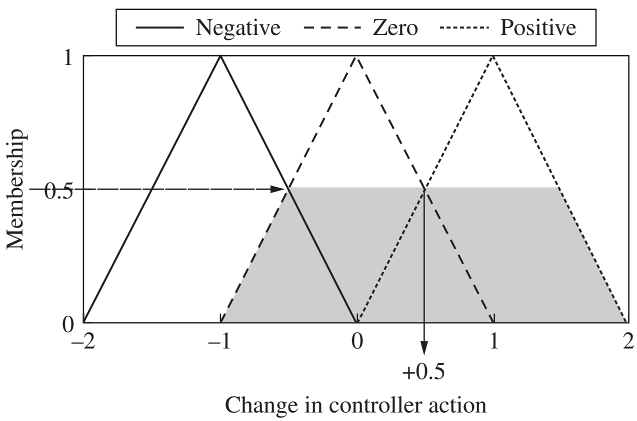

Figure 11.24 Resolution of output for Example 11.4.

The centroid of the "clipped" membership functions, the shaded area in Figure 11.24, is +0.5. This centroid becomes the term Δu in Equation (11.13). Because Δu is greater than 0.0, Equation (11.13) can be rewritten as

$$Q_i = \left(Q_{\max} - Q_{\mathrm{sp}}\right)\Delta u + Q_{\mathrm{sp}} \text{ or } Q_i = 0.5\left(Q_{\max} - Q_{\mathrm{sp}}\right),\ \text{since}\ \Delta u = 0.5. \tag{11.14}$$

This says that the new pump output, Q_i, should be adjusted to be halfway between the current or set-point output and the maximum heater output. After an appropriate time interval, corresponding to a predetermined sample rate, the same procedure will be repeated until the set-point level, 8 feet, is achieved. The set-point-tracking response curve for this problem will probably look something like the one shown in Figure 11.20. Hopefully, the overshoot will be reduced by the addition of the dead band in the input membership functions and a judicious choice of the sample interval time, Δt. In level control problems like this one, dead bands can be useful because the physical action of the liquid pouring from the pump outlet onto the liquid surface in the tank will cause the fluid in the tank to "slosh" around. A good sensor will pick up these level changes and overwork the controller. For the same reason, classical PI controllers are often used for level control problems like this, instead of PID controllers because the fluid movement keeps the derivative portion active.

Unfortunately, this simple fuzzy control system will not handle disturbance rejection problems very well. This is because of the method that we have chosen to solve this problem. The problem is the term Q_{sp}. This term is reasonably easy to measure or calculate, but it is no longer valid if there is a hole in the tank, or a plugged valve, which are probably the most likely causes of disturbances in this system. This is reasonably easy to fix with more rules, but the explanation is quite lengthy. The interested reader is referred to Parkinson (2001) or Ross, Booker, and Parkinson (2002). The PID controller will solve both the disturbance rejection problem and the set-point-tracking problem, with one set of control constants. Often, however, PID control constants that are optimized for one type of solution are not efficient for the other type.

Multi-Input, Multi-Output (MIMO) Control Systems

The classical MIMO control system is much more complicated. The textbook approach assumes linear systems, uses a lot of linear algebra, and often the best that can come out of these models is a set of proportional-only controllers. Phillips and Harbor (1996) have a good readable chapter devoted to this approach. There are also entire textbooks and graduate-level control courses devoted to linear control systems for MIMO systems. A common industrial approach, at least in the chemical industry where systems tend to be highly nonlinear, is to use multiple PID controllers. Because of this "brute-force" approach, these controllers tend to interact and "fight" one another. There are methods for decoupling multiple controllers, but it is a great deal of work. A good discussion for learning how to decouple multiple controllers is given in Ogunnaike and Ray (1994). One of the advantages of fuzzy controllers is that it is reasonably easy to write good MIMO control systems for highly nonlinear MIMO problems. The next example illustrates this.

Example 11.5
The Three-Tank MIMO Problem The three-tank system described here was a real experiment (Parkinson, 2001). It is an extension of the single-tank system discussed in Examples 11.3 and 11.4. The tanks are smaller, however. The experimental apparatus consisted of three Lucite tanks, in series, each holding slightly less than 0.01 m^3 of liquid; the system is shown in Figure 11.25. The tanks are numbered from left to right in the figure as tank 1, tank 3, and tank 2. All three tanks are connected, with the third tank in the series, tank 2, draining to the system exit. Liquid is pumped into the first and the third tanks to maintain their levels. The levels in the first and third tanks control the level in the middle tank. The level in the middle tank affects the levels in the two end tanks.

The differential equations that describe this experimental system are Equations (11.15) to (11.17). The tank flows are described in Equations (11.18) to (11.20). The symbols used in these equations have the same meaning as those used in Examples 11.3 and 11.4, except that they now have subscripts as described in Figure 11.25.

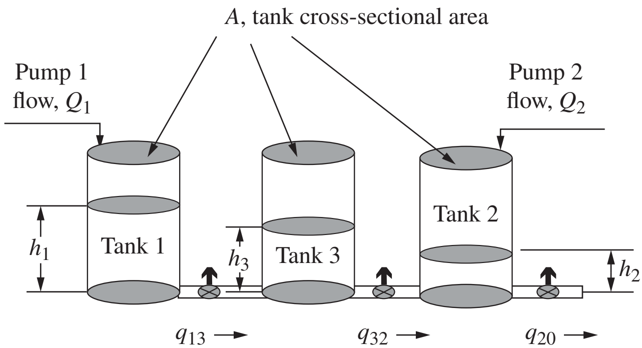

Figure 11.25 The experimental three-tank system.

$$A\frac{\mathrm{d}h_1}{\mathrm{d}t} = Q_1 - q_{13}, \tag{11.15}$$

$$A\frac{\mathrm{d}h_2}{\mathrm{d}t} = Q_2 - q_{32} - q_{20}, \tag{11.16}$$

$$A\frac{\mathrm{d}h_3}{\mathrm{d}t} = q_{13} - q_{32}, \tag{11.17}$$

$$q_{13} = \Phi_1 A_\mathrm{p}\,\mathrm{Sign}(h_1 - h_3)\sqrt{2g|h_1 - h_3|}, \tag{11.18}$$

$$q_{32} = \Phi_3 A_\mathrm{p}\,\mathrm{Sign}(h_3 - h_2)\sqrt{2g|h_3 - h_2|}, \tag{11.19}$$

$$q_{20} = \Phi_2 A_\mathrm{p}\sqrt{2gh_2}. \tag{11.20}$$

Again, we will describe only the set-point-tracking portion of the controller and the results for the set-point-tracking experiments. For the interested reader, the description of the disturbance rejection portion of the controller and the results of the disturbance rejection experiments are given in Parkinson (2001).

The fuzzy rules for this set-point-tracking module are given in Table 11.12. The rules are of the form

$$\text{If Error}(i) \text{ is} \ldots \text{ then Flow\_Change}(i) \text{ is} \ldots$$

The term Error(i) in Table 11.12 is defined as follows for both tanks 1 and 2:

$$\mathrm{Error}(i) = \frac{w_i - h_i}{Rangeh(i)}, \quad \text{for } i = 1 \text{ or } 2. \tag{11.21}$$

If $w_i > h_i$ then $Rangeh(i)$ equals w_i.

Else $Rangeh(i)$ equals $h_{i\ \max} - w_i$, for $i = 1$ or 2.

The variable h_i is the current level for tank i and w_i is the set point for tank i. The term $h_{i\ \max}$ is the maximum level for tank i, or the level when the tank is full. The Flow_Change(i) variable is defined as

Table 11.12 Fuzzy rules for set-point tracking.

Rule number	Error(i)	Flow_Change(i)
1	Error(1) = Negative	Flow_Change(1) = Negative
2	Error(1) = Zero	Flow_Change(1) = Zero
3	Error(1) = Positive	Flow_Change(1) = Positive
4	Error(2) = Negative	Flow_Change(2) = Negative
5	Error(2) = Zero	Flow_Change(2) = Zero
6	Error(2) = Positive	Flow_Change(2) = Positive

$$\text{Flow\_Change}(i)\,\frac{q_i - q_{iss}}{Rangeq(i)}. \tag{11.22}$$

If Flow_Change(i) > 0 then $Rangeq(i)$ equals q_{iss}.

Otherwise, $Rangeq(i)$ equals $q_{i\ \max} - q_{iss}$, for $i = 1$ or 2.

The term $q_{i\ \max}$ is the maximum possible flow from pump i, 6.0 L min^{-1} for each pump. The variable q_i is the current pump flow, from pump i, and q_{iss} is the steady state, or set point, flow for pump i, computed from a mass balance calculation. The mass balance calculation is a simple algebraic calculation. It is based on Equations (11.15) to (11.20) with derivatives set to zero and the tank levels set at the set points. The calculation simplifies to three equations with three unknowns. The three unknowns obtained from the solution are the two steady state pump flows and one independent steady state tank level. One difficulty is determining the flow coefficients, Φ_i. These coefficients can be easily measured or calculated using textbook values; measurement of these values would be best. However, if the coefficients are calculated and the calculation is not correct, the disturbance rejection mode will be invoked automatically and the values will be determined by the control system.

The membership function universes for Error(i) and for Flow_Change(i) have been normalized from −1 to 1. This was done to construct generalized functions. In this type of control problem, we want one set of rules to apply to all tank-level set points. Generalizing these functions makes it possible to use only six rules to handle all tank levels and all tank-level changes. Figure 11.26 shows the input membership functions for the fuzzy controller used to solve this three-tank problem. Figure 11.27 shows the output membership functions for the fuzzy control system. These membership function ranges have been expanded to the limits of −2 and 2 so that the pumps can be turned on full blast if needed. These inputs and outputs are connected by the rules shown in Table 11.12.

Example 11.6 Three-Tank Problem (continued)

For this example, we select the set points, w_1 and w_2, for tanks 1 and 2 to be 0.40 and 0.20 m, respectively. These are the same set points that were used in the actual experiment (Parkinson, 2001). We start with three empty tanks. The steady state flows required to maintain these levels are computed from the mass balance calculation to be $q_{1ss} = 1.99346$ L min^{-1} and $q_{2ss} =$ 2.17567 L min^{-1}. The maximum pump flows are $q_{1\ \max} = q_{2\ \max} = 6.0$ L min^{-1}. For the example, we assume that the tank levels have nearly reached their set-point values. The level in tank 1, h_1, is 0.38 m and the level in tank 2, h_2, is 0.19 m. The maximum tank levels are $h_{1\ \max} = h_{2\ \max} = 0.62$ m.

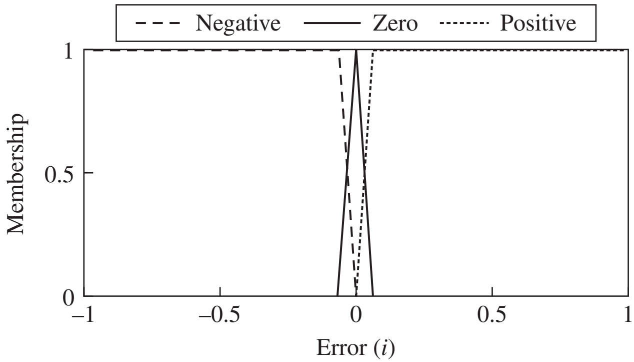

Figure 11.26 Membership functions for input Error(i) for $i =$ both 1 and 2 (pumps 1 and 2).

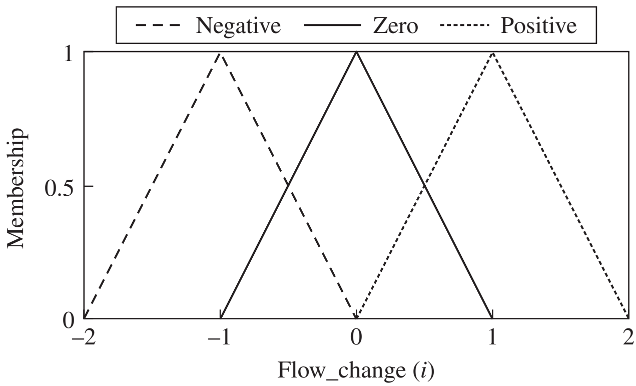

Figure 11.27 Membership functions for Flow_Change(i) for $i =$ both 1 and 2 (pumps 1 and 2).

Error(1) and Error(2) are both calculated to be –0.05 from Equation (11.21). In both cases the memberships from Figure 11.26 are computed to be ***Negative*** = 0.775 and ***Zero*** = 0.225. These values cause Rules 1 and 4 to be fired with a weight of 0.775, and Rules 2 and 5 to be fired with a weight of 0.225. The output membership functions shown in Figure 11.27 are truncated at these values. The defuzzification method used with the set-point-tracking control for this controller is the correlation minimum encoding (CME) technique. This method computes the areas and centroids of the entire truncated triangles. That is, the negative and positive Flow_Change membership functions are extended to –2 and +2, respectively, to compute the centroids and the areas. This is one of many techniques used with fuzzy output membership functions when it is important to use the centroid of the membership function to designate complete "shut off" or "go full blast." The crisp, or defuzzified, value used in the control equations, and obtained from firing these rules, is defined in Equation (11.22). Both Flow_Change(1) and Flow_Change(2) are computed to be –0.704. For both pumps, Flow_Change(i) is less than 0, so *Rangeq*(i) is

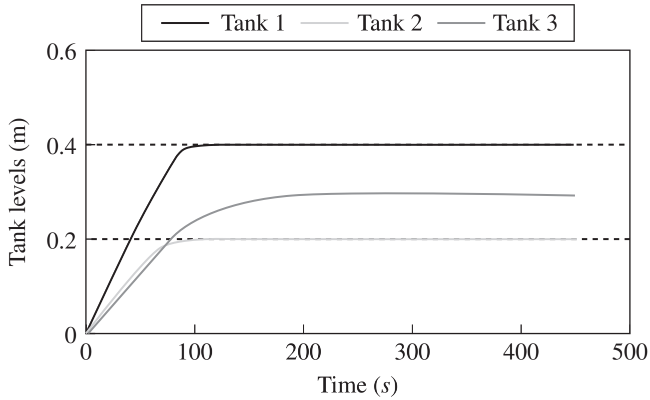

Figure 11.28 Tank levels versus time for the fuzzy controller.

equal to $q_{i\,\max} - q_{iss}$. $Rangeq(1)$ is then 4.00654 and $Rangeq(2)$ is equal to 3.82433. By manipulating Equation (11.22), we obtain the following relationship:

$$q_i = q_{iss} - \text{Flow\_Change}(i) * Rangeq(i). \tag{11.23}$$

From Equation (11.23), the value of q_1 is computed as

$$q_1 = 1.99346 + (0.704)(4.00654) = 4.8141$$

and q_2 is

$$q_2 = 2.17567 + (0.704)(3.82433) = 4.8680.$$

This set-point-tracking test case was a tank-filling problem. The set-point level was changed from 0 m, an empty condition, to 0.4 m for tank 1, and 0.2 m for tank 2. The level for tank 3 cannot be set independently. For this test, tank 3 found its own level at about 0.3 m. The run time was set to 7.5 min, or 450 s. The results for the fuzzy controller are shown in Figure 11.28.

Attention is called to the fact that there is no overshoot on any of the three tanks, even without a dead band in the input membership functions. Also, the filling time is rapid. In the actual experiment, Parkinson (2001) details several tests for both set-point tracking and disturbance rejection, and the fuzzy controller performed better than the classical controllers every time.

Fuzzy Statistical Process Control

There are two basic types of statistical process control (SPC) problems. One of the problem types deals with measurement data. An example would be the measurement of the diameter of a cylindrical part that is produced by a machining operation. The typical SPC method for dealing with measurement data is to use $\overline{X} - R$ charts (Shewhart, 1986).

These charts work well with one input variable and are discussed in the next section. The second problem type deals with attribute data. In this case, instead of dealing with the actual measurement information, the process control person assigns an attribute like "pass" or "fail" to the item. A common SPC technique is to use a p chart based on binary inputs and using the binomial distribution (Shewhart, 1986). This technique is discussed in the section on attribute data. These traditional techniques work very well as long as only a single input is required for the measurement data problems and as long as only binary input is required for the attribute data problems (Parkinson and Ross, 2002).

Fuzzy SPC is useful when multiple inputs are required for each of these two SPC problem types. Two separate studies were conducted to determine the usefulness of the fuzzy SPC technique (Parkinson and Ross, 2002). These case studies, illustrated here as Examples 11.7 and 11.8, were directed at beryllium part manufacture. Although the numbers presented here are not the actual values used in the studies, they are useful in illustrating the fuzzy SPC technique. The fuzzy technique using measurement data was for beryllium exposure, but would apply equally well to quality control. The other study used attribute data for quality control of the manufacturing of beryllium parts. The beryllium manufacturing process is quite interesting because the manufacturing process is atypical; it is atypical because the process almost always involves small lots, and often a different part is processed each time. In both Examples 11.7 and 11.8, a computer model of the beryllium plant was used. The purpose of the studies was to compare fuzzy SPC techniques with traditional SPC techniques for these atypical cases. The exposure control problem in the beryllium plant is atypical because it must consider several variables. The traditional SPC procedure with multiple-input data is to apply a least squares technique to regress the multiple data to one input and then use the $\overline{X}-R$ chart. The fuzzy technique, illustrated in Example 11.7, uses rules and membership functions to reduce the multiple variables to a single variable and then applies the $\overline{X}-R$ chart (Parkinson, 2001).

In the quality control study utilizing attribute data, illustrated in Example 11.8, instead of the usual binary pass–fail situation, we have multiple classifications such as firsts, seconds, recycle, and discard. The traditional SPC method of dealing with this problem is to use a generalized p chart based on the chi-square distribution (Shewhart, 1986). The fuzzy approach to this multiple-input problem is to use fuzzy rules to combine the multiple variables and then use a fuzzy chart that is somewhat similar to the standard p chart technique (Parkinson, 2001).

In both case studies, the fuzzy method proved slightly superior and much easier to use than the standard statistical techniques. The purpose of these illustrations is not to discuss the relative merits of the techniques, but to demonstrate the use of the fuzzy methods. Interested students are encouraged to study the various techniques and to make their own decisions about which technique to use in various applications.

Measurement Data: Traditional SPC

Suppose that the operation that produces the cylindrical parts mentioned previously produces a thousand parts a day. If an 8-hour day was used, this would be 125 parts an hour. Suppose, further, that the process control engineer decides to measure the diameter of five parts every 2 hours for a week. The engineer then plots the results on an $\overline{X}-R$ chart. Figure 11.29 is a simulated $\overline{X}$ chart for this situation. Figure 11.30 is the associated simulated R chart. The first point in Figure 11.29, or set number 1, is the average of the five diameter measurements taken in the first sample. In this case, the five measurements were 0.6, 0.59, 0.54, 0.57, and 0.58 inches. The average or $\overline{X}$ for set number 1 is 0.576 inches. The first point in Figure 11.30, or set number 1, is the range of the five diameter measurements taken in the first sample, $0.6 - 0.54 = 0.06$. In this example, four sets were sampled every day for 5 days for a total of 20 sets. The $\overline{X}$ average and

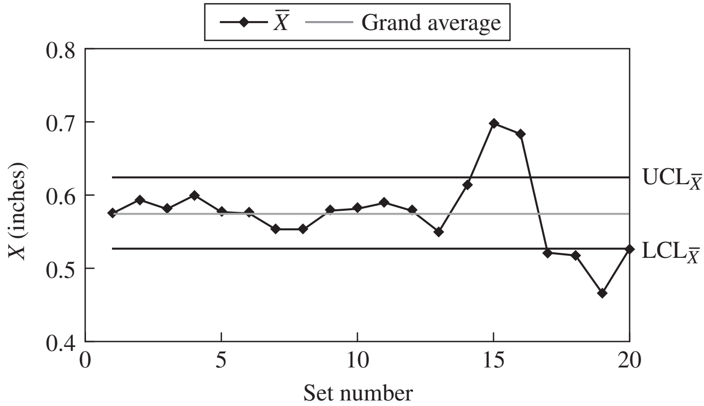

Figure 11.29 Simulated $\overline{X}$ chart for the measurement of cylinder diameters.

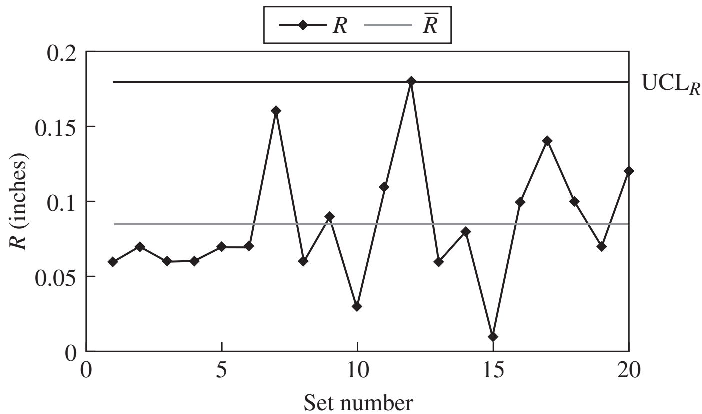

Figure 11.30 Simulated R chart for the measurement of cylinder diameters.

range, R, of every set were determined and plotted in Figure 11.29 and 11.30, respectively. The average of the 20 $\overline{X}$, or the grand average ($\overline{\overline{X}}$), is also plotted in Figure 11.29. The average of the 20 ranges, $\overline{R}$, is plotted in Figure 11.30. The upper and lower control limits shown in Figures 11.29 and 11.30 are computed using Table 11.13 and Equations (11.24) to (11.27):

$$\mathrm{UCL}_{\overline{X}} = \overline{\overline{X}} + A_2\overline{R}, \tag{11.24}$$

$$\mathrm{LCL}_{\overline{X}} = \overline{\overline{X}} + A_2\overline{R}, \tag{11.25}$$

$$\mathrm{UCL}_R = D_4\overline{R}, \tag{11.26}$$

$$\mathrm{LCL}_R = D_3\overline{R}, \tag{11.27}$$

Table 11.13 Factors for determining control limits. Adapted from Wheeler and Chambers, 1992.

N	A_2	D_3	D_4
2	1.880	–	3.268
3	1.023	–	2.574
4	0.729	–	2.282
5	0.577	–	2.114
6	0.483	–	2.004
7	0.419	0.076	1.924
8	0.373	0.136	1.864
9	0.337	0.184	1.816
10	0.308	0.223	1.777
11	0.285	0.256	1.744
12	0.266	0.283	1.717
13	0.249	0.307	1.693
14	0.235	0.328	1.672
15	0.223	0.347	1.653

where $UCL_{\overline{X}}$ and $LCL_{\overline{X}}$ are the upper and lower control limits, respectively, for the $\overline{X}$ chart and UCL_R and LCL_R are the upper and lower control limits, respectively, for the R chart (Shewhart, 1986). The symbols A_2, D_3, and D_4 are listed in Table 11.13 for various sample sizes, n. In the case of our example, the sample set size, n, is 5. A_2 is 0.577. There is no D_3, or lower limit for the range for set sizes smaller than 7. D_4 is 2.004.

The upper and lower control limits are roughly three standard deviations away from the average line. This means normal or random deviations will fall between the upper and lower control limits 99 to 100% of the time. Values falling outside these lines should nearly all be "special cause events" that need to be addressed. Values falling inside the control limits are normal events because of normal random deviations in the process. The $\overline{X}$ chart, Figure 11.29, shows process deviations over a period of time, the deviation between samples. The R chart, Figure 11.30, shows deviations within the sample set. An example of a problem that might be detected with the R chart would be a bad sensor. In this case, Figure 11.29 shows a process that is going out of control as time moves on. This could be caused from a cutting tool wearing out, a situation that needs to be corrected. Figure 11.30 shows one range point "out of control." This is probably the result of one bad measurement within the set. If this problem continues, it could indicate a bad set of calipers or other measuring device.

Our beryllium problem is a little more complex than the cylindrical parts example, but a similar technique can be applied. Exposure to beryllium particulate matter, especially small particles, has long been a concern to the beryllium industry because of potential human health problems (inhaled beryllium is extremely toxic). The industrial exposure limit has been 2 $\mu g\ m^{-3}$ per worker per 8-hour shift, but recently these limits have been reset to 0.2 $\mu g\ m^{-3}$, or 10 times lower than the previous industrial standard. The beryllium manufacturing facility investigated in our study has a workload that changes from day to day. Also the type of work done each day can vary dramatically. This makes the average beryllium exposure vary widely from day to day. This in turn makes it difficult to determine a degree of control with the standard statistical control charts.

Example 11.7

Measurement Data—Fuzzy SPC The simulated plant has four workers and seven machines. Each worker wears a device that measures the amount of beryllium inhaled during his or her shift. The devices are analyzed in the laboratory and the results are reported the next day after the exposure has occurred. An $\overline{X}-R$ chart can be constructed with these data and presumably answer the questions of control and quality improvement. Any standard text on SPC will contain a thorough discussion on control limits for these charts; for example, see Wheeler and Chambers (1992) or Mamzic (1995). Although such a chart can be useful, because of the widely fluctuating daily circumstances, these tests for controllability are not meaningful.

There are four variables that have a large influence on the daily beryllium exposure. They are the number of parts machined, the size of the part, the number of machine setups performed, and the type of machine cut (rough, medium, or fine). In our fuzzy model, a semantic description of these four variables and the beryllium exposure are combined to produce a semantic description of the type of day that each worker had had. The day type is then averaged and a distribution is found. These values are then used to produce fuzzy Shewhart-type $\overline{X}$ and R charts. These charts take into account the daily variability. They provide more realistic control limits than the traditional $\overline{X}-R$ charts.

The fuzzy system consists of five input variables or universes of discourse and one output variable. Each input universe has two membership functions and the output universe has five membership functions. The inputs and the outputs are related by 32 rules. The five input variables are:

1. Number of Parts: with a range of 0–10 and membership functions (a) Few and (b) Many.
2. Size of Parts: with a range of 0–135 and membership functions (a) Small and (b) Large.
3. Number of Setups: with a range of 0–130 and membership functions (a) Few and (b) Many.
4. Type of Cut: with a range of 1–5 and membership functions (a) Fine and (b) Rough.
5. Beryllium Exposure: with a range of 0–0.4 and membership functions (a) Low and (b) High.

The output variable is

1. The Type of Day: with range from 0 to 1 and membership functions (a) Good, (b) Fair, (c) OK, (d) Bad, and (e) Terrible.

For each of the five input variables, there are two membership functions represented in each case by two equal triangles. Figure 11.31 shows the five output membership functions. The

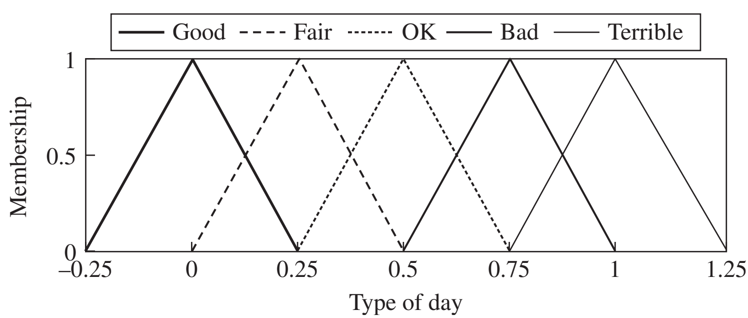

Figure 11.31 Output membership functions.

rules are based on some simple ideas. For example, if all of the four mitigating input variables indicate that the beryllium exposure should be low, and it is low, then the Type of Day is OK. Likewise, if all four indicate that the exposure should be high, and it is high, then the Type of Day is also OK. If all four indicate that the exposure should be low, and it is high, then the Type of Day is Terrible. If all four indicate that the exposure should be high, and it is low, then the Type of Day is Good. Fair and Bad days fall in between the OK days and the Good and Terrible extremes. The following is the form of the rules:

If (Number of Parts) is … and If (Size of Parts) is
… and If (Number of Setups) is … and If (Type of Cut) is
… and If (Beryllium Exposure) is … Then (The Type of Day) is …

The Size of Parts is determined as the number of parts multiplied by the average diameter of each part, measured in centimeters. The Type of Cut is determined by a somewhat complicated formula based on a roughness factor for each part, the number of parts and the size of those parts, and the number of setups required for each worker each day. A fine cut has a roughness factor (rf) of 1, a medium cut has an rf equal to 3, and a rough cut has an rf equal to 5. The calculation for Type of Cut is a bit complicated but it provides a daily number between 1 and 5 (fine to rough) for each worker, which is meaningful. An example of the use of the fuzzy technique will follow a discussion of the plant simulation.

Plant Simulation

The model has the following limitations or boundary conditions:

1. There are four machinists.
2. There are seven machines.
3. Machines 1 and 2 do only rough cuts.
4. Machines 3 and 4 do both rough cuts and medium cuts.
5. Machines 5, 6, and 7 do only fine cuts.
6. Machine 7 accepts only work from machines 3 and 4.
7. Machines 5 and 6 accept only work from machines 1 and 2.
8. Each machinist does all of the work on one order.
9. All machinists have an equally likely chance of being chosen to do an order.
10. There are 10 possible paths through the plant (at this point all are equally likely).

The simulation follows the below algorithm:

1. A random number generator determines how many orders will be processed on a given day (1–40).
2. Another random number generator picks a machinist.
3. A third random number generator picks a part size.
4. A fourth random number generator picks a path through the plant. For example, machine 1 to machine 3 to machine 7.

5. The machine and path decide the type of cut (rough, medium, or fine). Machines 1 and 2 are for only rough cuts, machines 5–7 are for only fine cuts, and machines 3 and 4 do rough cuts if they are the first machines in the path and medium cuts if they are the second machines in the path.
6. A random number generator picks the number of setups for each machine on the path.
7. Another random generator picks the beryllium exposure for the operator at each step.

This procedure is carried out for each part each day. The entire procedure is repeated the following day, until the required number of days has passed. For this study, the procedure was run for 30 days to generate some sample control charts. A description of the fuzzy control chart construction follows.

Establishing Fuzzy Membership Values

We follow each step of the process for a specific machinist for a given day; this process is then extended to the work for the entire day for all machinists. From the simulation, on day 1, 13 part orders were placed. Machinist 2 processed four of these orders, machinist 1 processed three, machinist 3 processed four, and machinist 4 processed two orders. Machinist 2 will be used to demonstrate the fuzzy system.

The cumulative size of the four parts that machinist 2 processed on day 1 was calculated to be 64.59. The number of setups that he or she performed was 45. The numeric value for the type of cuts he or she performed on that day was 1.63. Finally, the machinist's beryllium exposure was 0.181 $\mu g\, m^{-3}$ for that 8-hour period.

On inserting the input values into the simple input membership functions described, we obtain the following values. For Number of Parts = 4, the membership in Many is 0.4 and the membership in Few is 0.6. For Size of Parts = 64.59, the membership in Small is 0.52 and the membership in Large is 0.48. For Number of Setups = 45, the membership in Many is 0.35 and the membership in Few is 0.65. For Type of Cuts = 1.63, the membership in Rough is 0.16 and the membership in Fine is 0.84. The Beryllium Exposure is 0.181. The membership in High is 0.45 and the membership in Low is 0.55.

In this example, a max–min Mamdani inference was used on the rules and the centroid method was used for defuzzification. For example, Rule 1 is fired with the following weights:

- Number of Parts – Few = 0.6;
- Size of Parts – Small = 0.52;
- Number of Setups – Few = 0.65;
- Type of Cut – Fine = 0.84;
- Beryllium Exposure – Low = 0.55.

The consequent of Rule 1 is Fair and takes the minimum value, 0.52. Of the 32 rules that are all fired for this example, Fair is the consequent of 10 of them with membership values ranging from 0.16 to 0.52. The max–min rule assigns the maximum value of 0.52 to the consequent Fair. Similarly, the consequent Terrible appears five times with a maximum value of 0.45, OK appears twice with a maximum value of 0.35, Bad appears 10 times with a maximum value of 0.45, and Good appears five times with a maximum value of 0.4. The Mamdani inference process truncates the output membership functions at their maximum value (Chapter 5). In this

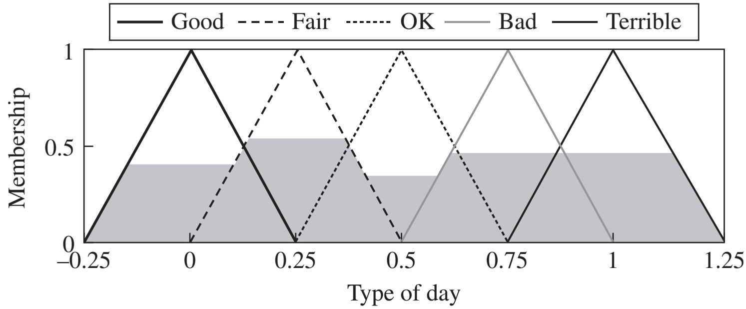

Figure 11.32 "Truncated" output membership functions for the example.

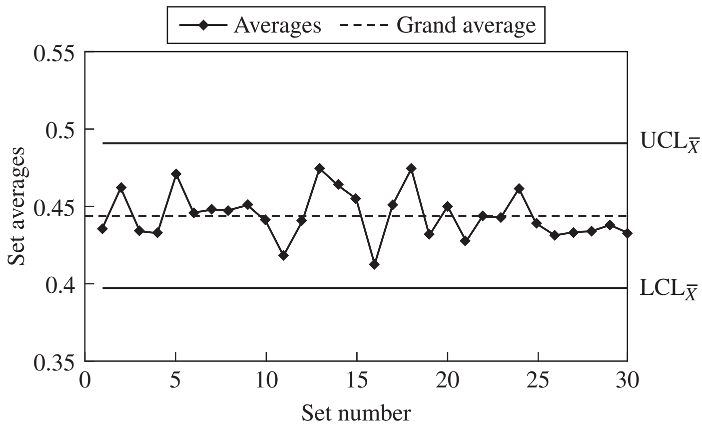

Figure 11.33 Fuzzy $\overline{X}$ chart for a "normal" 30-day beryllium plant run.

example, the membership functions are truncated as follows: Good = 0.4, Fair = 0.52, OK = 0.35, Bad = 0.45, and Terrible = 0.45. The shaded area in Figure 11.32 shows the results of the truncation in this example. The defuzzified value is the centroid (Chapter 4) of the shaded area in Figure 11.32, and is equal to 0.5036. So on day 1, machinist 2 had an OK Type of Day ($0.5036 \approx 0.5$).

The next step is to provide an average and a distribution for the entire day based on the results from each machinist. The procedure outlined here can be followed for each machinist, for day 1. The Type of Day results for the other machinists were as follows: machinist 1 = 0.4041, machinist 3 = 0.4264, and machinist 4 = 0.4088. The set average, or $\overline{X}$, for day 1 is 0.4357. This is the point for the first data set shown in Figure 11.33, the fuzzy Type of Day $\overline{X}$ chart. For the same 30-day run, the daily average beryllium exposure and beryllium exposure ranges were also plotted in the form of $\overline{X}-R$ charts. The $\overline{X}$ chart is presented in Figure 11.34. In Figure 11.33, all of the important variables are taken into account and the control chart indicates that nothing is out of control. This is the result that we would expect from this simulation because it is based on random numbers, representing only the normal or "common-cause" variation. In Figure 11.34, the traditional SPC technique, two points are above the upper control limit, representing an out-of-control situation. This represents two false alarms generated

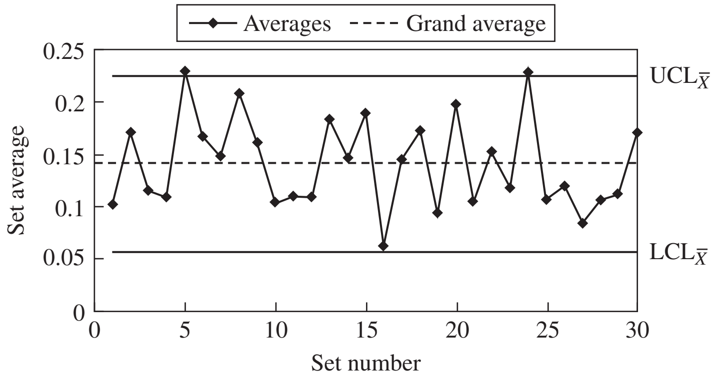

Figure 11.34 Beryllium exposure $\overline{X}$ chart for a "normal" 30-day beryllium plant run.

because all of the important variables are not factored into the solution of the problem. The corresponding R charts are not shown because, for this example, they did not add any information. Other simulations were run, in which the system was purposely perturbed. In all cases, where the significant variables influenced the outcome, the fuzzy SPC technique significantly outperformed the traditional SPC technique.

Attribute Data: Traditional SPC

The p chart is probably the most common test used with attribute data. Other common attribute data test charts are np charts, c charts, and u charts. Wheeler and Chambers (1992) and Mamzic (1995) both give good descriptions of these types of tests. We will use the p chart test as an illustration here. The p chart looks very much like the $\overline{X}$ chart. If we use the same example as we did for the cylindrical parts measurement data problem, we can construct a p chart. This time, suppose we accept only cylindrical parts with a measured diameter of 0.575 ± 0.020 inches; we will reject the rest. This creates the binary attribute system of "accept" and "reject." In the first sample set, we have five measurements of 0.60, 0.59, 0.54, 0.57, and 0.58 inches. We will reject the parts with diameters of 0.60 and 0.54 inches. The proportion rejected, p, is then 2/5 or 0.4. The p chart corresponding to the original example with this "accept–reject" criteria is shown in Figure 11.35, where the p value for set number 1 is 0.4. $\overline{p}$ is computed as the average proportion rejected for the entire 20 sets. The upper and lower control limits can be computed as follows:

$$\mathrm{UCL}_p = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}, \tag{11.28}$$

$$\mathrm{LCL}_p = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{\overline{n}}}. \tag{11.29}$$

These limits are ±3 standard deviations from the mean, based on having a true binomial distribution to represent the data. The reader will probably notice that these limits are wider than

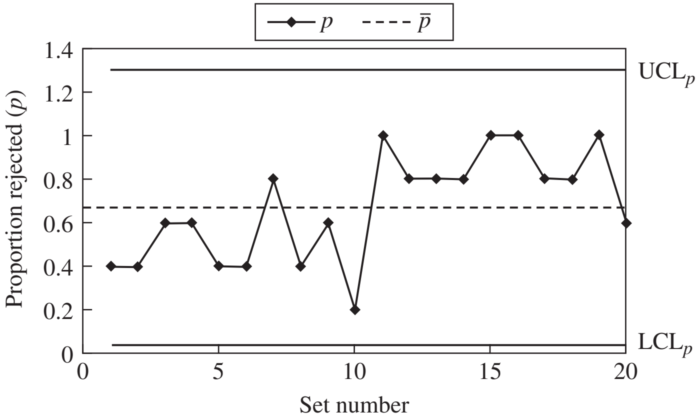

Figure 11.35 The p chart for the example problem.

those for the $\overline{X}$ chart. The reason for the tighter limits for the $\overline{X}$ chart is that the values for $\overline{X}$ have already been averaged, and the standard deviation must be again divided by $\sqrt{n}$. The p chart in Figure 11.35 would have much tighter bounds if it had been developed totally with data that were "in control"; the p chart would then tell us a different story. One advantage of the p chart is that the simple formulation for the control limits is based on the convenient properties of the binomial distribution. Our question is as follows: What happens if there are more than two attributes and the binomial distribution no longer applies? Our beryllium quality control study addresses this question.

The quality control issue in our beryllium manufacturing plant is a multivalued attribute problem. The beryllium material is expensive; therefore, care is taken not to waste it. Suppose the parts from this plant are required to satisfy two different applications: one application requires stringent quality control and the second application requires somewhat less quality control. An effort is made to rework parts that do not meet specification because of the cost of the material. Therefore, parts produced from the plant can fall into one of five categories: Premium or "Firsts," "Seconds," "Culls," "Possible Rework to a First," and "Possible Rework to a Second." In SPC terms, these are attribute data, normally analyzed with a p chart. Two papers by Raz and Wang (1990) presented some fuzzy solutions to multivalued or multinomial attribute problems. This work was criticized by Laviolette and Seaman (1992, 1994) and Laviolette, Seaman, Barrett, and Woodall (1995). But one problem with the work presented by Raz and Wang was that their example membership functions were not well defined. There has to be some physical justification for assigning the membership functions. In our beryllium simulation, we take special care in matching fuzzy membership functions to the physical world.

Our computer simulation is the same as that for Example 11.7 with the addition of the randomly generated flaws that represent the number of scratches, their length, and their depth. A 30-day simulation was accomplished to generate some sample control charts. This inspection technique is only one of several proposed, and the number and size of flaws in the material are strictly arbitrary.

Example 11.8
Attribute Data—Fuzzy SPC Our fuzzy system is divided into two parts. A fuzzy rule-base is used to assign inspected parts to the proper category. This is the connection to the physical world that provides realistic membership functions for each category. The second part is the fuzzy approach to developing a multinomial *p* chart. The development of the *p* chart works as follows:

- First, we choose our sample set size to be the total daily plant production.
- Then, we count the number of parts that fall into each category, Firsts, Seconds, and so on, from the sample set.
- We use the fraction in each category to define a fuzzy sample set in terms of membership functions.
- We defuzzify the sample to get a single representative value for the sample set and use that to construct a *p* chart.
- Finally, we use the fuzzy *p* chart just like a binomial *p* chart to determine the state of our process.

All parts going through the beryllium plant are inspected twice. The first inspector measures part dimensions with a machine to see if they are within the desired tolerance limits. This machine is a precision instrument, so parts are assigned to categories in a crisp manner. There are three major categories, "Firsts," "Seconds," and "Culls." Suppose there is a substantial demand for the "Seconds" and a tight tolerance for the "Firsts." Enough "Seconds" are generated by this procedure to satisfy demand. The "Firsts" are assigned a score of 0.0, the "Seconds" a score of 0.5, and the "Culls" a score of 1.0. Owing to the high cost of the beryllium, two more categories are added; they represent the possibility of reworking the part to qualify it to be either a "First" or a "Second." A second inspector looks at the surface finish of the part. This inspector visually checks for scratches, and records the number of scratches, the average length, and the average depth of any scratch. The numbers used in this example are not the true values for scratch sizes; they are representative values created by the plant simulator.

A fuzzy rule-based system determines how much to "downgrade" each part from the first inspection category based on the number, depth, and length of the scratches. The fuzzy membership functions that describe the beryllium parts after the inspections are shown in Figure 11.36.

The fuzzy system used to downgrade parts consists of three input variables and one output variable. The input variables are as follows:

- Number of Scratches with a range of 0–9.
- Length of Scratches with a range of 0–2.4 cm.
- Depth of Scratches with a range of 0–10 μm.

The output variable is the Amount of Downgrade.

Each input variable has three membership functions and the output variable has five membership functions. The inputs and the outputs are related by 27 rules. The input membership functions are quite simple; each variable is represented by three triangles; two equal-sized right triangles, with an equilateral triangle in the middle. The area of the equilateral triangle is twice the area of one of the right triangles. The configuration is such that the sum of membership

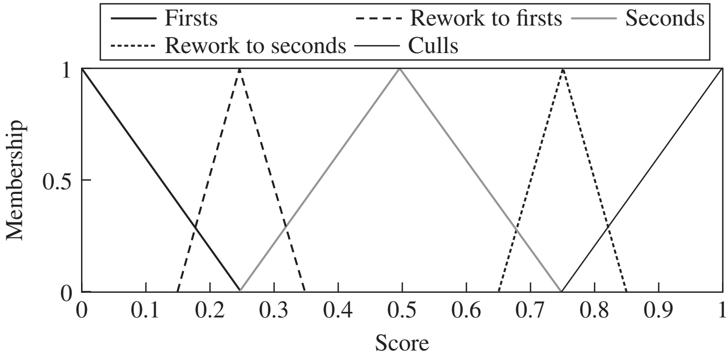

Figure 11.36 Fuzzy membership functions describing beryllium parts after inspections.

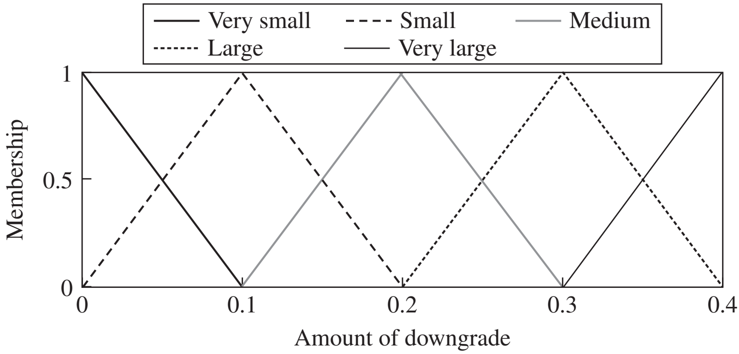

Figure 11.37 Output membership functions for rules to downgrade beryllium parts.

values for each variable adds to unity (the so-called orthogonal membership functions defined in Chapter 9). The output membership functions are a little bit more interesting and are shown in Figure 11.37. The following is the form of the rules:

If (Number of Scratches) is … and If (Depth of Scratches) is …

and If (Length of Scratches) is … Then the Amount of Downgrade is ….

The input variables are determined by measurements or estimates by the second inspector. The appropriate rules are fired using a max–min Mamdani procedure, and a centroid method is used to defuzzify the result. This defuzzified value is added to the score given by the first inspector producing a final score. The final score is used with Figure 11.36 to place the part in its final category (Firsts, Rework to Firsts, Seconds, Rework to Seconds, Culls). The categorization is accomplished by projecting a vertical line onto the score chart, Figure 11.36, at the point on the abscissa corresponding to the final score, and then picking the category with the highest membership value.

We used the information produced by our 30-day simulation of the plant operation to build a control chart for the beryllium manufacturing process. For each part, in each set, for each day, the categorization procedure described is carried out. The next step is to provide a fuzzy representation for each day based on the categorization of each part for that day. This is done using the *extension principle* (Chapter 12) and the concept of a triangular fuzzy number (TFN) (Kaufmann and Gupta, 1985). A TFN can be completely described by the vector $[t_1, t_2, t_3]^T$. The values t_1, t_2, and t_3 are the x values of the $x-y$ pairs representing the corners of a triangle with the base resting on the x axis ($y=0$) and the apex resting on the line $y=1$. For example, the triangular membership function Seconds in Figure 11.36 can be described as a TFN with $t_1=0.25$, $t_2=0.5$, and $t_3=0.75$, or $[0.25, 0.5, 0.75]^T$. The other four output membership functions are described as follows:

- Firsts = $[0.0, 0.0, 0.25]^T$
- Rework to Firsts = $[0.15, 0.25, 0.35]^T$
- Rework to Seconds = $[0.65, 0.75, 0.85]^T$
- Culls = $[0.75, 1.0, 1.0]^T$.

A matrix, called the *A matrix*, can be constructed with columns comprising the five output membership function TFNs. For this example, the A matrix is

$$\mathrm{A} = \begin{bmatrix} 0.0 & 0.15 & 0.25 & 0.65 & 0.75 \\ 0.0 & 0.25 & 0.5 & 0.75 & 1.0 \\ 0.25 & 0.35 & 0.75 & 0.85 & 1.0 \end{bmatrix}.$$

Next a five-element vector called **B** is constructed. The first element of the **B** vector is the fraction of the daily readings that were Firsts. The second element is the fraction of the readings that were Rework to Seconds, and so on. A **B** vector can be constructed for every day of the run. For example, for day 1, our simulation generated 13 parts and categorized them in the following manner:

- number of parts categorized as Firsts = 6;
- number of parts categorized as Rework to Firsts = 2;
- number of parts categorized as Seconds = 2;
- number of parts categorized as Rework to Seconds = 1;
- number of parts categorized as Culls = 2.

The following value is obtained for the **B** vector:

$$\mathbf{B} = [6/13, 2/13, 2/13, 1/13, 2/13]^T = [0.462, 0.154, 0.154, 0.077, 0.154]^T.$$

The product A**B** is a TFN that represents the fuzzy distribution for the day. For day 1 of the 30-day run, the TFN A**B** $\approx [0.227, 0.327, 0.504]^T$. This is a triangular distribution that is approximately halfway between Rework to Firsts and Seconds. Figure 11.38 shows how day 1 is distributed on a "Score chart" like Figure 11.36. The shaded area is the TFN, or fuzzy distribution for day 1. A different distribution is obtained every day. To construct a control chart, values for both a centerline and control limits must be determined. There are several metrics that can be used to represent the central tendency of a fuzzy set; the metric we used is the

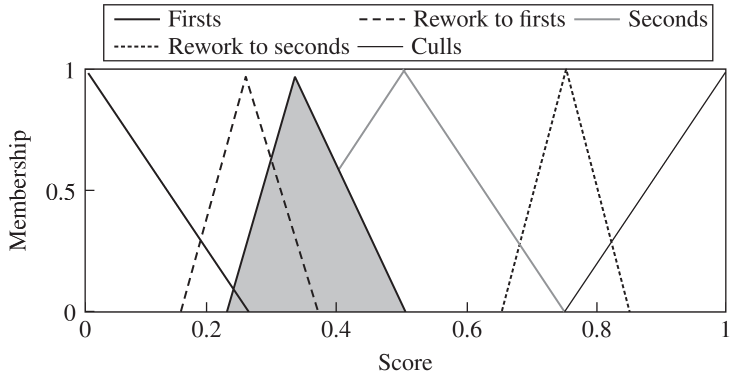

Figure 11.38 The fuzzy distribution for day 1, the shaded area, shown on the "Score chart."

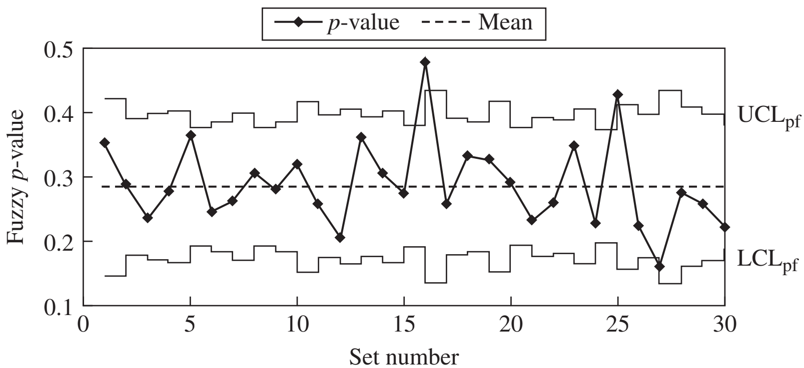

Figure 11.39 Fuzzy p chart using the α-level fuzzy midrange technique with $\alpha = 1/3$.

α-level (we called this a λ-level in Chapter 4) fuzzy midrange, as in Laviolette and Seaman (1994). The α-level fuzzy midrange, for $\alpha = 1/3$, for day 1 is 0.353. This is the midrange of the shaded triangle in Figure 11.38. The mean or the centerline for the 30-day run has a value of 0.285. These are the values shown in Figure 11.39 for the set 1 p value and the mean. Figure 11.39 is the p chart for the 30-day simulation run.

The α-level fuzzy midrange is defined for a fuzzy set as the midpoint of the crisp interval that divides the set into two subsets. One subset contains all of the values that have a membership in the original set of greater than or equal to α. The other subset contains all of the values with memberships less than α. This interval is called the α-cut (or the λ-cut in Chapter 4). The central tendency of the TFN A**B** described can be represented by using Equation (11.30).

$$R = \underline{\boldsymbol{\alpha}}^{\mathrm{T}} \mathrm{A}\mathbf{B}, \tag{11.30}$$

where R is the α-level fuzzy midrange and $\underline{\alpha}$ is a vector defined as

$$\underline{\boldsymbol{\alpha}} = \left[\frac{1-\alpha}{2} \alpha \frac{1-\alpha}{2}\right]^{\mathrm{T}}, \tag{11.31}$$

where α is the scalar value chosen for the α-cut. If we pick α equal to 1/3, then the vector $\underline{\alpha} = [1/3, 1/3, 1/3]$. For sufficiently large sample size, n, the vector **B** constitutes an observation from a multivariate normal distribution with a rank of $c - 1$, where c is the number of categories in the problem. The scalar R is then an observation from a univariate normal distribution with a mean $\mu = \underline{\boldsymbol{\alpha}}^{\mathrm{T}} \mathrm{A}\pi$ and a variance $\sigma^2 = \underline{\boldsymbol{\alpha}}^{\mathrm{T}} \mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}} \underline{\boldsymbol{\alpha}}$ where $\boldsymbol{\pi}$ and $\sum$ are the respective mean vector and covariance matrix of the set of **B** vectors. The covariance matrix $\sum$ is defined as

$$\Sigma = \left[\sigma_{ij}\right] = \begin{cases} \pi_i(1-\pi_i)/n, & i=j, \\ -\pi_i\pi_j/n, & i \neq j. \end{cases} \tag{11.32}$$

This is convenient because it provides upper and lower control limits, $\mathrm{UCL_{pf}}$ and $\mathrm{LCL_{pf}}$, for the fuzzy p chart:

$$\mathrm{UCL_{pf}} = \mu + z_c\sigma, \tag{11.33}$$

$$\mathrm{LCL_{pf}} = \mu - z_c\sigma. \tag{11.34}$$

The factor z_c is a function of the confidence level for the normalized Gaussian random variable. These factors are readily available; for example, see Williams (1991). For this problem, we picked $z_c = 1.96$, for a 95% confidence level. The control limits for each sample will be a function of n, the sample size. This can be seen in Figure 11.39 where the irregular shapes of the upper and lower control limits, $\mathrm{UCL_{pf}}$ and $\mathrm{LCL_{pf}}$, were computed using the technique described. Figure 11.39 shows that two points are beyond the 95% confidence limit; these are points 16 and 25. Of 30 points, 28 are within the control limits; this represents 93.33% of the data, which is close to 95% level that we would expect.

Industrial Applications

Two papers have provided an excellent review of the wealth of industrial products and consumer appliances that are bringing fuzzy logic applications to the marketplace. One paper describes fuzzy logic applications in a dozen household appliances (Quail and Adnan, 1992) and the other deals with a large suite of electronic components in the general area of image processing equipment (Takagi, 1992). A conference in 1992 dealt entirely with industrial applications of fuzzy control (Yen, Langari, and Zadeh, 1992).

Few of us could have foreseen the revolution that fuzzy set theory has already produced. Zadeh himself predicts that fuzzy logic will be part of every appliance when he says that we will "see appliances rated not on horsepower but on IQ" (Rogers and Hoshai, 1990). In Japan, the revolution has been so strong that "fuzzy logic" has become a common advertising slogan (Reid, 1990). Although the Eastern world equates the word *fuzzy* with a form of computer intelligence, the Western world still largely associates the word derisively within the context of "imprecise or approximate science."

The consumer generally purchases new appliances based on their ability to streamline housework and to use the consumer's available time more effectively. Fuzzy logic is being incorporated worldwide in appliances to accomplish these goals, primarily in the control mechanisms designed to make them work. Appliances with fuzzy logic controllers provide the consumer with optimum settings that more closely approximate human perceptions and reactions than those associated with standard control systems. Products with fuzzy logic monitor user-defined settings, and then automatically set the equipment to function at the user's preferred level for a given task. For example, fuzzy logic is well suited to making adjustments in temperature, speed, and other control conditions found in a wide variety of consumer products (Loe, 1991) and in image-processing applications (Takagi, 1992). Ross (1995) provided a good summary of several industrial applications using fuzzy control, including blood pressure control during anesthesia (Meier, Nieuwland, Zbinden, and Hacisalihzade, 1992), autofocusing for a 35-mm camera (Shingu and Nishimori, 1989), image stabilization for video camcorders (Egusa, Akahori, Morimura, and Wakami, 1992), adaptive control of a home heating system (Altrock et al., 1993), and adaptive control of an automobile's throttle system (Cox, 1993). The literature abounds in papers and books pertaining to fuzzy control systems (see, for example, Passino and Yurkovitz, 1998).

A recent search on the internet to look at the uses of Fuzzy Logic from the Internet (Nov. 15, 2012, Quora.com) has revealed a number of industrial applications:

- Air Conditioners: Old ACs used simple on-off mechanism. When the temperature dropped below a preset level, the AC was turned off. When it rose above a preset level, the AC was turned on. There was a slight gap between the two preset values to avoid high frequency on-off cycling. An example would be "When the temperature rises above 25 °C, turn on the unit, and when temperature falls below 20 °C, turn off the unit." Using Fuzzy Rules like "If the ambient air is getting warmer, turn the cooling power up a little; if the air is getting chilly, turn the cooling power down moderately" etc., the machine will become smoother as a result of this and give more consistently comfortable room temperatures.
- Automatic Gear Transmission System: It uses several variables like speed, acceleration, throttle opening, rate of change of throttle opening, engine load, and assigns a weight to each of these. A fuzzy aggregate is calculated from these weights and is used to decide whether to shift gears.
- Washing Machine: It senses the load size, detergent amount, etc. Keep track of the water clarity. At the start of a cycle, the water will be clean and will allow light to pass through it easily. As the wash cycle proceeds, the water becomes discolored and allows less light to pass through it. This information is used and control decisions are made.
- Reading hand written input and interpreting the characters for data entry.
- Television: A fuzzy logic scheme uses sensed variables such as ambient lighting, time of the day, and user profile to adjust parameters such as screen brightness, color, contrast and sound.
- Criminal Search System: Helps in criminal investigation by analyzing photos of the suspects along with their characteristics like "short, young-looking and so forth" to form witnesses to determine the most likely criminals.
- Online Disease Diagnostic System: Analyses the user's symptoms and tries to identify the disease he or she may be suffering from.

- Error Correction in Information Reception: Is used for limited-bandwidth communication links which are affected by data-corrupting noise. The front-end of a decoder produces a likelihood measure for the value intended by the sender (0-or-1) for each bit in the data stream. The likelihood measures might use a scale of 256 values between extremes of "certainly 0" and "certainly 1." The two decoders may analyze the data in parallel, arriving at different likelihood results for the values intended by the sender. Each can then use as additional data the other's likelihood results and repeat the process to improve the results until consensus is reached as to the most likely values.
- A massive engine used in the movie "The Lord of the Rings," which helped show huge-scale armies create random, yet orderly movements.

A recent quote from the popular writer, Dr. Laura Schlessinger, highlights the differences between standard control systems (based on binary logic) and fuzzy control systems (123HelpMe.com, Feb., 2, 2016) to wit: "It's funny how when we're the recipient of pain we're clear that it's black and white. But when we've got something to gain there are shades of gray." Almost any control system can be replaced with a fuzzy logic based control system. This may be overkill in many places however it simplifies the design of many more complicated cases. So fuzzy logic is not the answer to everything; it must be used when appropriate to provide better control. If a simple closed loop or PID controller works fine then there is no need for a fuzzy controller. There are many cases when tuning a PID controller or designing a control system for a complicated system is overwhelming, and this is where fuzzy logic gets its chance to shine.

One of the most famous applications of fuzzy logic is that of the Sendai Subway system in Sendai, Japan. This control of the Nanboku line, developed by Hitachi, used a fuzzy controller to run the train all day long. This made the line one of the smoothest running subway systems in the world and increased efficiency as well as improved stopping time. This is also an example of the earlier acceptance of fuzzy logic in the far-east since the subway went into operation in 1988.

The most tangible applications of fuzzy logic control have appeared in commercial appliances. Specifically, but not limited to, heating ventilation and air conditioning (HVAC) systems. These systems use fuzzy logic thermostats to control the heating and cooling; this saves energy by making the system more efficient and it keeps the temperature more steady than a traditional thermostat.

Another significant area of application of fuzzy control is in industrial automation. Fuzzy logic based programmable logic controllers (PLCs) have been developed by companies like Moeller. These PLCs, as well as other implementations of fuzzy logic, can be used to control any number of industrial processes.

Fuzzy logic also finds applications in many other systems. For example, the MASSIVE3D animation system for generating crowds uses fuzzy logic for artificial intelligence. This program was used extensively in the making of the "Lord of the Rings" trilogy as well as "The Lion, The Witch and the Wardrobe" films.

As a final example of fuzzy logic, it can be used in areas other than simply control. Fuzzy logic can be used in any decision making process such as signal processing or data analysis. An example of this is a fuzzy logic system that analyzes a power system and diagnoses any harmonic disturbance issues. The system analyzes the fundamental voltage, such as third, fifth and seventh harmonics as well as the temperature to determine if there is cause for concern in the operation of the system.

Summary

New generations of fuzzy logic controllers are based on the integration of conventional and fuzzy controllers. Fuzzy clustering techniques have also been used to extract the linguistic IF–THEN rules from the numerical data. In general, the trend is toward the compilation and fusion of different forms of knowledge representation for the best possible identification and control of ill-defined complex systems. The two new paradigms—artificial neural networks and fuzzy systems—try to understand a real-world system starting from the very fundamental sources of knowledge, that is, patient and careful observations, measurements, experience, and intuitive reasoning and judgments, rather than starting from a preconceived theory or mathematical model. Advanced fuzzy controllers use adaptation capabilities to tune the vertices or supports of the membership functions or to add or delete rules to optimize the performance and compensate for the effects of any internal or external perturbations. Learning fuzzy systems try to learn the membership functions or the rules. In addition, principles of genetic algorithms, for example, have been used to find the best string representing an optimum class of input or output symmetrical triangular membership functions (Chapter 4).

It would take an entire book to thoroughly discuss the subject of classical control theory and there are many good ones available. The interested reader should see, for example, Phillips and Harbor (1996), Shinskey (1988), Ogunnaike and Ray (1994), or Murrill (1991). For more information on the fuzzy logic control aspects, the reader is referred to Ross and colleagues (2002), Passino and Yurkovich (1998), Wang (1997), and Parkinson (2001).

All the engineering process control examples are for set-point-tracking control. For a problem involving the tougher problem of disturbance rejection, the reader is referred to Problem 11.10 at the end of this chapter. But, in terms of set-point control, it is interesting to compare fuzzy control with classical (e.g., PID) control. A fuzzy controller is often, but not always, simpler to write, especially if the engineer understands the physics of the system. Fuzzy control is often faster than PID control, and with little or no overshoot. As an example, Figure 11.40 shows PID versus fuzzy control results for a large-scale two-tank problem discussed in Parkinson (2001). Fuzzy control is invariably easier and better with MIMO or nonlinear systems. But sometimes, fuzzy control can get reasonably complicated if the membership functions are difficult to determine. On the other hand, classical control has elegant mathematics, it is quite accurate for ideal problems, and it gives a basis for determining stability. Classical control algorithms can be quite simple to write for "ideal" problems, with "one-size fitting all." Unfortunately, classical control rests fundamentally on the engineer's ability to determine good PID constants, and sometimes this is difficult for many nonlinear problems.

The previous comparison between classical and fuzzy control provokes a more general discussion of these two paradigms. Figure 11.41 shows a schematic comparing the two paradigms—fuzzy and classical—from a high-level perspective. In both paradigms, we start with the real world; some call this the "plant," that is, the object that we wish to control. The introduction to this chapter describes the "plant" as a simple mechanical device or something as large and pervasive as a country's economy. For this discussion, we will assume the plant is some piece of hardware. In classical control, the first step is to develop a mathematical model of the plant; this step is one of abstraction, because to model some complex object we must make some simplifying assumptions before we can describe a system with, say, differential equations. And, in accomplishing this abstraction, we encounter the most difficult step in this parade: being able to justify the mathematical model in terms of the simplifying

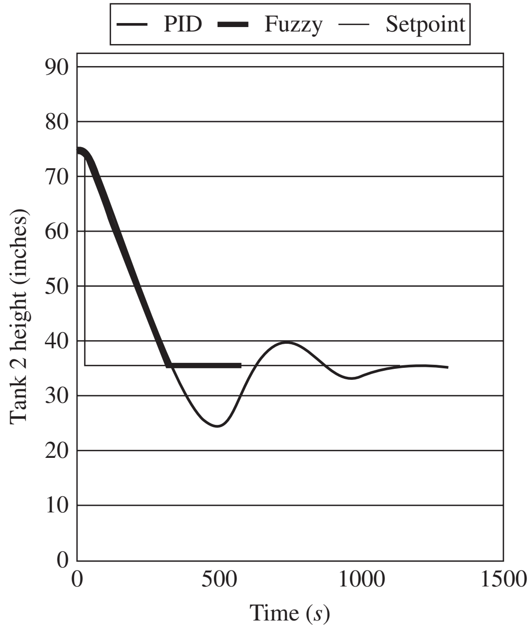

Figure 11.40 PID versus fuzzy control for a two-Tank problem. Adapted from Parkinson, 2001.

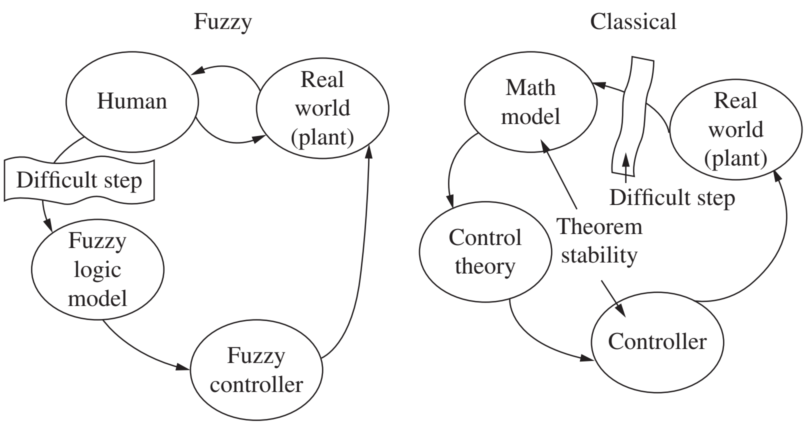

Figure 11.41 Conceptual differences between classical and fuzzy control (personal communication, Dorato, 2008).

assumptions that are inherently required. After the math model is developed, we invoke control theory on the equations of the model, which results in the description of the controller—usually another mathematical model. The controller is then applied to the plant, and we will see if the controller actually works. In most cases, our controller will be stabilizing as some theorems can prove stability if the mathematical model is well behaved. For the fuzzy paradigm, the real world, or plant, is often a device that is controlled by a human, or some other kind of "intelligent agent." In Figure 11.41, there are two arrows cycling between the human and the plant; this is because there is a strong feedback between the human and the plant in the human-control process. Humans offer their expertise, usually in the form of linguistic rules, about how they control the device. Using these rules, and associated membership functions (which presumably also come from the same intelligent agent), a fuzzy rule-based controller is developed. Using fuzzy systems theory, an automated control model will result. Again, the test of the controller's value is to verify that it controls the plant in an acceptable way, or least in a way that is as "good" as a human controller.

Both paradigms shown in Figure 11.41 illustrate the process involved in developing an automated controller. In the classical case, the control model results from a mathematical abstraction of the real world, and in the fuzzy case the control model results from the expertise of a human controller. In both cases, there is the burden to show that the automated controller actually works; in both cases the measures of "goodness" can be the same: for example, time to a stable configuration, size of overshoot, and amount of energy expended in the control action. In the case of classical control, there may be an advantage if the controller can be shown to be stable. The question of which controller is the best may be a difficult one to answer for some plants; this question may be analogous to asking a group of people to agree on which "painting they prefer" among a Rembrandt, a Picasso, and a Ruben.

For the illustrations of fuzzy SPC, the fuzzy "type of day" Shewhart-type $\overline{X}-R$ control chart has the potential to take into account task-dependent beryllium exposure for beryllium plant operations. On the basis of the studies completed to this point, we believe that these control charts will provide more realistic information than the standard single-variable $\overline{X}-R$ chart using only beryllium exposure information. Because of the ability to take into account task-dependency, "the type of day" chart can be used to determine the significance of plant improvements as well as initiate "out-of-control" alarms. This fuzzy technique should work well with many other task-dependent problems, which are characteristic of small-lot problems, as long as they are well defined semantically. A least-squares approach, which was studied but not presented here, will also work for this type of problem, but in many cases will not be as descriptive as the fuzzy approach. The least-squares approach can produce problems if the data used to develop a control chart have many out-of-control points in them. This is because the technique squares the difference between the expected value and the measured value.

The fuzzy technique for dealing with multinomial attribute data works quite well if the problem is defined well. With our simulation, we have also compared the fuzzy technique with individual p charts that deal with multinomial attribute data. Although not presented here, these comparisons have shown the fuzzy technique to be superior. We have also compared our fuzzy technique with the chi-square technique for multinomial data and have discovered that the chi-square technique works nearly as well as the fuzzy technique, but has only one control limit. The single control limit is not a problem in the examples illustrated here because all efforts were directed toward keeping the process below the upper limit. If both upper and lower limits are important, additional work will be required to determine the meaning given by the chi-square chart.

The computer models of the beryllium plant operation described here were built from a semantic description of the process as was the fuzzy rule-base and membership functions. Consequently, the correlation between the fuzzy model and the plant simulation was quite good. Both models come from the same description. It is important when developing a fuzzy model of a process that a lot of care is taken to listen to the experts and get the best model possible. If the domain expert is knowledgeable this task is usually not that difficult, but it may require several iterations to achieve validity. The fuzzy control chart will only be as good as the fuzzy rules and membership functions that comprise the system.

References

Altrock, C., Arend, H., Krause, B., et al. (1993). *Customer-Adaptive Fuzzy Control of Home Heating System* (Piscataway, NJ: IEEE Press).

Cox, E. (1993). Adaptive fuzzy systems. *IEEE Spectr.*, **30** (2), 27–31.

Egusa, Y., Akahori, H., Morimura, A., and Wakami, N. (1992). An electronic video camera image stabilizer operated on fuzzy theory. *IEEE International Conference on Fuzzy Systems* (San Diego, CA, IEEE Press).

Kaufmann, A., and Gupta, M. M. (1985). *Introduction to Fuzzy Arithmetic: Theory and Applications* (New York: Van Nostrand Reinhold).

Kiszka, J. B., Gupta, M. M., and Nikfrouk, P. N. (1985). Some properties of expert control systems, in M. M. Gupta, A. Kendal, W. Bandler, and J. B. Kiszka, eds., *Approximate Reasoning in Expert Systems*, pp. 283–306. (Amsterdam, Elsevier Science).

Kuhn, T. (1962). *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).

Laviolette, M., and Seaman, J. W. Jr. (1992). Evaluating fuzzy representations of uncertainty. *Math. Sci.*, **17**, 26–41.

Laviolette, M., and Seaman, J.W. Jr. (1994). The efficacy of fuzzy representations of uncertainty. *IEEE Trans. Fuzzy Syst.*, **2**, 4–15.

Laviolette, M., Seaman, J. W. Jr., Barrett, J. D., and Woodall, W. H. (1995). A probabilistic and statistical view of fuzzy methods. *Technometrics*, **37** (3), 249–261.

Loe, S. (1991). SGS-Thomson launches fuzzy-logic research push. *Electron World News,* **August 12**, p. 1.

Mamdani, E. (1974). Application of fuzzy algorithms for control of simple dynamic plant. *Proc. IEEE*, **121**, 1585–1588.

Mamdani, E. H., and Gaines, R. R., eds. (1981). *Fuzzy Reasoning and Its Applications* (London: Academic Press).

Mamzic, C. L., ed. (1995). *Statistical Process Control*, (Research Triangle Park, NC: Instrument Society of America Press).

Meier, R., Nieuwland, J., Zbinden, A., and Hacisalihzade, S. (1992). Fuzzy logic control of blood pressure during anesthesia. *IEEE Control Syst. Mag.*, **12**, 12–17.

Murrill, P. W. (1991). *Fundamentals of Process Control Theory*, 2nd ed. (Research Triangle Park, NC: Instrument Society of America).

Ogunnaike, B. A., and Ray, W. H. (1994). *Process Dynamics, Modeling, and Control* (New York: Oxford University Press).

Olsen, R. M. (1961). *Essentials of Engineering Fluid Mechanics* (Scranton, PA: International Textbook).

Pappas, C., and Mamdani, E. (1976). *A Fuzzy Logic Controller for a Traffic Junction*, Research report (London: Queen Mary College).

Parkinson, J., and Ross, T. (2002) Control charts for statistical process control, in T. J. Ross, J. M. Booker, and W. J. Parkinson, eds., *Fuzzy Logic and Probability Applications: Bridging the Gap*, pp. 263–323. (Philadelphia, PA: Society for Industrial and Applied Mathematics).

Parkinson, W. J. (2001). Fuzzy and probabilistic techniques applied to the chemical process industries. PhD dissertation. Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM.

Passino, K., and Yurkovich, S. (1998). *Fuzzy Control* (Menlo Park, CA: Addison-Wesley).

Phillips, C. L., and Harbor, R. D. (1996). *Feedback Control Systems*, 3rd ed. (Englewood Cliffs, NJ: Prentice Hall).

Quail, S., and Adnan, S. (1992). State of the art in household appliances using fuzzy logic, in J. Yen, R. Langari, and L. Zadeh, eds., *Proceedings of the Second International Workshop—Industrial Fuzzy Control and Intelligent Systems*, pp. 204–213 (College Station, TX: IEEE Press).

Raz, T., and Wang, J. (1990). Probabilistic and membership approaches in the construction of control charts for linguistic data. *Prod. Plan. Control*, **1**, 147–157.

Reid, T. (1990). The future of electronics looks 'fuzzy'; Japanese firms selling computer logic products. *Washington Post*, p. H–1.

Rogers, M., and Hoshai, Y. (1990). The future looks 'fuzzy'. *Newsweek*, **115** (22), 46–47.

Ross, T. (1995). *Fuzzy Logic with Engineering Applications*, McGraw-Hill, New York.

Ross, T. J., Booker, J. M., and Parkinson, W. J., eds. (2002). *Fuzzy Logic and Probability Applications: Bridging the Gap* (Philadelphia, PA: Society for Industrial and Applied Mathematics).
Shewhart, W. A. (1986). *Statistical Method from the Viewpoint Economic of Quality Control* (New York: Dover).
Shingu, T., and Nishimori, E. (1989). Fuzzy-Based Automatic Focusing System for Compact Camera. *Proceedings of the Third International Fuzzy Systems Association Congress*, pp. 436–439 (Seattle, WA).
Shinskey, F. G. (1988). *Process Control Systems—Applications, Design, and Tuning*, 3rd ed. (New York: McGraw-Hill).
Sugeno, M. (1985). *Industrial Application of Fuzzy Control* (New York: North Holland).
Takagi, H. (1992). Survey: Fuzzy logic applications to image processing equipment, in J. Yen, R. Langari, and L. Zadeh, eds., *Proceedings of the Second International Workshop—Industrial Fuzzy Control and Intelligent Systems*, pp. 1–9 (College Stations, TX: IEEE Press).
Vadiee, N. (1993). Fuzzy rule-based expert systems—I, in M. Jamshidi, N. Vadiee, and T. Ross, eds., *Fuzzy Logic and Control: Software and Hardware Applications*, pp. 51–85 (Englewood Cliffs, NJ: Prentice Hall).
Wang, L. X. (1997). *A Course in Fuzzy Systems and Control* (Upper Saddle River, NJ: Prentice Hall).
Wheeler, D. J., and Chambers, D. S. (1992). *Understanding Statistical Process Control* 2nd ed. (Knoxville, TN: SPC Press).
Williams, R. H. (1991). *Electrical Engineering Probability* (St. Paul, MN: West Publishing Company).
Yen, J., Langari, R., and Zadeh, L., eds. (1992). *Proceedings of the Second International Workshop on Industrial Fuzzy Control and Intelligent Systems* (College Station, TX: IEEE Press).

Problems

In some of the problems to follow, differential equations are reduced to difference equations. The resulting difference equations will generally involve a parameter (Δt), but this parameter is not shown explicitly because it is assigned a value of unity, that is $\Delta t = 1$.

11.1 The interior temperature of an electrically heated oven is to be controlled by varying the heat input, u, to the jacket. The oven is shown in Figure P11.1a. Let the heat capacities of the oven interior and of the jacket be c_1 and c_2, respectively. Let the interior and the exterior jacket surface areas be a_1 and a_2, respectively. Let the radiation coefficients of the interior and exterior jacket surfaces be r_1 and r_2, respectively. Assume that there is uniform and instantaneous distribution of temperature throughout, and the rate of loss of heat is proportional to the area and the excess of temperature over that of the surroundings. If the external temperature is T_0, the jacket temperature is T_1, and the oven interior temperature is T_2, then we have

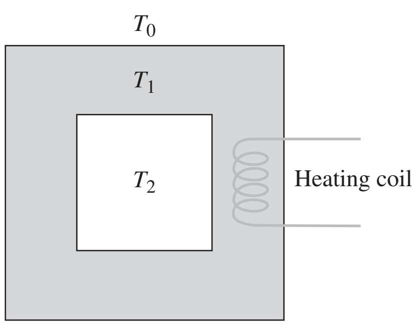

Figure P11.1a

$$c_1\dot{T}_1 = -a_2r_2(T_1 - T_0) - a_1r_1(T_1 - T_2) + u,$$
$$c_2\dot{T}_2 = -a_1r_1(T_1 - T_2).$$

Let the state variables be the excess of temperature over the exterior, that is, $x_1 = T_1 - T_0$ and $x_2 = T_2 - T_0$. With these substituted into the preceding equations, we find that they can be written as

$$\dot{x}_1 = -\frac{(a_2r_2 + a_1r_1)}{c_1}\cdot x_1 + \frac{a_1r_1}{c_1}\cdot x_2 + \frac{1}{c_1}\cdot u$$

and

$$\dot{x}_2 = \frac{a_1r_1}{c_2}\cdot x_1 - \frac{a_1r_1}{c_2}\cdot x_2.$$

Assuming that

$$\frac{a_2r_2}{c_1} = \frac{a_1r_1}{c_1} = \frac{a_1r_1}{c_2} = \frac{1}{c_1} = 1,$$

we have

$$\dot{x}_1(t) = -2x_1(t) + x_2(t) + u(t)$$

and

$$\dot{x}_2(t) = x_1(t) - x_2(t),$$

and

$$\dot{x}_i(t+1) = x_i(t) + \alpha\dot{x}_i(t), i = 1, 2.$$

Let $\alpha = 1/10$.

The membership functions for each of x_1, x_2, and u, each given on the same universe, are shown in Figure P11.1b. For each of the variables, membership functions are taken to be low (L), medium (M), and high (H) temperatures.

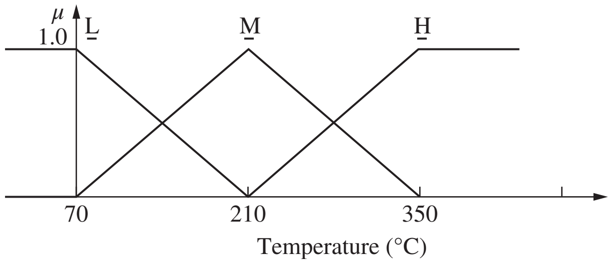

Figure P11.1b

Using the accompanying FAM table, conduct a graphical simulation of this control problem. The entries in the table are the control actions (u). Conduct at least four simulation cycles similar to Example 11.2. Use initial conditions of x_1 (0) = 80° and x_2 (0) = 85°.

	Interior temperature excess, x_2		
Jacket temperature excess, x_1	L	M	H
L	H	M	L
M	H	–	L
H	H	M	L

11.2 Conduct a simulation of an automobile cruise control system. The input variables are speed and angle of inclination of the road, and the output variable is the throttle position. Let speed = 0 to 100 (mph), incline = −10° to +10°, and throttle position = 0 to 10. The dynamics of the system are given as follows:

$$T = k_1 v + \theta k_2 + m\dot{v},$$

$$\dot{v} = v(n+1) - v(n),$$

$$T(n) = k_1 v(n) + \theta(n) k_2 + m(v_{n+1} - v_n),$$

$$v_{n+1} = \left(1 - \frac{k_1}{m}\right) v(n) + T(n) - \frac{k_2}{m}\theta(n),$$

$$v_{n+1} = k_a v(n) + [1 - k_b] \begin{bmatrix} T(n) \\ \theta(n) \end{bmatrix},$$

where

T = throttle position
k_1 = viscous friction
v = speed
θ = angle of incline
$k_2 = mg \sin\theta$
$.v^{\cdot}$ = acceleration
m = mass
$k_a = 1 - \dfrac{k_1}{m}$ and $k_b = \dfrac{k_2}{m}$.

Assign $k_1/m = k_2/m = 0.1$. The membership function for speed is determined by the cruise control setting, which we will assume to be 50 mph. The membership functions are shown in Figure P11.2.

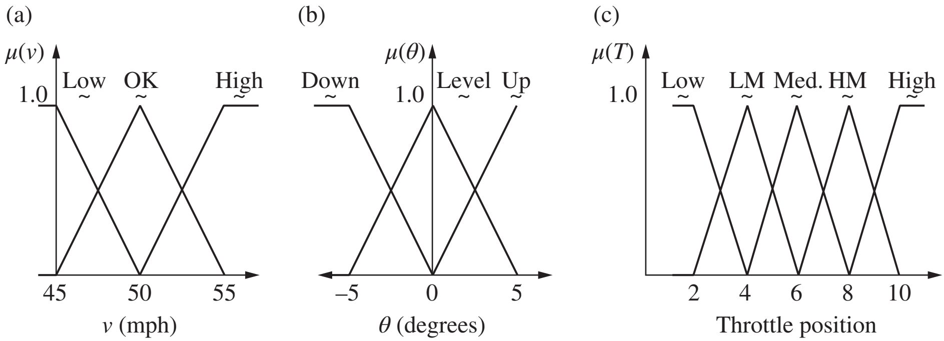

Figure P11.2

The FAM table is shown next:

	Inclination of the road		
Speed	Up	Level	Down
High	LM	LM	Low
OK	HM	Medium	LM
Low	High	HM	HM

Use initial conditions of speed = 52.5 mph and angle of incline = –5°. Conduct at least four simulation cycles.

11.3 A printer drum is driven by a brushless DC motor. The moment of inertia of the drum is, $J = 0.00185\ \text{kg m}^2$. The motor resistance is, $R = 1.12\ \Omega$. The torque constant for the motor is $K_T = 0.0363\ \text{N m}\ \underset{\sim}{\text{A}}^{-1}$. The back electromagnetic field (EMF) constant is $k = 0.0363\ \text{V}\ (\text{rad s}^{-1})^{-1}$. The equation of the system is

$$j\ddot{\theta} = \frac{K_{\text{T}}\left(V - \dot{\theta}\right)}{R},$$

where

$$\frac{\left(V - \dot{\theta}k\right)}{R} = I = \text{motor current}$$

$$\theta = \text{rotational angle}$$

$$V = \text{motor control voltage}$$

The state variables are $x_1 = \theta$ and $x_2 = \dot{\theta}$. Also,

$$\ddot{\theta} = \frac{K_{\text{T}}}{JR}V - \frac{K_{\text{T}}k}{JR}\dot{\theta}.$$

Now

$$x_2 = \dot{x}_1$$

Therefore,

$$\dot{x}_2 = \frac{K_{\mathrm{T}}}{JR}V - \frac{K_{\mathrm{T}}k}{JR}x_2.$$

Substituting in the values of the constants, we find

$$\dot{x}_2 + 0.64x_2 = 17.5V.$$

The resulting difference equations will be

$$x_1(k+1) = x_2(k) + x_1(k),$$
$$x_2(k+1) = 17.5V(k) + 0.36x_2(k).$$

The motor can be controlled to run at constant speed or in the position mode. The membership functions for x_1, x_2, and V are shown in Figure P11.3. The rule-based system is summarized in the following FAM table:

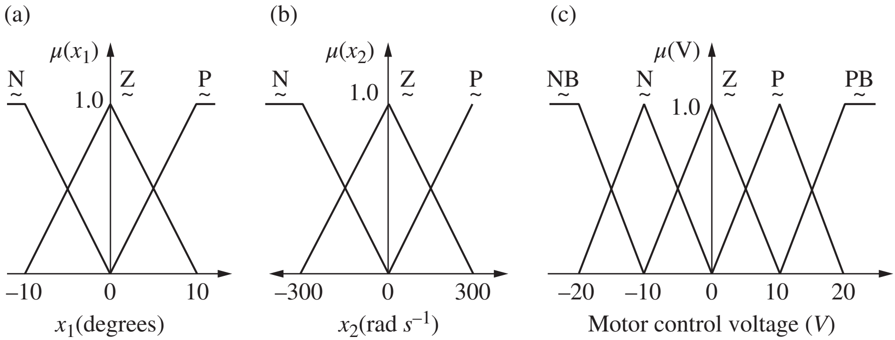

Figure P11.3

	x_2		
x_1	Negative	Zero	Positive
N	PB	P	N
Z	P	Z	N
P	Z	Z	NB

Using the initial conditions of $x_1 = 7.5°$ and $x_2 = -150\ \mathrm{rad\ s^{-1}}$ and the difference equations, conduct at least four graphical simulation cycles.

11.4 The basic mechanical system behind clocks that are enclosed in glass domes is the torsional pendulum. The general equation that describes the torsional pendulum is

$$J\frac{\mathrm{d}^2\theta(t)}{\mathrm{d}t^2} = \tau(t) - \underset{\sim}{B}\frac{\mathrm{d}\theta(t)}{\mathrm{d}t} - k\theta(t).$$

The moment of inertia of the pendulum bob is represented as J, the elasticity of the brass suspension strip is represented as k, and the friction between the bob and the air is represented as $\underset{\sim}{B}$. The controlling torque $\tau(t)$ is applied at the bob. When this device is used in clocks the actual torque is not applied at the bob but is applied through a complex mechanism at the main spring. The foregoing differential equation is the sum of the torques of the pendulum bob. The numerical values are $J = 1$ kg m^2, $k = 5$ N m rad^{-1}, and $\underset{\sim}{B} = 2$ N m s rad^{-1}.

The final differential equation with the foregoing constants incorporated is given as

$$\frac{d^2\theta(t)}{dt^2} + 2\frac{d\theta(t)}{dt} + 5\theta(t) = \tau(t).$$

The state variables are

$$x_1 = \theta(t) \text{ and } x_2(t) = \dot{\theta}(t),$$
$$\dot{x}_1 = \dot{\theta}(t) \text{ and } \dot{x}_2(t) = \ddot{\theta}(t).$$

Rewriting the differential equation using state variables, we have

$$\dot{x}_2(t) = \tau(t) - 2x_2(t) - 5x_1(t), \tag{P11.4.1}$$
$$\dot{x}_1(t) = x_2(t). \tag{P11.4.2}$$

Using these equations,

$$\dot{x}_1(t) = x_1(k+1) - x_1(k)$$
$$\dot{x}_2(t) = x_2(k+1) - x_2(k)$$

in Equations (P11.4.1) and (P11.4.2), and rewriting the equations in terms of θ and $\dot{\theta}$ in matrix form, we have

$$\begin{bmatrix} \theta(k+1) \\ \dot{\theta}(k+1) \end{bmatrix} = \begin{bmatrix} \theta(k) + \dot{\theta}(k) \\ -5\theta(k) - \dot{\theta}(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \tau(k) \end{bmatrix}.$$

The membership values for θ, $\dot{\theta}$, and τ are shown in Figure P11.4. The rules for the control system are summarized in the accompanying FAM table.

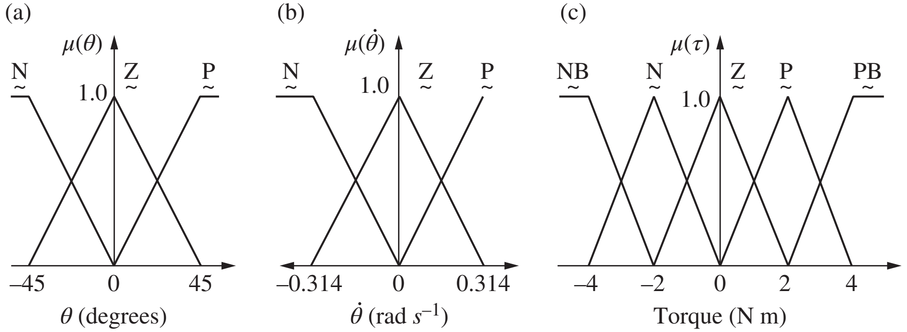

Figure P11.4

	$\dot{\theta}$		
Θ	Positive	Zero	Negative
P	NB	N	Z
Z	Z	Z	Z
N	Z	P	PB

The initial conditions are given as

$$\theta(0) = 0.7^\circ,$$
$$\dot{\theta}(0) = -0.2\,\text{rad}\,\text{s}^{-1}.$$

Conduct a graphical simulation for the control system.

11.5 On the electrical circuit shown in Figure P11.5a, it is desired to control the output current through inductor L_2 by using a variable voltage source, V. By using Kirchhoff's voltage law, the differential equation for loop 1 is given in terms of the state variables as

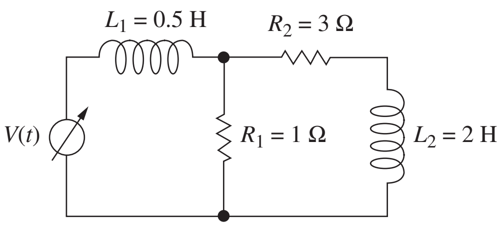

Figure P11.5a

$$\frac{\mathrm{d}I_1(t)}{\mathrm{d}t} = -2I_1(t) + 2I_2(t) + 2V(t) \tag{P11.5.1}$$

and that for loop 2 is

$$\frac{\mathrm{d}I_2(t)}{\mathrm{d}t} = 0.5I_1(t) - 2I_2(t). \tag{P11.5.2}$$

Converting the system of differential equations into a system of difference equations, we get

$$I_1(k+1) = -I_1(k) + 2I_2(k) + 2V(k),$$
$$I_2(k+1) = 0.5I_1(k) - I_2(k).$$

Rewriting the equations in matrix form, we get

$$\begin{bmatrix} I_1(k+1) \\ I_2(k+1) \end{bmatrix} = \begin{bmatrix} -I_1(k)+2I_2(k) \\ 0.5I_1(k)-I_2(k) \end{bmatrix} + \begin{bmatrix} 2V(k) \\ 0 \end{bmatrix}.$$

The membership functions are given in Figure P11.5b and the rules are presented in the following table:

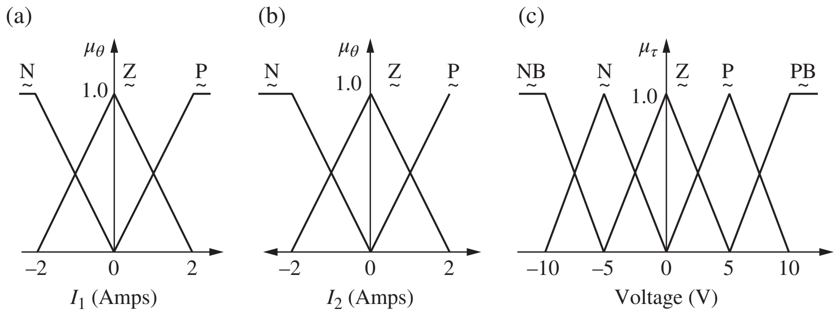

Figure P11.5b

I_1	I_2: Negative	Zero	Positive
N	PB	P	Z
Z	Z	Z	Z
P	Z	N	NB

The initial conditions are $I_1(0)=1$ $\underset{\sim}{A}$ and $I_2(0)=-1$ $\underset{\sim}{A}$. Conduct a simulation of the system.

11.6 We have a cylindrical tank with cross-sectional area, A_c. Liquid flows in at a rate F_i and liquid flows out at a constant rate F_o. We want to control the tank liquid level h using a level controller to change the liquid level set height h_s. The available tank liquid height is H_T. The flow rate in the tank (F_i) is proportional to the percentage that the value is opened. We call this set flow into the tank F_{is}.

$$F_i - F_o = A_c \frac{dh}{dt},$$

$$\frac{dh}{dt} = \frac{F_i - F_o}{A_c}.$$

The difference between the liquid-level set point and the actual tank liquid level is

$$e = h - h_s$$

and the percentage difference is

$$e = \frac{h - h_s}{h}.$$

The percentage difference is used to govern the flow into the system with the following rules:

$$\text{If } e=0\% \text{ then } \Delta Fi=0.$$
$$\text{If } e>10\% \text{ then } \Delta Fi=4\%.$$
$$\text{If } e<-10\% \text{ then } \Delta Fi=-4\%.$$

The percentage change in flow into the system (ΔF_i) is

$$\Delta F_i\% = \frac{F_{is}-F_i}{F_{is}}.$$

The initial values are

$$F_{is}=0.3\,\text{m}^3\,\text{s}^{-1}$$
$$F_{is}=0.3\,\text{m}^3\,\text{s}^{-1}$$
$$H_T=2\,\text{m}$$
$$A_c=3\,\text{m}^2$$
$$h_s=1\,\text{m}$$

At $t=0$, the disturbance in the inlet flow is

$$F_i=0.4\,\text{m}^3\,\text{s}^{-1}$$
$$F_o=0.3\,\text{m}^3\,\text{s}^{-1}$$
$$e=0$$

At $t=0.5$ s, $\Delta h=(0.4-0.3)/3=0.03$ m; thus, $h=1.03$ m and $e=(1.03-1)/1=3\%$.

We now make use of the fuzzy controller. The single-input membership function is as shown in Figure P11.6a,b whereas the output membership function is as shown in

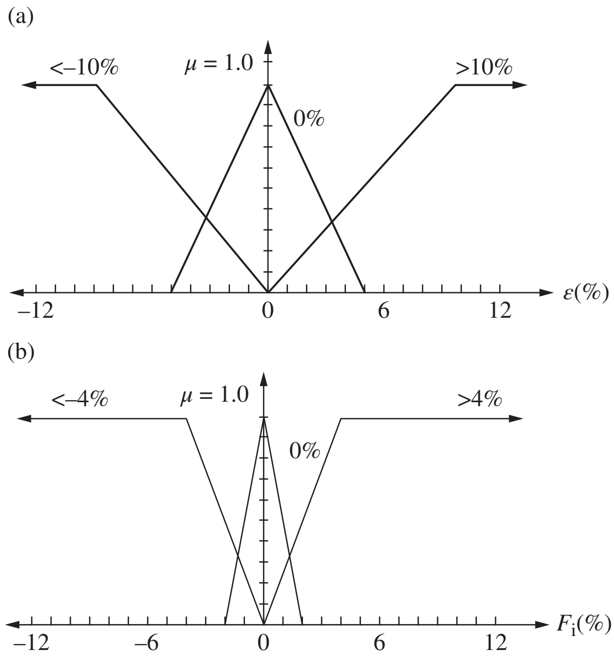

Figure P11.6a and b

Using a weighted average defuzzification, conduct a three-cycle simulation of this system.

11.7 The transport of toxic chemicals in water principally depends on two phenomena: advection and dispersion. In advection, the mathematical expression for time-variable diffusion is a partial differential equation accounting for concentration difference in space and time, which is derived from Fick's first law:

$$J = -DA \cdot \frac{\Delta C}{\Delta x},$$

$$V \cdot \frac{\Delta C}{\Delta t} = -DA \cdot \frac{\Delta C}{\Delta x} \quad \text{and} \quad V = A \cdot \Delta\alpha,$$

$$\frac{\Delta C}{\Delta t} = -D \cdot \frac{\Delta C}{\Delta x \Delta x} \quad \Rightarrow \quad \frac{\partial C}{\partial t} = -D \frac{\partial^2 C}{\partial x^2},$$

where

J = the mass flux rate due to molecular diffusion, mg s^{-1}
D = the molecular diffusion coefficient, cm^2 s^{-1}
A = the area of the cross section, cm^2
$\frac{\Delta C}{\Delta t}$ = the concentration gradient in time, mg cm^{-3} s^{-1}
Δx = movement distance, cm

So if we want to control $\Delta C/\Delta t$, we can set a control with the following inputs:

$$W_1 = C(\text{concentration, mg cm}^{-3}\,\text{cm}),$$

$$W_2 = \frac{\Delta C}{\Delta x}\left(\text{concentration gradient in space, mg cm}^{-3}\,\text{s}\right),$$

and the output: $\Delta C/\Delta t = \alpha$. So

$$\frac{\mathrm{d}W_1}{\mathrm{d}x} = W_2 \text{ and } \frac{\mathrm{d}W_2}{\mathrm{d}x} = -\alpha \quad \left(\text{if } D = 1.0\,\text{cm}^2\,\text{s}^{-1}\right)$$

Therefore,

$$W_1(k+1) = W_1(k) + W_2(k) \quad \text{and} \quad W_2(k+1) = W_2(k) - \alpha(k).$$

For this problem, we assume

$$0 \le W_1 \le 2000\,\text{mg cm}^{-3}$$

$$-400 \le W_2 \le 0\,\text{mg cm}^{-3}$$

$$0 \le \alpha \le 80\,\text{mg cm}^{-3}$$

(W_2 is negative because flow direction is from high concentration to low concentration).

Step 1: Partition W_1 to zero (PZ), low (PL), high (PH) (Figure P11.7a). Partition W_2 to zero (Z), low (NL), high (NH) (Figure P11.7b).

Step 2: Partition α to low (L) and high (H) (Figure P11.7c).

Step 3: Construct rules based on experience as in the following FAM table:

	W_2		
W_1	NZ	NL	NH
PZ	L	L	L
PL	L	L	H
PH	L	H	H

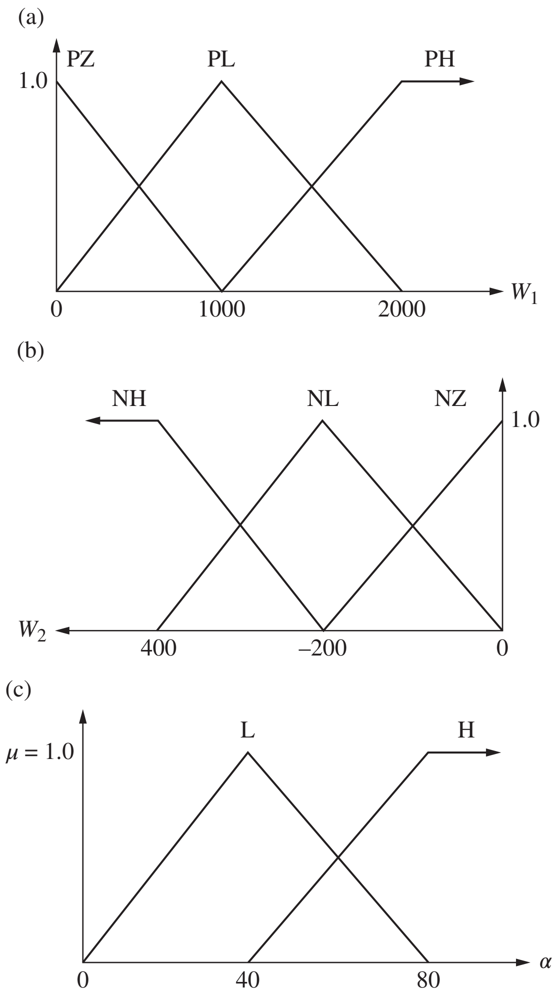

Figure P11.7a, b, c

Step 4: Initial conditions:

$$W_1(0) = 800\,\text{mg cm}^{-3} \text{ and } W_2(0) = -280\,\text{mg cm}^{-3}\,\text{cm}$$

Now conduct a two-cycle simulation using a centroidal defuzzification method.

11.8 Global information system (GIS) is a powerful tool in environmental modeling. It integrates geographical information with data stored in databases. The main issue in using GIS is selecting the appropriate spatial resolution. If the spatial resolution selected is low, then the mapping tool cannot fully represent the true topography. On the other hand, if the spatial resolution selected is too high, then the database size will be larger than necessary, thus increasing storage requirement and processing speed.

Two parameters can be used in a fuzzy control system to govern the GIS. The first one is the digital elevation (DE) value. This value is the difference between the highest and lowest elevation in a certain geographical area. The second parameter is the area of coverage (AC). The update equation is defined as follows:

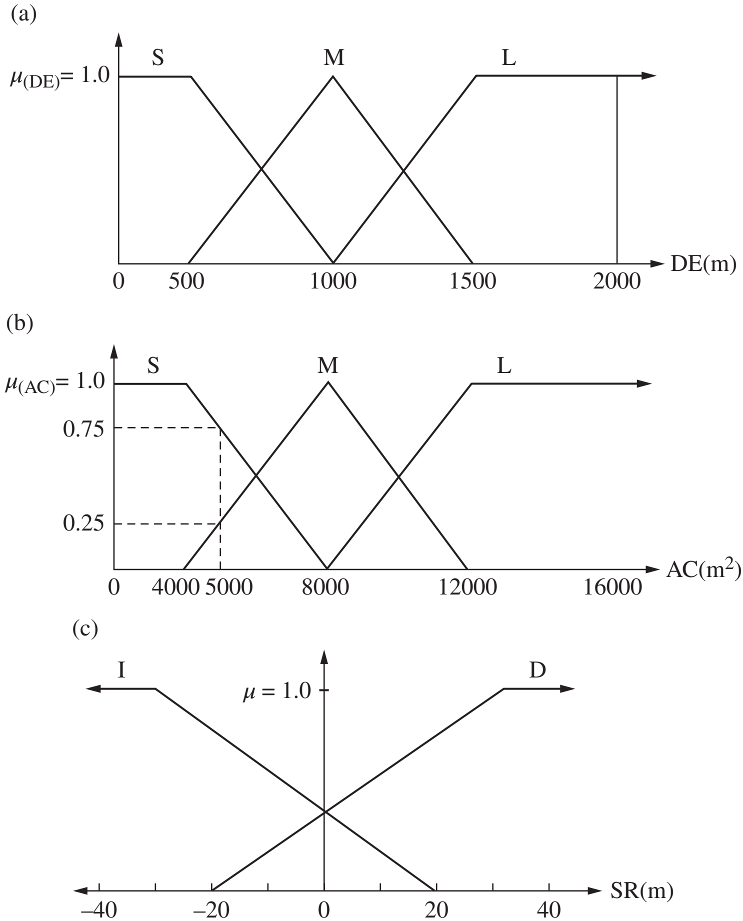

Figure P11.8a, b, c

$$DE_{new} = \frac{SR^2}{AC} DE_{old} + DE_{old}.$$

In the equation, SR represents spatial resolution (meters).

The first input is DE and can be either {small, medium, large} (meters) as shown in Figure P11.8a, the second input is AC and can be either {small, medium, large} (square meters) as shown in Figure P11.8b, and SR is the output, which can either be {increase (I), decrease (D)} as seen in Figure P11.8c and the FAM table, below.

	DE		
AC	L	M	S
L	D	I	I
M	I	D	I
S	D	D	D

$$\text{Initial condition for DE is DE}(0) = 2000\,\text{m}.$$

$$\text{Initial condition for AC is AC}(0) = 5000\,\text{m}^2.$$

Using a weighted average defuzzification, conduct a two-cycle simulation.

11.9 A businessman employs five people: one engineer to do his SPC work and four woodcarvers. The woodcarvers sit by the side of the road and carve figures of small animals for tourists. For the purposes of this problem, each figure is equally hard to carve. The tourist picks the type of figure and the type of wood that the figure is to be carved from. There are several types of wood with ratings of 0 for very soft to 10 for very hard. The tourist pays a price for the carving based on the number of flaws in the final product. The businessman wants to be able to keep track of the quality of the work, but knows that number of flaws alone is not a good metric. The number of flaws per worker per day is a function of the hardness of the wood and the number of carvings each worker has to do each day. The businessman decides to use the fuzzy "type of day" approach discussed in this chapter for his SPC work. He develops rules of the form

If the wood hardness is…and the number of carvings is…
and the number of flaws is…
Then the type of day is….

The input membership functions are described by the following TFNs:

Wood hardness: Soft (0.0, 0.0, 10.0); Hard (0.0, 10.0, 10.0).
Number of carvings: Small (0.0, 0.0, 5.0); Medium (0.0, 5.0, 10.0);
Large (5.0, 10.0, 10.0).
Number of flaws: Small (0.0, 0.0, 50.0); Medium (0.0, 50.0, 100.0);
Large (50.0, 100.0, 100.0).

The output membership functions are the day types Good, Fair, OK, Bad, and Terrible, and are exactly the same as those shown in the body of the text.

There are 18 rules and they are given in Table P11.9a.

Table P11.9a Woodcutters' rules.

Rule number	Wood hardness	Number of carvings	Number of flaws	Type of day
1	Soft	Small	Small	Fair
2	Soft	Small	Medium	OK
3	Soft	Small	Large	Bad
4	Soft	Medium	Small	OK
5	Soft	Medium	Medium	OK
6	Soft	Medium	Large	Bad
7	Soft	Large	Small	Good
8	Soft	Large	Medium	Fair
9	Soft	Large	Large	OK
10	Hard	Small	Small	OK
11	Hard	Small	Medium	Bad
12	Hard	Small	Large	Terrible
13	Hard	Medium	Small	Fair
14	Hard	Medium	Medium	OK
15	Hard	Medium	Large	OK
16	Hard	Large	Small	Fair
17	Hard	Large	Medium	OK
18	Hard	Large	Large	Bad

The businessman uses $\overline{X}-R$ charts to gain information about the quality of his product. For these charts, he computes a type of day for each worker, each day, using his fuzzy rule-based system. He then uses the four type of day readings to compute his set average and set range. He does this for about 20 working days and then computes his grand average, average range, and control limits. Because he is paying his woodcarvers the minimum wage, there is quite a bit of turnover. For this reason, he keeps his R charts to see if a statistical difference between workers develops. He also keeps the $\overline{X}$ charts to see if the average type of day is changing with any statistical significance over a period of time. Because the turnover rate is high, he does not know his carver's names. They are just called $\underset{\sim}{A}$, $\underset{\sim}{B}$, C, and D. One other thing that the businessman is looking for is, Has there been an out-of-control situation during the last control period? He makes his carvers work out of doors because it attracts tourists. But the number of flaws in the carvings also influence the price of the carvings and his profit.

The 20-day period has ended and the SPC engineer has nearly completed the analysis. There was some bad weather during this period and the businessman wants to know if there was a statistically significant effect on the quality of work, or on the type of day. Unfortunately, his engineer left work early before the calculation was completed. Table P11.9b lists the partially completed work of the SPC engineer.

Day 20 was nearly completed but not quite. Worker A had a type of day of 0.45, worker B 0.45, and worker C 0.43. The type of day calculation was not finished for worker D. Worker D had the following statistics: the number of carvings was 5, the total number of flaws was 35, and the average wood hardness for these wood carvings for the day was 8.

Table P11.9b Partially completed work.

Day or set number	Worker average type of day ($\overline{X}$)	Range (R)
1	0.43	0.10
2	0.45	0.07
3	0.41	0.06
4	0.41	0.08
5	0.62	0.12
6	0.59	0.06
7	0.58	0.03
8	0.44	0.03
9	0.44	0.06
10	0.43	0.02
11	0.40	0.08
12	0.43	0.08
13	0.47	0.08
14	0.46	0.05
15	0.45	0.06
16	0.40	0.04
17	0.45	0.06
18	0.47	0.02
19	0.42	0.10
20	–	–

Assume that you are the businessman. Finish the calculations by computing the following:

a. The type of day for worker D.
b. The set average and set range for day 20.
c. The grand average, the average range, and all of the control limits for the 20 $\overline{X}$ and R values.
d. Determine from the $\overline{X}$ chart if the system was ever "out of control."
e. Is there anything in the R chart that would indicate a significant difference between the workers at any time?

11.10 In this problem, we have a hot liquid that is a product stream coming from a poorly mixed stirred tank reactor. The reaction is exothermic so that the fluid leaving the reactor can become hot. This fluid is cooled with cold water flowing through a countercurrent heat exchanger, with the hot fluid on the shell side. The situation is depicted in Figure P11.10. The hot fluid temperature needs to be maintained at about 110 °F, because it is used in another process. There is some leeway. The process that is accepting the new hot fluid can easily handle a fluctuation of ±5 °F. Differences much larger than this start to become a problem. The cold water comes from a cooling tower where the temperature of the cold water is maintained at 85 °F. Because this temperature is constant, the only way of controlling the hot fluid temperature is to allow more or less cold water to flow through the heat exchanger. The flow is controlled by opening and

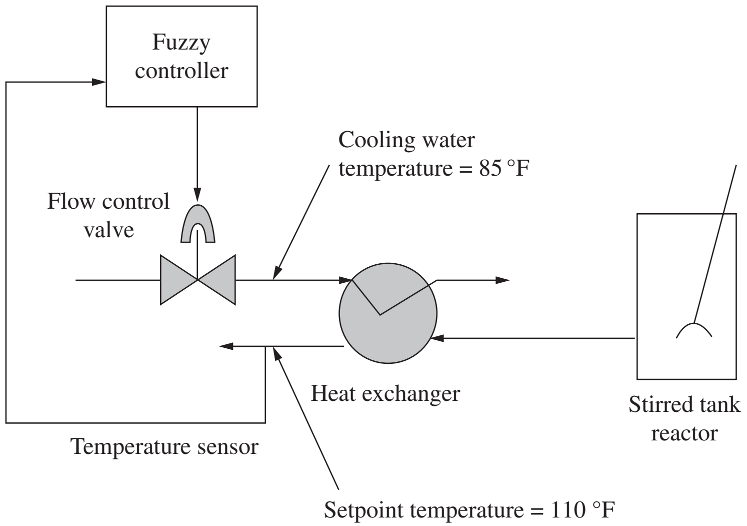

Figure P11.10

closing a control valve. The amount of flow through the valve is relative to the valve stem position. A stem position of 1 represents fully opened and a position of 0 equals fully closed. The system is controlled using an SISO fuzzy control system. The fuzzy rules for the system are of the following form:

If the ΔT is…Then the valve fractional change is … ,

where ΔT is $T - T_s$, or the current hot fluid temperature minus the set-point temperature.

The valve fractional change, f, is a fraction defined by the output membership functions of a Range. If the fraction, f, is greater than 0, the Range is defined as "f9211 open" (1.0) minus the current valve position. If f is less than 0, the Range is defined as the current valve position. The control action described by the fuzzy controller is as follows:

New valve position = old valve position + $f * $ Range.

The fuzzy rules are given in Table P11.10.

Table P11.10 Fuzzy rules for hot fluid problem.

Rule number	ΔT	Valve fractional change, f
1	Large positive	Large positive
2	Small positive	Small positive
3	Zero	Zero
4	Small negative	Small negative
5	Large negative	Large negative

There are five input membership functions for ΔT. Because they are not all triangles, we will give the (x, y) coordinates rather than the TFNs:

- Large negative: (−20.0, 1.0) (−15.0, 1.0) (−10.0, 0.0)
- Small negative: (−15.0, 0.0) (−10.0, 1.0) (−5.0, 0.0)
- Zero: (−10.0, 0.0) (−5.0, 1.0) (+5.0, 1.0) (+10.0, 0.0)—note the dead band
- Small positive: (+5.0, 0.0) (+10.0, 1.0) (+15.0, 0.0)
- Large positive: (+10.0, 0.0) (+15.0, 1.0) (+20.0, 1.0)

There are five output membership functions for valve fractional change. Because they are all triangles, we will give their TFNs:

- Large negative: (−1.5, −1.0, −0.5)
- Small negative: (−1.0, −0.5, 0.0)
- Zero: (−0.5, 0.0, 0.5)
- Small positive: (0.0, 0.5, 1.0)
- Large positive: (0.5, 1.0, 1.5)

Assuming that the current valve position is 0.6, calculate the following:

a. If the hot fluid temperature suddenly increases to 113 °F, what is the new valve position recommended by the fuzzy controller?
b. If the hot fluid temperature suddenly rises to 122 °F, what is the new valve position recommended by the fuzzy controller?
c. If the hot fluid temperature suddenly drops to 98 °F, what would be the new valve position recommended by the fuzzy controller?

12

Applications of Fuzzy Systems Using Miscellaneous Models

Knowing ignorance is strength, and ignoring knowledge is sickness.
Lao Tsu, Chinese philosopher, Tao Te Ching, circa 600 B.C.

This chapter exposes the reader to a few of the additional application areas that have been extended with fuzzy set theory and fuzzy logic. These few areas cannot cover the wealth of other applications, but they give to the reader an appreciation of the potential influence of fuzzy logic in almost any technology area. Addressed in this chapter are five additional application areas: optimization, fuzzy cognitive mapping, fuzzy agent-based models, fuzzy arithmetic, and fuzzy data fusion. The chapter summary gives the reader some references for other applications.

Fuzzy Optimization

Most technical fields, including all those in engineering, involve some form of optimization that is required in the process of design. Because design is an open-ended problem with many solutions, the quest is to find the "best" solution according to some criterion. In fact, almost any optimization process involves trade-offs between costs and benefits because finding optimum solutions is analogous to creating designs; there can be many solutions, but only a few might be optimum, or useful, particularly where there is a generally nonlinear relationship between performance and cost. Optimization, in its most general form, involves finding the most optimum solution from a family of reasonable solutions according to an optimization criterion. For all but a few trivial problems, finding the global optimum (the best optimum solution) can never be guaranteed. Hence, optimization in the last three decades has focused on methods to achieve the best solution per unit computational cost.

In cases where resources are unlimited and the problem can be described analytically and there are no constraints, solutions found by exhaustive search (Akai, 1994) can guarantee global optimality. In effect, this global optimum is found by setting all the derivatives of the criterion function to zero, and the coordinates of the stationary point, which satisfy the resulting simultaneous equations, represent the solution. Unfortunately, even if a problem can be described analytically, there are seldom situations with unlimited search resources. If the optimization problem also requires the simultaneous satisfaction of several constraints and the solution is known to exist on a boundary, then constraint boundary search methods such as Lagrangian multipliers are useful (deNeufville, 1990). In situations where the optimum is not known to be located on a boundary, methods such as the steepest gradient, Newton–Raphson, and penalty function have been used (Akai, 1994), and some promising methods have used genetic algorithms (Goldberg, 1989).

For functions with a single variable, search methods such as Golden section and Fibonacci are quite fast and accurate. For multivariate situations, search strategies such as parallel tangents and steepest gradients have been useful in some situations. But most of these classical methods of optimization (Vanderplaats, 1984) suffer from one or more disadvantages: the problem of finding higher-order derivatives of a process, the issue of describing the problem as an analytic function, the problem of combinatorial explosion when dealing with many variables, the problem of slow convergence for small spatial or temporal step sizes, and the problem of overshoot for step sizes too large. In many situations, the precision of the optimization approach is greater than the original data describing the problem, so there is an *impedance mismatch* in terms of resolution between the required precision and the inherent precision of the problem itself.

In the typical scenario of an optimization problem, fast methods with poorer convergence behavior are used first to get the process near a solution point, such as a Newton method; then slower but more accurate methods, such as gradient schemes, are used to converge to a solution. Some current successful optimization approaches are now based on this hybrid idea: fast, approximate methods first and slower and more precise methods second. Fuzzy optimization methods have been proposed as the first step in hybrid optimization schemes. One of these methods is introduced here. More methods can be found in Sakawa (1993).

One-Dimensional Optimization

Classical optimization for a one-dimensional (one independent variable) relationship can be formulated as follows. Suppose we wish to find the optimum solution, x^*, which maximizes the objective function $y = f(x)$, subject to the constraints

$$g_i(x) \leq 0, \quad i = 1, \ldots, m. \tag{12.1}$$

Each of the constraint functions $g_i\,(x)$ can be aggregated as the intersection of all the constraints. If we let $C_i = \{x | g_i(x) \leq 0\}$, then

$$\mathrm{C} = \mathrm{C}_1 \cap \mathrm{C}_2 \cap \ldots \cap \mathrm{C}_m = \{x | g_1(x) \leq 0, g_2(x) \leq 0, \ldots, g_m(x) \leq 0\}, \tag{12.2}$$

which is the feasible domain described by the constraints C_i. Thus, the solution is

$$f(x^*) = \max_{x \in \mathrm{C}} \{f(x)\}. \tag{12.3}$$

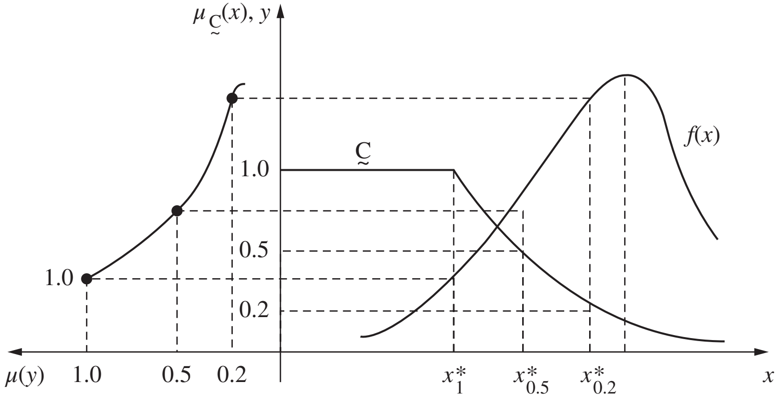

Figure 12.1 Function to be optimized, $f(x)$, and fuzzy constraint, C.

In a real environment, the constraints might not be so crisp, and we could have fuzzy feasible domains (Figure 12.1) such as "x could exceed x_0 a little bit." If we use λ-cuts on the fuzzy constraints $\underset{\sim}{C}$, fuzzy optimization is reduced to the classical case. Obviously, the optimum solution x^* is a function of the threshold level λ, as given in Equation (12.4):

$$f\left(x_\lambda^*\right) = \max_{x \in C_\lambda} \{f(x)\}. \tag{12.4}$$

Sometimes, the goal and the constraint are more or less contradictory, and some trade-off between them is appropriate. This can be done by converting the objective function $y = f(x)$ into a pseudogoal $\underset{\sim}{G}$ (Zadeh, 1972) with membership function

$$\mu_{\underset{\sim}{G}}(x) = \frac{f(x) - m}{M - m}, \tag{12.5}$$

where

$$m = \inf_{x \in X} f(x)$$

$$M = \sup_{x \in X} f(x).$$

Then, the fuzzy solution set $\underset{\sim}{D}$ is defined by the intersection

$$\underset{\sim}{D} = \underset{\sim}{C} \cap \underset{\sim}{G}, \tag{12.6}$$

membership is described as

$$\mu_{\underset{\sim}{D}}(x) = \min\{\mu_{\underset{\sim}{C}}(x), \mu_{\underset{\sim}{G}}(x)\}, \tag{12.7}$$

and the optimum solution will be x^* with the condition

$$\mu_{\underset{\sim}{D}}(x^*) \geq \mu_{\underset{\sim}{D}}(x), \quad \text{for all } x \in X, \tag{12.8}$$

where $\mu_{\underset{\sim}{C}}(x), \underset{\sim}{C}, x$ should be substantially greater than x_0. Figure 12.2 shows this situation.

Example 12.1
Suppose we have a deterministic function given as

$$f(x) = xe^{(1-x/5)},$$

for the region $0 \leq x \leq 5$, and a fuzzy constraint given as

$$\mu_{\underset{\sim}{C}}(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ \dfrac{1}{1+(x-1)^2}, & x > 1. \end{cases}$$

Both these functions are illustrated in Figure 12.3. We want to determine the solution set $\underset{\sim}{D}$ and the optimum solution x^*, that is, find $f(x^*) = y^*$. In this case, we have $M = \sup[f(x)] = 5$ and $m = \inf[f(x)] = 0$; hence, Equation (12.5) becomes

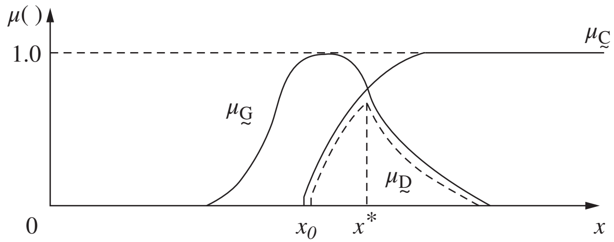

Figure 12.2 Membership functions for goal and constraint. Adapted from Zadeh, 1972.

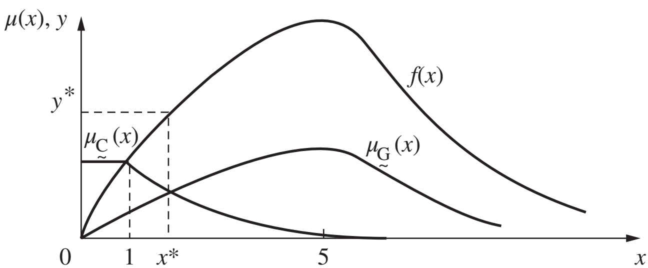

Figure 12.3 Problem domain for Example 12.1.

$$\mu_{\underset{\sim}{G}}(x)=\frac{f(x)-0}{5-0}=\frac{x}{5}e^{(1-x/5)},$$

which is also shown in Figure 12.3. The solution set membership function, using Equation (12.7), then becomes

$$\mu_{\underset{\sim}{D}}(x)=\begin{cases}\frac{x}{5}e^{(1-x/5)}, & 0\le x\le x^*,\\ \frac{1}{1+(x-1)^2}, & x>x^*,\end{cases}$$

and the optimum solution x^*, using Equation (12.8), is obtained by finding the intersection

$$\frac{x^*}{5}e^{(1-x^*/5)}=\frac{1}{1+(x-1)^2},$$

and is shown in Figure 12.3.

When the goal and the constraint have unequal importance, the solution set $\underset{\sim}{D}$ can be obtained by the convex combination, that is,

$$\mu_{\underset{\sim}{D}}(x)=\alpha\mu_{\underset{\sim}{C}}(x)+(1-\alpha)\mu_{\underset{\sim}{G}}(x). \tag{12.9}$$

The single-goal formulation expressed in Equation (12.9) can be extended to the multiple-goal case as follows. Suppose we want to consider n possible goals and m possible constraints. Then, the solution set $\underset{\sim}{D}$ is obtained by the aggregate intersection, that is, by

$$\underset{\sim}{D}=\left(\bigcap_{i=1,m}\underset{\sim}{C}_i\right)\cap\left(\bigcap_{j=1,n}\underset{\sim}{G}_j\right). \tag{12.10}$$

Example 12.2

A beam structure is supported at one end by a hinge and at the other end by a roller. A transverse concentrated load P is applied at the middle of the beam, as in Figure 12.4. The maximum bending stress caused by P can be expressed by the equation $\sigma_b=Pl/\mathrm{w}_z$, where w_z is a coefficient decided by the shape and size of a beam and l is the beam's length. The deflection at the centerline of the beam is $\delta=Pl^3/(48EI)$, where E and I are the beam's modulus of elasticity

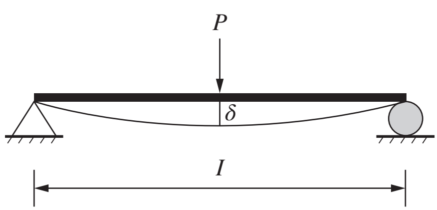

Figure 12.4 Simply supported beam with a transverse concentrated load.

and cross-sectional moment of inertia, respectively. If $0 \le \delta \le 2$ mm, and $0 \le \sigma_b \le 60$ MPa, the constraint conditions are as follows: the span length of the beam is

$$l = \begin{cases} l_1, & 0 \le l_1 \le 100\,\text{m}, \\ 200 - l_1, & 100 < l_1 \le 200\,\text{m} \end{cases}$$

and the deflection is

$$\delta = \begin{cases} 2 - \delta_1, & 0 \le \delta_1 \le 2\,\text{mm}, \\ 0, & \delta_1 > 2\,\text{mm}. \end{cases}$$

To find the minimum P for this two-constraint and two-goal problem (the goals are the stress, σ_b, and the deflection, δ), we first find the membership function for the two goals and two constraints.

1. The μ_{G1} for bending stress σ_b is given as follows:

$$P(0) = 0, P(60\text{MPa}) = \frac{w_z 60}{l}, P(\sigma_b) = \frac{w_z \sigma_b}{l}; \text{thus}, \mu_{G_1} = \frac{\sigma_b}{60}.$$

   To change the argument in μ_{G1} into a unitless form, let $x = \sigma_b/60$, where $0 \le x \le 1$. Therefore, $\mu_{G1}\ (x) = x$ when $0 \le x \le 1$.
2. The μ_{G2} for deflection δ is as follows:

$$P(\delta) = \frac{48EI\delta}{l^3}, P(0) = 0, P(2) = \frac{48EI \times 2}{l^3}; \text{thus}, \mu_{G_2} = \frac{\delta}{2}.$$

   Let $x = \delta/2$, so that the argument of μ_{G2} is unitless. Therefore, $\mu_{G2} = x$, $0 \le x \le 1$.
3. Using Equation (12.10), we combine $\mu_{G1}\ (x)$ and $\mu_{G2}\ (x)$ to find $\mu_G\ (x)$:

$$\mu_G(x) = \min\left(\mu_{G_1}(x), \mu_{G_2}(x)\right) = x, \quad 0 \le x \le 1.$$

4. The fuzzy constraint function μ_{C1} for the span is

$$\mu_{C_1}(x) = \begin{cases} 2x, & 0 \le x \le 0.5, \\ 2 - 2x, & 0.5 < x \le 1, \end{cases}$$

   where $x = l_1/200$. Therefore, the constraint function will vary according to a unitless argument x.
5. The fuzzy constraint function μ_{C2} for the deflection δ can be obtained in the same way as in point 4:

$$\mu_{C_2}(x) = \begin{cases} 1 - x, & 0 \le x \le 1, \\ 0, & x > 1, \end{cases}$$

   where $x = \delta/2$.

6. The fuzzy constraint function $\mu_C(x)$ for the problem can be found by the combination of $\mu_{C1}(x)$ and $\mu_{C2}(x)$, using Equation (12.10):

$$\mu_C(x) = \min\left(\mu_{C_1}(x), \mu_{C_2}(x)\right),$$

and $\mu_C(x)$ is shown as the bold line in Figure 12.5.

Now, the optimum solutions P can be found by using Equation (12.10):

$$D = (G \cap C),$$

$$\mu_D(x) = \mu_C(x) \wedge \mu_G(x).$$

The optimum value can be determined graphically, as seen in Figure 12.6, to be $x^* = 0.5$. From this, we can obtain the optimum span length, $l = 100$ m, optimum deflection, $\delta = 1$ mm, and optimum bending stress, $\sigma_b = 30$ MPa. The minimum load P is

$$P = \min\left(\frac{\sigma_b w_z}{l}, \frac{48EI\delta}{l^3}\right).$$

Suppose that the importance factor for the goal function $\mu_G(x)$ is 0.4. Then, the solution for this same optimization problem can be determined using Equation (12.9) as

$$\mu_D = 0.4\mu_G + 0.6\mu_C,$$

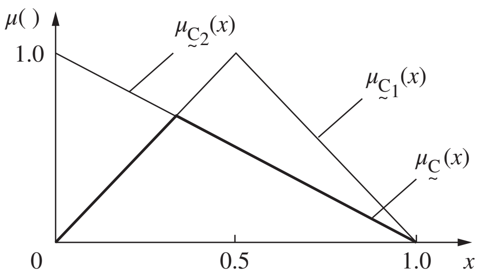

Figure 12.5 Minimum of two-constraint functions.

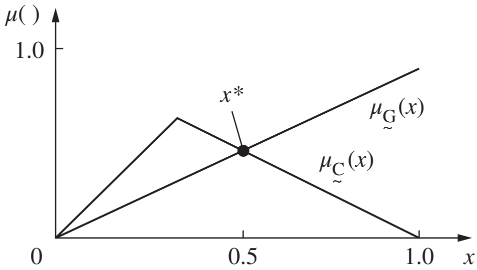

Figure 12.6 Graphical solution to minimization problem of Example 12.2.

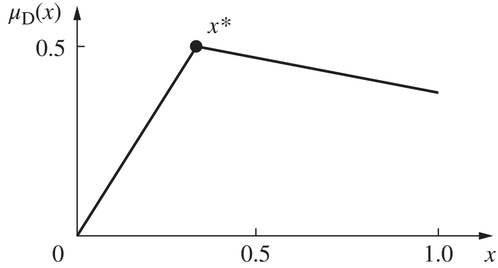

Figure 12.7 Solution of Example 12.2 considering an importance factor.

where μ_C can be expressed by the function (Figure 12.5)

$$\mu_C(x) = \begin{cases} 2x, & 0 \le x \le \frac{1}{3}, \\ 1-x, & \frac{1}{3} < x \le 1. \end{cases}$$

Therefore,

$$0.6\mu_C(x) = \begin{cases} 1.2x, & 0 \le x \le \frac{1}{3}, \\ 0.6-0.6x, & \frac{1}{3} < x \le 1, \end{cases}$$

and $0.4\mu_G(x) = 0.4x$. The membership function for the solution set, from Equation (12.9), is then

$$\mu_D(x) = \begin{cases} 1.6x, & 0 \le x \le \frac{1}{3}, \\ 0.6-0.2x, & \frac{1}{3} < x \le 1. \end{cases}$$

The optimum solution for this is $x^* = 0.33$, which is shown in Figure 12.7.

Fuzzy Cognitive Mapping

Cognitive maps (CMs) were introduced by Axelrod (1976) as a formal means of modeling decision making in social and political systems. CMs are a type of directed graph that offers a means to model interrelationships or causalities among concepts; there are various forms of CMs, such as signed digraphs, weighted graphs, and functional graphs. The differences among these various forms can be found in Kardaras and Karakostas (1999). CMs can also be used for strategic planning, prediction, explanation, and for engineering concept development. The use of simple binary relationships (i.e., increase and decrease) is done in a *conventional* (crisp) CM. All CMs offer a number of advantages that make them attractive as models for engineering planning and concept development. CMs have a clear way to visually represent causal relationships, they expand the range of complexity that can be managed, they allow users to rapidly compare their mental models with reality, they make evaluations easier, and they promote new ways of thinking about the issue being evaluated.

Concept Variables and Causal Relations

CMs graphically describe a system in terms of two basic types of elements: concept variables and causal relations. Nodes represent concept variables, C_x, where $x = 1, \dots, N$. A concept variable at the origin of an arrow is a cause variable, whereas a concept variable at the endpoint of an arrow is an effect variable. For example, for $C_h \rightarrow C_i$, C_h is the cause variable that impacts C_i, which is the effect variable. Figure 12.8 represents a simple CM, in which there are four concept variables (C_h represents utilization of waste steam for heat, C_i represents the amount of natural gas required to generate heat, C_k represents economic gain for the local economy, and C_j represents the market value of waste steam [which is dependent on the price of natural gas]).

Arrows represent the causal relations between concept variables, which can be positive or negative. For example, for $C_n \xrightarrow{-} C_i, C_h$ has a negative causal relationship on C_i. Therefore, an increase in C_h results in a decrease in C_i.

Paths and Cycles

A path between two concept variables, C_h and C_k, denoted by $P(h, k)$, is a sequence of all the nodes that are connected by arrows from the first node (C_h) to the last node (C_k) (Figure 12.8; Kosko, 1986). A cycle is a path that has an arrow from the last point of the path to the first point.

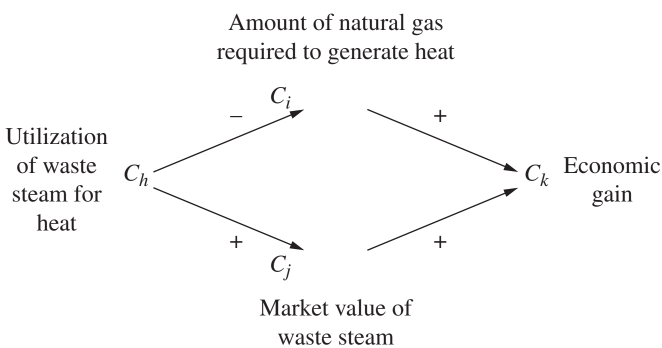

Figure 12.8 A conventional cognitive map for the utilization of waste steam.

Indirect Effect

The indirect effect of a path from the cause variable C_h to the effect variable C_k, which is denoted by $I\,(h, k)$, is the product of the causal relationships that form the path from the cause variable to the effect variable (Axelrod, 1976). If a path has an even number of negative arrows, then the indirect effect is positive. If the path has an odd number of negative arrows, then the indirect effect is negative. In Figure 12.8, the indirect effect of cause variable C_h on the effect variable C_k through path $P(h, i, k)$ is negative and the indirect effect of the cause variable C_h on the effect variable C_k through path $P(h, j, k)$ is positive.

Total Effect

The total effect of the cause variable C_h on the effect variable C_k, which is denoted by $T\,(h, k)$, is the union of all the indirect effects of all the paths from the cause variable to the effect variable (Axelrod, 1976). If all the indirect effects are positive, the total effect is positive. If all the indirect effects are negative, so is the total effect. If some indirect effects are positive and some are negative, the sum is indeterminate (Kosko, 1986). A large CM, that is one with a large number of concepts and paths, will therefore be dominated by the characteristic of being indeterminate. In Figure 12.8 the total effect of cause variable C_h to effect variable C_k is the collection of the indirect effect of C_h to C_k through the paths $P(h, i, k)$ and $P(h, j, k)$. Since one indirect effects is positive and the other is negative in this case, this means that the total effect is indeterminate.

Indeterminacy

The character of a conventional CM being indeterminate can be resolved, but it comes at a computational and conceptual price. To do so, the CM must accommodate a numerical weighting scheme (Kosko, 1986). If the causal edges are weighted with positive or negative real numbers, then the indirect effect is the product of each of the weights in a given path, and the total effect is the sum of the path products. This scheme of weighting the path relationships removes the problem of indeterminacy from the total effect calculation, but it also requires a finer causal discrimination. Such fineness may not be available from the analysts or experts who formulate the CM. This finer discrimination between concepts in the CM would make knowledge acquisition a more onerous process–forced numbers from insufficient decision information, different numbers from different experts or from the same expert on different days, and so on. However, causal relationships could be represented by linguistic quantities as opposed to numerical ones. Such is the context of a *fuzzy* cognitive map (FCM).

Fuzzy Cognitive Maps

If one were to emphasize that the simple binary relationship of a CM needed to be extended to include various degrees of increase or decrease (small decrease, large increase, almost no increase, etc.), then an FCM is more appropriate. An FCM extends the idea of conventional CMs by allowing concepts to be represented linguistically with an associated fuzzy set, rather than requiring them to be precise. Extensions by Taber (1994) and Kosko (1992) allow fuzzy numbers or linguistic terms to be used to describe the degree of the relationship between concepts in the FCM. FCMs are analyzed either geometrically or numerically (Pelaez and Bowles, 1996). A geometric analysis is used primarily for small FCMs, where it simply traces the increasing and decreasing effects along all paths from one concept to another. For larger FCMs, such as those illustrated later in this section, a numerical analysis is required, where the concepts are represented by a state vector and the relations between concepts are represented by a fuzzy relational matrix, called an *adjacency* matrix. This, along with a few other key features of FCMs that distinguish them from CMs, are mentioned in this chapter.

Adjacency Matrix

A CM can be transformed using a matrix called an *adjacency matrix* (Kosko, 1986). An adjacency matrix is a square matrix that denotes the effect that a cause variable (row) given in the CM has on the effect variable (column). Figure 12.9 is an adjacency matrix for the CM displayed in Figure 12.8. In other words, the adjacency matrix for a CM with n nodes uses an $n \times n$ matrix in which an entry in the (h, i) position of the matrix denotes an arrow between nodes C_h and C_i. This arrow (as shown in Figure 12.8) simply represents the "strength" of the effect between the two nodes (i.e., a "+1" represents that the effect is to increase, whereas a "−1" represents that the effect is to decrease).

Threshold Function

Concept states are held within defined boundaries through the threshold function. The type of threshold function chosen determines the behavior of a CM. A bivalent threshold function requires concepts to have a value of 1 or 0, which is equivalent to "on" or "off":

$$f(x_i) = 0, \quad x_i \le 0,$$
$$f(x_i) = 1, \quad x_i > 0.$$

The trivalent threshold function includes negative activation. Therefore, concepts have a value of 1, 0, or −1, which is equivalent to "positive effect," "no effect," and "negative effect," respectively:

$$f(x_i) = -1, \quad x_i \le -0.5.$$

$$\mathrm{E} = \begin{array}{c} \\ C_h \\ C_i \\ C_k \\ C_j \end{array} \begin{array}{c} \begin{array}{cccc} C_h & C_i & C_k & C_j \end{array} \\ \begin{bmatrix} 0 & -1 & 0 & +1 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 \end{bmatrix} \end{array}$$

Figure 12.9 The adjacency matrix for the cognitive map in Figure 12.8.

$$E = \begin{array}{c|cccc} & C_h & C_i & C_k & C_j \\ \hline C_h & 0 & -1 & 0 & +1 \\ C_i & 0 & 0 & +1 & 0 \\ C_k & 0 & 0 & 0 & 0 \\ C_j & 0 & 0 & +1 & 0 \end{array}$$

$$f(x_i) = 0, \quad -0.5 < x_i < 0.5.$$
$$f(x_i) = 1, \quad x_i \geq 0.5.$$

Concepts are multiplied by their connecting causal relation weights to give the total input to the effect concept. In cases where there are multiple paths connecting a concept, the sum of all the causal products is taken as the input (Tsadiras and Margaritis, 1996)

$$x_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} C_J w_{ji}, \tag{12.11}$$

where

x_i = input
C_j = concept state
w_{ji} = weight of the causal relations

Feedback

For FCMs, we can model dynamic systems that are cyclic, and therefore feedback within a cycle is allowed. Each concept variable is given an initial value based on the belief of the expert(s) of the current state. The FCM is then free to interact until an equilibrium is reached (Kosko, 1997). An equilibrium is defined to be the case when a new state vector is equal to a previous state vector.

Min–Max Inference Approach

The min–max inference approach is a technique that can be used to evaluate the indirect and total effects of an FCM. The causal relations between concepts are often defined by linguistic variables, which are words that describe the strength of the relationship. The min–max inference approach can be used to evaluate these linguistic variables (Pelaez and Bowles, 1995). The minimum value of the links in a path is considered to be the path strength. If more than one path exists between the cause variable and the effect variable, the maximum value of all the paths is considered to be the overall effect. In other words, the indirect effect amounts to specifying the weakest linguistic variable in a path, and the total effect amounts to specifying the strongest of the weakest paths. This is analogous to the "max–min composition operator" discussed in Chapter 3.

Example 12.3

Figure 12.10 depicts an FCM with five concept variables (C_1 represents *utilization of waste steam for heat*, C_2 represents amount of *natural gas required* to produce heat, C_3 represents the resulting carbon dioxide [CO_2] *emissions* produced from the burning of a methane-based gas, C_4 represents *carbon credits* that would need to be purchased, and C_5 represents the *economic gain*). Carbon credits are credits a company would receive from reducing its CO_2 emissions below the required level stipulated by the government's Kyoto implementation plan. Those companies not meeting their required level may need to purchase credits from others. In the FCM, the "effects" of the paths, P, are linguistic instead of simple binary quantities like a "+1" (increase) or a "−1" (decrease). However, the numerical quantities +1, 0, and −1 for a trivalent FCM are still used to convey the signs of the linguistic term. For example, a linguistic effect of "significant, +1" means that the effect is "significantly positive." A linguistic effect of "a lot −1" means that the effect is "negatively a lot." The values of the paths can be in matters of degree such as "none," "some," "much," or "a lot." In this example, then, $P = \{\text{none} < \text{some} < \text{much} < \text{a lot}\}$. These P values would be the linguistic values that would be contained within the adjacency matrix of the FCM:

$$\begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & 0 & -1 \\ 0 & 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

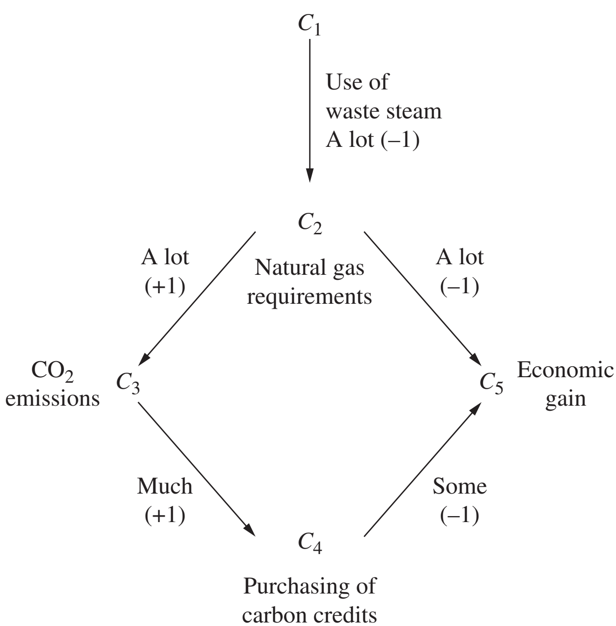

Figure 12.10 A fuzzy cognitive map involving waste steam and greenhouse gas emissions.

To implement the FCM, we start by activating C_1 (i.e., we begin the process by assessing the impact of an increase in waste steam for a facility); this results in the initial state vector

$$[1, 0, 0, 0, 0].$$

This state vector is activating only concept C_1 in Figure 12.10. Causal flow in the FCM was determined with repeated vector–matrix operations and thresholding (Pelaez and Bowles, 1995).

The new state is the old state multiplied by the adjacency matrix (Pelaez and Bowles, 1996):

$$[C_1 C_2 \ldots C_n]_{\text{new}} = [C_1 C_2 \ldots C_n]_{\text{old}} * \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \cdots & C_{nn} \end{bmatrix}. \tag{12.12}$$

The values of the state vector were thresholded to keep their values in the set {−1, 0, 1}, and the activated concept (in this case C_1) was reset to 1 after each matrix multiplication. Using the algorithm developed by Pelaez and Bowles (1995), we premultiply the trivalent adjacency matrix shown above by this initial state vector. At each iteration of this multiplication, the trivalent threshold function is invoked. This multiplication is continued until the output vector reaches a limiting state (i.e., it stabilizes). For this simple example, the resulting state vector stabilized after four iterations to the following form:

$$[1.0, -1.0, -1.0, -1.0, 1.0].$$

This stabilized output vector can be understood in the following sense. For an *increase* in the waste steam (+1), the natural gas requirements will *decrease* (–1), the CO_2 emissions will *decrease* (−1), and carbon credits also *decrease* (−1). Finally, there is an *increase* in economic gain (+1).

With a conventional CM, we would get the following results. First, we see that path $I_1 = (1, 2, 5)$ has two negative causal relationships, and its indirect path effect would be positive (two negatives yield a positive). For path $I_2 = (1, 2, 3, 4, 5)$, we see that it has two negative causal relationships (between C_1 and C_2 and between C_4 and C_5) and two positive relationships for the other two elements of the path; hence, the indirect effect of this path is positive (two negatives and two positives yield a positive effect). Hence, in the conventional CM characterization of this simple example, the results would be *determinate* (two positive indirect). If, however, one of the causal relationships in the I_2 path was positive, instead of negative (say, C_4–C_5 was positive), then the indirect effect of this path would have been negative, and the results would have been *indeterminate*. For an FCM, we can accommodate linguistic characterizations of the elements, as discussed previously and as seen in Figure 12.10. Hence, the issue of *indeterminacy* is not generally a problem. Two unique paths that exist from the cause variable (C_1) to the effect variable (C_5) are $I_1 = (1, 2, 5)$ and $I_2 = (1, 2, 3, 4, 5)$. The indirect effects of C_1 on C_5, expressed in terms of the linguistic values of P, are (Kosko, 1992)

$$I_1(C_1, C_5) = \min\{e_{12}, e_{25}\} = \min\{\text{alot}, \text{alot}\}$$
$$= \text{alot}.$$

$$I_2(C_1, C_5) = \min\{e_{12}, e_{23}, e_{34}, e_{45}\} = \min\{\text{alot, alot, much, some}\}$$
$$= \text{some}.$$

Therefore, the linguistic total effect is expressed as

$$T(C_1, C_5) = \max\{I_1(C_1, C_5), I_2(C_1, C_5)\}$$
$$= \max\{\text{alot}, \text{some}\} = \text{alot}.$$

Applying these linguistic results to the stabilized vector above, that is, [1.0, –1.0, –1.0, –1.0, 1.0], we come to the conclusion that an increase in waste steam for this facility results in "*a lot of increase*" in economic gain.

A more complete example of a FCM that was investigated at the University of New Mexico (UNM) is presented next. In this example, there is a comparison between a crisp CM and a fuzzy CM to give the readers an appreciation for the additional modeling capacity of an FCM.

Example 12.4

Figure 12.11 shows a simplified CM representing the UNM strategy as applied to the campus while it undergoes an energy demand response (DR) event. Energy DR refers to voluntary actions taken by the customers (in this case UNM) who change their consumption of electric power in response to price signals, incentives, or directions from electric power grid operators at times of high wholesale market prices or when electric system reliability is jeopardized. DR-driven changes in electricity use are designed to be short-term in nature, centered on critical hours during the day when demand is high or when the electricity supplier's reserve margins are low (see Federal Energy Regulatory Commission, 2007). In this example, the concepts shown in Figure 12.11 represent events and decisions related to energy consumption. Concept activation levels were ranged from −1 to +1 and their causal connections were weighted from −1 to +1. The following is the description of each of the seven concepts that make up this CM:

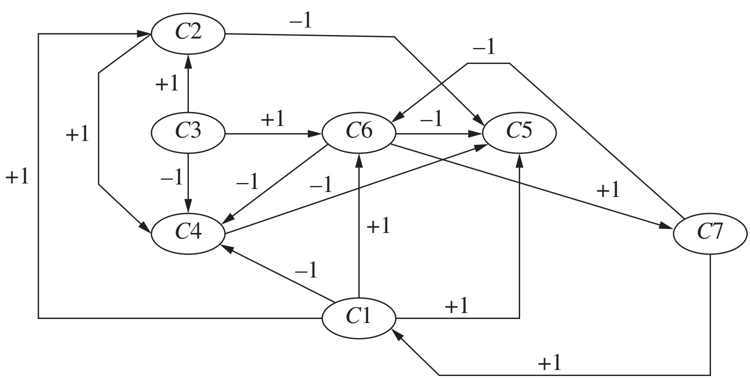

Figure 12.11 Conventional CM showing concepts and crisp activation levels for energy demand.

- *C1—Level of UNM electricity demand:* This is a measure of the quantity of electricity the UNM campus is consuming in total. An activation level of +1 indicates a high level of electric consumption while an activation level of 0 represents a low level.
- *C2—UNM generates electricity internally:* UNM has a cogeneration gas turbine in its Utilities Center capable of producing 6000 kW of electricity, approximately a third of the maximum campus usage, and supplying this to the campus electricity distribution system. An activation level of +1 represents full power output for the cogeneration turbine, 0 represents one half power output, and −1 represents no internal power generation.
- *C3—Public Service Company of New Mexico (PNM) sends UNM a curtailment signal through high pricing:* PNM supplies electricity to UNM through "time-of-use" pricing with kilowatt hour costs announced in advance for each hour. An activation level of +1 represents that a high price signal is present and 0 represents a low price signal.
- *C4—UNM stores energy:* This concept assumes that the UNM facilities are able to store energy by storing hot or cold water in insulated storage tanks for use at a later time. This is commonly done by commercial facilities at night when electric rates are low for use during the day when prices are higher. An activation level of +1 represents energy storage at the maximum rate, 0 represents storage at a moderate rate, and −1 represents no storage.
- *C5—UNM external (supplied by PNM) electric energy usage:* This concept represents the amount of the campus energy usage that is supplied by PNM as opposed to that supplied by the UNM Utilities Center cogeneration turbine or stored hot or cold water. An activation level of +1 represents a high level of PNM electricity usage, 0 represents a moderate level of usage, and −1 represents a low level of usage.
- *C6—DR by UNM to curtail PNM supplied electric energy use:* This concept measures the amount of curtailment that UNM performs. An activation level of +1 represents the maximum level of curtailment and 0 represents no curtailment.
- *C7—Occupant dissatisfaction:* This concept indicates the amount of dissatisfaction registered by the campus occupants due to the energy curtailment; an example would be complaints from occupants because the temperature in the buildings is too hot or too cold. An activation level of +1 represents a high level of dissatisfaction and a level of 0 represents no dissatisfaction.

Initially, UNM investigators analyzed the conventional CM in Example 12.4 using *crisp* initiating activation levels and interconnecting weights. Figure 12.11 shows the crisp interconnecting weights as supplied by the subject matter expert. The *crisp* adjacency matrix for these weights is shown in Figure 12.12. The four input vectors were used to determine the four stable output concept activation levels using the $(-1\ 0+1)$ approach described in Equation (12.12). The output vector is thresholded by replacing the activation levels for the output vector elements with the input concept activation levels; that is, activation levels less than −1 are replaced with the value −1 and activation levels greater than +1 are replaced with the value +1. The normalized output vector replaces the input vector and the process of Equation (12.12) is repeated until the output vector is stabilized (CM converges), or until the system becomes cyclic (CM fails). For this problem, an investigation using the $(0+1)$ approach (see Kosko, 1992, or Vasantha and Smarandache, 2003) was performed, and similar results were achieved.

Concepts $C1$, $C3$, and $C7$ were used to normalize (threshold) the output vector elements as the fixed inputs. Given here are the results of the crisp analysis of the conventional CM shown in Figure 12.11.

$$
\begin{array}{c c}
 & \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{bmatrix}
0 & 1 & 0 & -1 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & -1 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & -1 & 0
\end{bmatrix}
\end{array}
$$

Figure 12.12 Adjacency matrix for the CM in Figure 12.11.

The initial input vector [1, 0, 1, 0, 0, 0, 1] represented

- C1—level of UNM electricity demand high,
- C3—PNM sends UNM curtailment signal, and
- C7—occupant dissatisfaction is high.

A stable output was achieved after three iterations and resulted in [1, 1, 1, −1, 0, 1, 1]. This represented, in addition to the input above,

- C2—UNM generates electricity internally at the maximum level,
- C4—UNM stores no energy,
- C5—UNM external (PNM) electric energy usage is moderate because of the Ford Utilities Center cogeneration turbine operating at the maximum level, and
- C6—full DR by UNM to curtail electric energy use even though occupant dissatisfaction was so high.

The next input [1, 0, 1, 0, 0, 0, 0] represented

- C1—level of UNM electricity demand high,
- C3—PNM sends UNM curtailment signal, and
- C7—there is no occupant dissatisfaction.

A stable output was also achieved after three iterations and resulted in [1, 1, 1, −1, 0, 1, 0]. This represented

- C2—UNM generates electricity internally at the maximum level,
- C4—UNM stores no energy,
- C5—UNM external (PNM) electric energy usage is moderate, and
- C6—full DR by UNM to curtail electric energy use as occupant dissatisfaction was nonexistent.

The third input [0, 0, 1, 0, 0, 0, 0] represented

- C1—level of UNM electricity demand low,
- C3—PNM sends UNM curtailment signal, and
- C7—there is no occupant dissatisfaction.

The stable output was achieved after four iterations and resulted in [0, 1, 1, −1, −1, 1, 0]. This represented

- C2—UNM generates electricity internally at the maximum level,
- C4—UNM stores no energy,
- C5—UNM external (PNM) electric energy usage is low as the campus demand is low, and
- C6—full DR by UNM to curtail electric energy use as occupant dissatisfaction was nonexistent.

The fourth input [1, 0, 0, 0, 0, 0, 0] represented

- C1—level of UNM electricity demand high,
- C3—PNM sends UNM no curtailment signal, and
- C7—there is no occupant dissatisfaction.

The stable output was achieved after three iterations and resulted in [1, 1, 0, −1, 0, 1, 0]. This represented

- C2—UNM generates electricity internally at the maximum level,
- C4—UNM stores no energy,
- C5—UNM external (PNM) electric energy usage is low as the campus demand is low, and
- C6—full DR by UNM to curtail electric energy use as occupant dissatisfaction was nonexistent.

The next step in this investigation was to conduct a fuzzy CM by developing linguistic modifiers for the causal weights connecting the concepts. The linguistic modifiers and their weight definitions used in the FCM are as follows:

• Definitely causes	1.0
• Strongly causes	0.8
• Moderately causes	0.5
• Weakly causes	0.2
• Does not cause	0
• Weakly causes the negative of the concept	−0.2
• Moderately causes the negative of the concept	−0.5
• Strongly causes the negative of the concept	−0.8
• Definitely causes the negative of the concept	−1.0

Figure 12.13 shows the FCM with the causal weights labeled, w1 through w15. The subject matter expert assigned the weights to each of these connections, as seen in Table 12.1.

The adjacency matrix for this FCM is shown in Figure 12.14

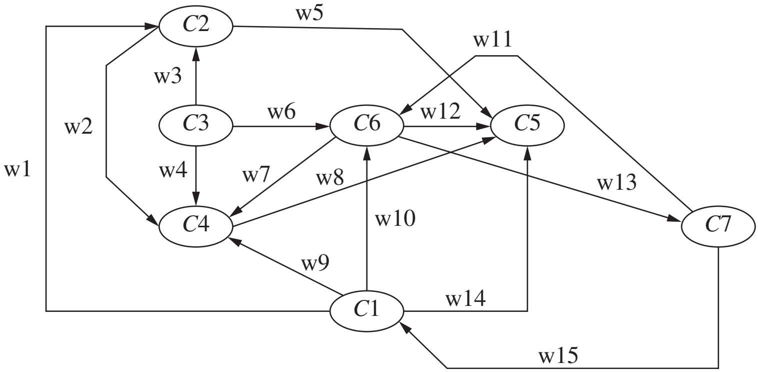

Figure 12.13 Fuzzy CM showing the concepts and causal weights for energy demand.

Table 12.1 Weight assignments.

w1	1.0	w9	–0.5
w2	0.2	w10	0.5
w3	1.0	w11	–0.5
w4	–0.5	w12	–0.8
w5	–0.8	w13	0.5
w6	1.0	w14	0.5
w7	–0.8	w15	0.2
w8	–0.5		

$$
\begin{array}{c|ccccccc|}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\hline
1 & 0 & \mathbf{w1} & 0 & \mathbf{w9} & \mathbf{w14} & \mathbf{w10} & 0 \\
2 & 0 & 0 & 0 & \mathbf{w2} & \mathbf{w5} & 0 & 0 \\
3 & 0 & \mathbf{w3} & 0 & \mathbf{w4} & 0 & \mathbf{w6} & 0 \\
4 & 0 & 0 & 0 & 0 & \mathbf{w8} & 0 & 0 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
6 & 0 & 0 & 0 & \mathbf{w7} & \mathbf{w12} & 0 & \mathbf{w13} \\
7 & \mathbf{w15} & 0 & 0 & 0 & 0 & \mathbf{w11} & 0
\end{array}
$$

Figure 12.14 Adjacency matrix for the FCM in Figure 12.13.

Again, following the basic method of analysis used in the crisp CM, the input vector is multiplied by the adjacency matrix, Figure 12.14, to generate the output vector. And again, the output vector is normalized by replacing the activation levels for the output vector elements with the input concept activation levels, replacing activation levels less than −1 with the value −1, and replacing activation levels greater than +1 with the value +1. This process is repeated until the FCM stabilizes. In this example, the criteria for stabilization because the values are now on the unit interval, was a convergence threshold of 1% (i.e., 0.01). The same four input vectors as used in the crisp CM were also used for the fuzzy CM; the results were as follows:

The initial input [1, 0, 1, 0, 0, 0, 1] represented

- C1—level of UNM electricity demand *high*,
- C3—PNM sends UNM curtailment signal, and
- C7—occupant dissatisfaction is *high*.

The stable output was achieved after three iterations and resulted in [1, 1, 1, −1, −0.6, 1, 1]. This represented, in addition to the input above,

- C2—UNM generates electricity internally at the maximum level,
- C4—UNM stores *no* energy,
- C5—UNM external (PNM) electric energy usage is *moderately low*, and
- C6—*full* DR by UNM to curtail electric energy use even though occupant dissatisfaction was so *high*.

The second input [1, 0, 1, 0, 0, 0, 0] represented

- C1—level of UNM electricity demand *high* and
- C3—PNM sends UNM curtailment signal.

The stable output was also achieved after three iterations and resulted in [1, 1, 1, −1, −0.6, 1, 0.5]. This represented

- C2—UNM generates electricity internally at the *maximum* level,
- C4—UNM stores *no* energy,
- C5—UNM external (PNM) electric energy usage is *moderately low*,
- C6—full DR by UNM to curtail electric energy use as occupant dissatisfaction was *nonexistent*, and
- C7—there is *moderate* occupant dissatisfaction.

The third input [0, 0, 1, 0, 0, 0, 0] represented

- C3—PNM sends UNM curtailment signal only.

The stable output was achieved after 11 iterations and resulted in [0.08, 1.00, 1.00, −1.00, −0.93, 0.83, 0.42]. This represented

- C1—level of UNM electricity demand *light*,
- C2—UNM generates electricity internally at the *maximum* level,

- C4—UNM stores *no* energy,
- C5—UNM external (PNM) electric energy usage is *low* as the campus demand is *low*,
- C6—*nearly full* DR by UNM to curtail electric energy use, and
- C7—there is *moderately low* occupant dissatisfaction.

The forth input [1, 0, 0, 0, 0, 0, 0] represented

- C1—level of UNM electricity demand *high*.

The stable output was achieved after 11 iterations and resulted in [1.00, 1.00, 0, −0.63, −0.31, 0.40, 0.20]. This represented

- C2—UNM generates electricity internally at the *maximum* level,
- C3—PNM sends UNM *no* curtailment signal,
- C4—UNM stores *some* energy,
- C5—UNM external (PNM) electric energy usage is *moderate* as the campus demand is *high* and but the cogeneration turbine is at the *maximum* level,
- C6—*moderate* DR by UNM to curtail electric energy use, and
- C7—there is *minor* occupant dissatisfaction.

From this analysis, it is clear that the use of fuzzy weights for the causal connections in the FCM yields a greater range of outcomes for the concept activation levels even when the input levels are crisp. In fact, the defuzzification of the fuzzy output concepts was simply interpolated from the concepts' linguistic definitions described at the beginning of this example.

Genetically Evolved Fuzzy Cognitive Maps

Decision support modeling using FCMs has been applied to a vast array of problems from technical to political. These include biological processes, electric circuits, control systems, equipment failure modes and effects analysis, social and political situations, and organizational strategic planning. CMs were initially introduced as a means of formalizing and improving the decision process for policy makers. FCMs are a form of neural networks that lend themselves to analysis using neural network techniques as well as genetic algorithms (GAs). A relatively new development from investigators at the University of Cyprus (see Mateou and Andreou, 2008) has pioneered the development of decision support modeling using (GAs) in the evaluation of FCMs. Their approach, termed a *genetically evolved fuzzy cognitive map* (GEFCM), optimizes the potential solutions of an FCM thereby giving the decision maker some choices among the various near-optimal solutions.

In a GEFCM, a GA is used to calculate optimum causal connection weights in the FCM based on desired activation levels of one or more of the FCM concepts. These desired activation levels of one or more concepts act as constraints on the FCM that come from political, social, physical, or economic requirements. The desired activation levels establish the goal for the analysis. The GEFCM then calculates the activation levels of the rest of the concepts in the FCM given the desired activation levels. Through the defuzzification process, the meaning of the analysis is developed. This is a powerful approach because it adds an evolutionary advantage to FCMs. Now, not only can an FCM provide for a reasonable approach to planning and decision making, but also the GEFCM now makes it possible to achieve optimum strategies involving resource allocation for typical problems in decision making, policy, and planning.

Example 12.5

Lincoln and Ross (2012) have adapted a GEFCM to evaluate the policies of UNM as applied to the campus energy demand response. Here, a GEFCM was used to determine an optimal energy policy within the constraints of public utility and on-campus energy providers. The GEFCM was applied to the simplified model shown in Figure 12.13, as well as a more complicated map that addressed 20 subjective concepts (see Lincoln and Ross, 2012 for details). In this work, the authors started with an initial input vector of the following (C1 = C3 = C7 = 1.0), or

C1—A high level of UNM electricity consumption
C3—PNM sends a UNM curtailment signal
C7—Occupant dissatisfaction is high

The output of the GEFCM produced the following results for the other four concepts:

C2—UNM generates very little electricity internally (activation level = −0.86)
C4—UNM stores very little energy (activation level = −0.65)
C5—UNM external (PNM) electric energy use is low due to the release of stored energy and the energy reduction from the demand response event (activation level = −0.79)
C6—UNM curtails electric energy use to support demand response (activiation level = +0.94)

Fuzzy cognitive maps allow us to model the human and political reactions to electric power reduction that occur in a demand response event and estimate the impact these reactions can have on the event. The results found in Lincoln and Ross (2012) show that facility occupancy levels have a strong impact on the level of human dissatisfaction and therefore on the imposition of limits by the university administration on the amount of power reduction allowed during the demand response event. This analysis suggests that the organizers of the demand response event will achieve greater success if occupancy level is understood and considered in planning the demand response event. It also implies that once dissatisfaction emerges, that dissatisfaction tends to drive a strong political intervention that would be absent with a low dissatisfaction level. In other words, if the dissatisfaction level is kept low, the demand response event will not be impacted negatively and power reduction goals will likely be met. The analysis indicates that high occupancy facilities should receive less power reduction than those with moderate occupancy levels and low occupancy facilities will tolerate the highest power reduction.

An interesting idea in the use of the GEFCM for predicting electricity demand response would be to modify the algorithms to use time dependent functions as the desired activation levels, that is, $C_i(t)$. This would allow analyzing the dynamic effects in non-steady state conditions. As an example, assume the wholesale price of electricity, C1, varies in time according to Figure 12.15.

From the information in Figure 12.15, the fuzzy concept activation levels for C1 could be defined as follows:

- The very high pricing level is \$500 per MWH with an activation level of +1.
- The moderate pricing level is \$150 per MWH with an activation level of 0, and
- The low pricing level is \$25 per MWH with an activation level of −1.

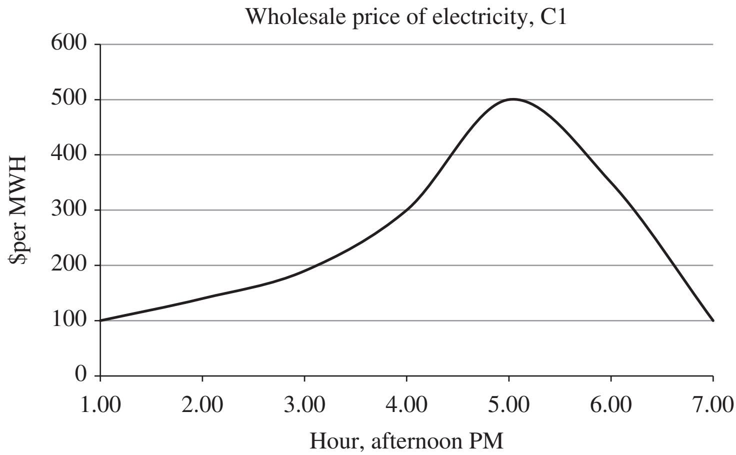

Figure 12.15 Time carrying wholesale price of electricity, Concept C1.

This would result in the activation level, C1, for the wholesale price of electricity varying with time, and a GEFCM could be run for the various times of the day, resulting in a cumulative time dependent response.

To conclude this section fuzzy cognitive mapping does suffer somewhat in comparison with other digraph methods in that there is a large degree of subjectivity. But FCM does allow for varying degrees of magnitude or significance of relationships, which is a limitation of other standard methods. Therefore, much of the grayness in subjectivity is captured and accounted for, resulting in a more balanced assessment. Thus, with appropriate expert-based professional judgment (likely by a panel of experts in the field of the electricity demand), FCM can be an effective assessment tool. Fuzzy cognitive maps provide a heuristic approach for analyzing complex situations to improve the decision-making process. The genetically evolved method for FCM analysis adds additional dimensions to this analysis. It allows the analyst to fix the activation level of one or more concepts in the FCM and resolve the activation levels of the rest of the concepts in the map. The solutions to the FCM appear in families or groups and are identified using clustering techniques.

Agent-Based Models

Agent-based modeling (ABM) provides a methodology to model emergent macrostructure patterns resulting from complex microstructure interactions within a system (Macy and Flache, 2002). In other words, the large-scale patterns of a complex system can be modeled by assigning interaction rules to a system's individual components. This is considered a bottom-up method as the component interactions are specified, providing a natural description and realistic representation of a complex system (Bonabeau, 2002).

The origins of ABM can likely be traced to Craig Reynold's attempt in 1987 to mimic the flocking behavior of birds. Reynolds, a software engineer, wrote a computer program to model flocking behavior under the assumption that local perception was the key element that allowed a group of birds to remain in a coherent flock (Ball, 2004). The program consisted of individual agents called *boids*, whose movements were governed by their separation, alignment, and cohesion to nearby "boids," as well as separation to other objects in the environment (Macy and Flache, 2002). Through these simple rules governing the component interactions of the system, the macrostructure, or flocking, of the system could be visualized. The program worked so well that it was picked up for use in simulating flocking and herding behavior in computer animation for movies such as *Batman Returns* and *The Lion King* (Ball, 2004). The key breakthrough that Reynold's boids have provided in modeling complex systems is that the macrostructures observed in complex systems are not governed by system-wide rules, but emerge from rules governing the interaction and behavior of individual components (Macy and Flache, 2002).

ABM is a methodology that allows a researcher to model complex systems using a natural description (Bonabeau, 2002) that is particularly useful in situations where the details of a system can be accurately characterized only by a model as complex as the system itself (Bankes, 2002). ABM presents a unique approach compared to the standard deductive and inductive research methods (Axelrod, 2003). Of course, ABM is not a total departure from deductive and inductive reasoning, and actually uses aspects of both. ABM is similar to deductive reasoning in formulating explicit rules, or axioms, governing agent interactions and behaviors, deviating from deductive reasoning after this step as theorems are not deduced from these rules. Aspects of inductive reasoning are used in the final steps of ABM to provide insights into the model results (Axelrod, 2003). It is in this way that ABM provides a unique approach to understanding the rules governing complex systems using aspects of both deductive and inductive reasoning. In the end, it is the use of inductive reasoning which gives a deep understanding of the complex system (Chapter 1) by inferring how microstructure interactions and component behavior lead to macrostructure patterns. This deep understanding due to the use of inductive reasoning can be likened to the level of understanding necessary to play chess well, while deductive, or shallow, reasoning provides a level of reasoning suited to playing checkers well.

One of the most significant improvements provided by ABM is the ability to model irrational agents (Ball, 2004; Axelrod, 2003; Macy and Willer, 2002) with the ability to adapt as the simulation progresses (Bonabeau, 2002; Axelrod, 2003). It is precisely this ability that has allowed ABM to realistically model phenomena, such as market fluctuations and growth of firms, in a manner previously impossible with existing methods based on the assumption that agents are "rational maximizers" (Ball, 2004). The "rational maximizer" should be recognized as a theoretical limiting case of social interactions with a negligible probability of occurrence. The concept of a "rational maximizer" is analogous to the concept of an ideal gas, or frictionless motion, and as such is only relevant as a simplification of reality used for demonstrating first principles (Axtell, 2000).

Irrational decision making is often attributed to an individual's lack of knowledge concerning the consequences of their decisions. Although irrational behavior has commonly been introduced into ABMs as noise affecting an agent's interaction and decision rules randomly, this behavior could be included in an ABM more naturally by explicitly modeling this irrationality using principles of fuzzy logic. Fuzzy logic provides a theoretical construct for the explicit inclusion of the uncertainty associated with irrational decision making due to imprecise information concerning the consequences of an agent's decisions.

Example 12.6

This example explores the application of fuzzy logic to ABM through the implementation of fuzzy rules to the Recursive Porous Agent Simulation Toolkit (REPAST)'s model *SugarScape*, which is a partial implementation of the SugarScape model from Epstein and Axtell (1996). In this model, a population of agents collects sugar from a sugar field. The agents are assigned a metabolism from a uniform distribution that governs how much sugar they must consume each time step to survive. In the sugar field, the maximum amount of sugar at a given location is constant. Once the amount of sugar at a given location has been reduced below its maximum value due to an agent collecting some or all of it, the sugar will begin to regenerate according to a specified growth rate. At each time step, each agent decides where to move within his vision, which is assigned to each individual agent at their inception from a uniform distribution that ranges from one space to a user-defined maximum number of spaces, allowing the agent to consider a specified number of spaces in the negative and positive x and y directions. In the standard SugarScape model, the best spot to move to is decided by identifying the space with the greatest sugar at the nearest location. This is considered the crisp SugarScape model here as it is governed by crisp rules. In the fuzzy SugarScape model developed by Harp (2009), the best spot to move to is decided by identifying the spot where the agent has the greatest desire to move determined by a fuzzy rule-base where the amount of sugar and the distance are the inputs to the rule-base and the desire is the output. This gives the agents the ability to reason in a way that is more similar to human reasoning, where the factors used to make a decision may interact in a complicated fashion. This differs from the crisp SugarScape model, where the decision process is in a hierarchical structure, where the amount of sugar takes precedence over the distance.

Crisp Sugar Agents. As mentioned, the crisp Sugar Agents make their decision on the best spot to move to within their vision based on a hierarchical structure. This structure can be visualized in outline form using the following pseudocode:

```
IF sugar at space is greater than the best space so far
   Make this space the best space
ELSE if sugar at space is equal to best space so far
   IF distance to space is less than distance to best space so far
      Make this space the best space
   ELSE if distance to space is equal to distance to best space so far
      Include this space and previous best space or spaces as possible best spaces
```

If multiple best spaces have been collected, the space to move to is selected randomly from among these spaces. This hierarchical structure places the emphasis of the decision on the amount of sugar at the space, where distance is only considered after it has been determined that the sugar is equal to the sugar at the previous best spot. These crisp rules produce "rational maximizers," which does not allow for irrational decisions.

Fuzzy Sugar Agents. The best way to fuzzify the example is to build fuzzy rules. The fuzzy rules consist of a fuzzy inference system (FIS) that maps the amount of sugar to a value of sugar desire and an FIS that maps the distance to a value of distance desire. These fuzzy rules can be expressed linguistically as IF–THEN rules. The crisp input has to be fuzzified into one or more fuzzy sets, which are mapped to one or more corresponding output fuzzy sets.

The output fuzzy sets are defuzzified to a crisp value of desire using centroid defuzzification. The sugar desire and distance desire are combined with a user-defined weighting factor (fuzzy weight) to obtain the total desire the agent has to move to a given space. The fuzzy rules are as follows:

IF sugar is low THEN sugar desire is low
IF sugar is medium THEN sugar desire is medium
IF sugar is high THEN sugar desire is high
IF distance is near THEN distance desire is high
IF distance is medium THEN distance desire is medium
IF distance is far THEN distance desire is low

And these rules and associated membership functions are shown for the *sugar* FIS and the *distance* FIS in a graphical way in Figure 12.16. The resulting values for sugar and distance desire are combined to determine the total desire as

$$wf * \text{sugar desire} + (1 - wf) * \text{distance desire} = \text{total desire},$$

where *wf* is the fuzzy weight controlling the relative dependence of the sugar and distance desires on the total desire. A fuzzy weight of 1 would provide total dependence on sugar and 0 would provide total dependence on distance.

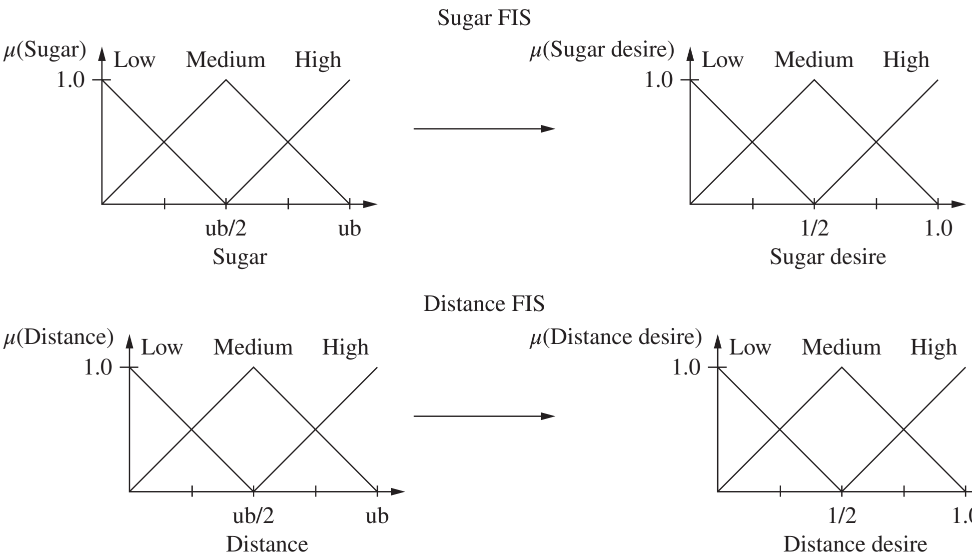

Figure 12.16 Fuzzy SugarScape rule-base where *ub* is the upper bound for the input domains, corresponding to the largest value of sugar within the agent's vision in the sugar FIS and the value of the agent's vision in the distance FIS.

In practice, the fuzzy rule-base would have to be much more complex than what is shown here to allow a full representation of an agent's reasoning. This simplistic fuzzy rule-base has been used here to allow the effects of fuzzy logic on an ABM to be explored while keeping the complexity of the ABM to a minimum. The fuzzy rule-base presented here could be extended to produce more sophisticated agents for a complex and realistic ABM.

A thorough simulation of the fuzzy ABM SugarScape model was performed. The simulation modeled two issues: (1) the effects of limiting the resource by varying the growth rate of sugar and (2) the agent metabolism. The output of the model had two metrics: (1) the "wealth" of each agent, that is, the average amount of sugar controlled by an agent and (2) the average distance traveled by all agents to obtain sugar in a given time step. In addition, to evaluate the capacity of agents to deal with imprecision in their perceptions, the simulation introduced, in a random way, bias to the agents perceptions of the amount of sugar in a particular location. This allowed for an evaluation of the use of fuzzy rules compared with crisp rules in the ABM where perception was imprecise. The simulation showed that the average step distance for the fuzzy model was unaffected by the introduced imprecision, but the crisp model showed a significant increase in step distance when imprecision in perceptions was introduced. It was also found that as an agent's dependence on sugar was decreased the ability of the fuzzy sugar agents to survive was reduced, conceivably because of an increase in irrational decisions. Hence, fuzzy logic can be used to improve the decision-making process used by agents, providing irrational decision-making capability and a rule-base that mimics human reasoning. Even in a simplistic simulation, fuzzy logic can be used to provide flexibility in agent decision making. The results of the simulation provide some indication that fuzzy logic has the ability to realistically introduce and control the level of irrationality in agent decision making. The inclusion of irrationality in a realistic manner is important to enhancing ABM as the idea of the "rational maximizer," often used to model society is a limiting case rarely if ever seen in reality. Although the results presented are elementary and limited, they do indicate that ABM can be enhanced and enriched through the use of fuzzy logic.

Fuzzy Arithmetic and the Extension Principle

> *Said the Mock Turtle with a sigh, "I only took the regular course." "What was that?" inquired Alice. "Reeling and Writhing, of course, to begin with," the Mock Turtle replied; "and the different branches of Arithmetic – Ambition, Distraction, Uglification, and Derision."*
>
> Lewis Carroll, *Alice in Wonderland*, 1865

As Lewis Carroll so cleverly implied as early as 1865 (he was, by the way, a brilliant mathematician), there possibly could be other elements of arithmetic; consider those of ambition, distraction, uglification, and derision. Certainly, fuzzy logic has been described in worse terms by many people over the last four decades! Perhaps Carroll had a presage of fuzzy set theory exactly 100 years before Zadeh; perhaps, possibly.

In this section we see that standard arithmetic and algebraic operations, which are based after all on the foundations of classical set theory, can be extended to fuzzy arithmetic and fuzzy algebraic operations. This extension is accomplished with Zadeh's extension principle (Zadeh, 1975). Fuzzy numbers, briefly described in Chapter 4, are used here because such numbers are the basis for fuzzy arithmetic. In this context the arithmetic operations are not fuzzy; the numbers on which the operations are performed are fuzzy and, hence, so too are the results of these operations. This section provides only an overview of this subject, as it is treated in more detail in the many texts in the literature (see Ross, 2010, for an introduction and references).

Extension Principle

In engineering, mathematics, and the sciences, functions are ubiquitous elements in modeling. Consider a simple relationship between one independent variable and one dependent variable as shown in Figure 12.17. This relationship is a single-input, single-output process where the transfer function (the box in Figure 12.17) represents the mapping provided by the general function f. In the typical case, f is of analytic form, for example, $y = f(x)$, the input, x, is deterministic, and the resulting output, y, is also deterministic.

How can we extend this mapping to the case where the input, x, is a fuzzy variable or a fuzzy set, and where the function itself could be fuzzy? That is, how can we determine the fuzziness in the output, y, based on either a fuzzy input or a fuzzy function or both (mapping)? An extension principle developed by Zadeh (1975) and later elaborated by Yager (1986) enables us to extend the domain of a function on fuzzy sets.

The material in this section is just a summary of the extension principle and how it is useful in fuzzy arithmetic. Additional examples to illustrate the power of the extension principle, especially in the area of fuzzy arithmetic, is provided in (Ross, 2010).

Crisp Functions, Mapping, and Relations

Functions (also called *transforms*), such as the logarithmic function, $y = \log(x)$, or the linear function $y = ax + b$, are mappings from one universe, X, to another universe, Y. Symbolically, this mapping (function, f) is sometimes denoted f: $\mathrm{X} \rightarrow \mathrm{Y}$. Other terminology calls the mapping $y = f(x)$ the *image of x under f* and the inverse mapping, $x = f^{-1}(y)$, is termed the *original image of y*. A mapping can also be expressed by a relation R (as described in Chapter 3), on the Cartesian space $\mathrm{X} \times \mathrm{Y}$. Such a relation (crisp) can be described symbolically as $\mathrm{R} = (x, y)|y = f(x)$, with the characteristic function describing the membership of specific x, y pairs to the relation R as

$$\chi_{\mathrm{R}}(x, y) = \begin{cases} 1, & y = f(x), \\ 0, & y \neq f(x). \end{cases} \tag{12.13}$$

Now, because we can define transform functions, or mappings, for specific elements of one universe (x) to specific elements of another universe (y), we can also do the same thing for

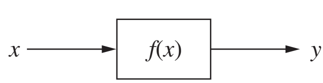

Figure 12.17 A simple single-input, single-output mapping (function).

collections of elements in X mapped to collections of elements in Y. Such collections have been referred to in this text as *sets*. Presumably, then, all possible sets in the power set of X can be mapped in some fashion (there may be null mapping for many of the combinations) to the sets in the power set of Y, that is, $f: \mathrm{P(X)} \rightarrow \mathrm{P(Y)}$. For a set A defined on universe X, its image, set B on the universe Y, is found from the mapping, $\mathrm{B} = f(\mathrm{A}) = y|$ for all $x \in \mathrm{A}$, $y = f(x)$, where B will be defined by its characteristic value

$$\chi_{\mathrm{B}}(y) = \chi_{f(\mathrm{A})}(y) = \bigvee_{y=f(x)} \chi_{\mathrm{A}}(x). \tag{12.14}$$

Example 12.7

Suppose we have a crisp set A using Zadeh's notation,

$$\mathrm{A} = \left\{ \frac{0}{-2} + \frac{0}{-1} + \frac{1}{0} + \frac{1}{1} + \frac{0}{2} \right\}$$

defined on the universe $\mathrm{X} = \{-2, -1, 0, 1, 2\}$ and a simple mapping $y = |4x| + 2$. We wish to find the resulting crisp set B on an output universe Y using the extension principle. From the mapping, we can see that the universe Y will be $\mathrm{Y} = \{2, 6, 10\}$. The mapping described in Equation (12.14) will yield the following calculations for the membership values of each of the elements in universe Y:

$$\begin{aligned}
\chi_{\mathrm{B}}(2) &= \vee\{\chi_{\mathrm{A}}(0)\} = 1, \\
\chi_{\mathrm{B}}(6) &= \vee\{\chi_{\mathrm{A}}(-1), \chi_{\mathrm{A}}(1)\} = \vee\{0, 1\} = 1, \\
\chi_{\mathrm{B}}(10) &= \vee\{\chi_{\mathrm{A}}(-2), \chi_{\mathrm{A}}(2)\} = \vee\{0, 0\} = 0.
\end{aligned}$$

Notice that there is only one way to get the element 2 in the universe Y, but there are two ways to get the elements 6 and 10 in Y. Written in Zadeh's notation this mapping results in the output

$$\mathrm{B} = \left\{ \frac{1}{2} + \frac{1}{6} + \frac{0}{10} \right\};$$

or, alternatively, $\mathrm{B} = \{2, 6\}$.

Suppose we want to find the image B on universe Y using a relation that expresses the mapping. This transform can be accomplished by using the composition operation described in Chapter 3 for finite universe relations, where the mapping $y = f(x)$ is a general relation. Again, for $\mathrm{X} = \{-2, -1, 0, 1, 2\}$ and a generalized universe $\mathrm{Y} = \{0, 1, \ldots, 9, 10\}$, the crisp relation describing this mapping ($y = |4x| + 2$) is

$$\mathrm{R} = \begin{array}{c} \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

The image B can be found through composition (because X and Y are finite): that is, $B = A \circ R$ (we note here that any set, say A, can be regarded as a one-dimensional relation), where, again using Zadeh's notation,

$$A = \left\{ \frac{0}{-2} + \frac{0}{-1} + \frac{1}{0} + \frac{1}{1} + \frac{0}{2} \right\}$$

and B is found by means of Equation (3.9) to be

$$\chi_B(y) = \bigvee_{x \in X} (\chi_A(x) \wedge \chi_R(x, y)) = \begin{cases} 1, & \text{for } y = 2, 6, \\ 0, & \text{otherwise,} \end{cases}$$

or in Zadeh's notation on Y,

$$B = \left\{ \frac{0}{1} + \frac{0}{1} + \frac{1}{2} + \frac{0}{3} + \frac{0}{4} + \frac{0}{5} + \frac{1}{6} + \frac{0}{7} + \frac{0}{8} + \frac{0}{9} + \frac{0}{10} \right\}.$$

Functions of Fuzzy Sets: Extension Principle

Again, we start with two universes of discourse, X and Y, and a functional transform (mapping) of the form $y = f(x)$. Now, suppose that we have a collection of elements in universe x that form a fuzzy set $\underset{\sim}{A}$. What is the image of fuzzy set $\underset{\sim}{A}$ on X under the mapping f? This image will also be fuzzy, say we denote it fuzzy set $\underset{\sim}{B}$ and it will be found through the same mapping, that is, $\underset{\sim}{B} = f(\underset{\sim}{A})$.

The membership functions describing $\underset{\sim}{A}$ and $\underset{\sim}{B}$ will now be defined on the universe of a unit interval [0, 1], and for the fuzzy case Equation (12.14) becomes

$$\mu_{\underset{\sim}{B}}(y) = \bigvee_{f(x) = y} \mu_{\underset{\sim}{A}}(x). \tag{12.15}$$

More generally, suppose our input universe comprises the Cartesian product of many universes. Then, the mapping f is defined on the power sets of this Cartesian input space and the output space, or

$$f : P(X_1 \times X_2 \times \cdots \times X_n) \to P(Y). \tag{12.16}$$

Let fuzzy sets $\underset{\sim}{A}_1, \underset{\sim}{A}_2, \ldots, \underset{\sim}{A}_n$ be defined on the universes X_1, X_2, …, X_n. The mapping for these particular input sets can now be defined as $\underset{\sim}{B} = f(\underset{\sim}{A}_1, \underset{\sim}{A}_2, \ldots, \underset{\sim}{A}_n)$, where the membership function of the image $\underset{\sim}{B}$ is given by

$$\mu_{\underset{\sim}{B}}(y) = \max_{y = f(x_1, x_2, \ldots, x_n)} \left\{ \min\left[\mu_{\underset{\sim}{A}_1}(x_1), \mu_{\underset{\sim}{A}_2}(x_2), \ldots, \mu_{\underset{\sim}{A}_n}(x_n) \right] \right\}. \tag{12.17}$$

In the literature Equation (12.17) is generally called Zadeh's *extension principle*. Equation (12.17) is expressed for a discrete-valued function, *f*. If the function, *f*, is a continuous-valued expression, the max operator is replaced by the sup (suprenum) operator (the suprenum is the least upper bound).

Heretofore, we have discussed features of fuzzy sets on certain universes of discourse. Suppose there is a mapping between elements, *u*, of one universe, U, onto elements, *v*, of another universe, V, through a function *f*. Let this mapping be described by $f: u \rightarrow v$. Define $\underset{\sim}{A}$ to be a fuzzy set on universe U; that is, $\underset{\sim}{A} \subset U$. This relation is described by the membership function

$$\underset{\sim}{A} = \left\{ \frac{\mu_1}{u_1} + \frac{\mu_2}{u_2} + \ldots + \frac{\mu_n}{u_n} \right\}. \tag{12.18}$$

Then, the extension principle, as manifested in Equation (12.15), asserts that, for a function *f* that performs a one-to-one mapping (i.e., maps one element in universe U to one element in universe V), an obvious consequence of Equation (12.15) is

$$\begin{aligned} f\left(\underset{\sim}{A}\right) &= f\left(\frac{\mu_1}{u_1} + \frac{\mu_2}{u_2} + \cdots + \frac{\mu_n}{u_n}\right). \\ &= \left\{ \frac{\mu_1}{f(u_1)} + \frac{\mu_2}{f(u_2)} + \cdots + \frac{\mu_n}{f(u_n)} \right\}. \end{aligned} \tag{12.19}$$

The mapping in Equation (12.19) is said to be *one-to-one*.

Example 12.8

Let a fuzzy set $\underset{\sim}{A}$ be defined on the universe $U = \{1, 2, 3\}$. We wish to map elements of this fuzzy set to another universe, V, under the function

$$v = f(u) = 2u - 1.$$

We see that the elements of V are $V = \{1, 3, 5\}$. Suppose the fuzzy set $\underset{\sim}{A}$ is given as

$$\underset{\sim}{A} = \left\{ \frac{0.6}{1} + \frac{1}{2} + \frac{0.8}{3} \right\}.$$

Then, the fuzzy membership function for $v = f(u) = 2u - 1$ would be

$$f\left(\underset{\sim}{A}\right) = \left\{ \frac{0.6}{1} + \frac{1}{2} + \frac{0.8}{5} \right\}.$$

For cases where this functional mapping *f* maps products of elements from two universes, say U_1 and U_2, to another universe V, and we define $\underset{\sim}{A}$ as a fuzzy set on the Cartesian space $U_1 \times U_2$, then

$$f\left(\underset{\sim}{A}\right) = \left\{ \sum \frac{\min[\mu_1(i), \mu_2(j)]}{f(i,j)} \middle| i \in U_1, j \in U_2 \right\}, \tag{12.20}$$

where $\mu_1(i)$ and $\mu_2(j)$ are the separable membership projections of $\mu(i, j)$ from the Cartesian space $U_1 \times U_2$ when $\mu(i, j)$ cannot be determined. This projection involves the invocation of a condition known as *noninteraction* between the separate universes. It is analogous to the assumption of independence employed in probability theory, which reduces a joint probability density function to the product of its separate marginal density functions. In the fuzzy noninteraction case, we are doing a kind of intersection; hence, we use the minimum operator (some logics use operators other than the minimum operator) as opposed to the product operator used in probability theory.

The complexity of the extension principle increases when we consider more than one of the combinations of the input variables, U_1 and U_2, mapped to the same variable in the output space, V, that is, the mapping is not one-to-one. In this case, we take the maximum membership grades of the combinations mapping to the same output variable, or, for the following mapping, we get

$$\mu_{\underset{\sim}{A}}(u_1, u_2) = \max_{v=f(u_1,u_2)} [\min\{\mu_1(u_1), \mu_2(u_2)\}]. \tag{12.21}$$

Example 12.9

We have two fuzzy sets $\underset{\sim}{A}$ and $\underset{\sim}{B}$, each defined on its own universe as follows

$$\underset{\sim}{A} = \left\{\frac{0.2}{1} + \frac{1}{2} + \frac{0.7}{4}\right\} \text{ and } \underset{\sim}{B} = \left\{\frac{0.5}{1} + \frac{1}{2}\right\}.$$

We wish to determine the membership values for the algebraic product mapping

$$\begin{aligned}
f(\underset{\sim}{A}, \underset{\sim}{B}) &= \underset{\sim}{A} \times \underset{\sim}{B} (\text{arithmetic product}) \\
&= \left\{ \min\left(\frac{0.2, 0.5}{1} + \frac{\max[\min(0.2, 1), \min(0.5, 1)]}{2}\right.\right. \\
&\qquad \left.\left. + \frac{\max[\min(0.7, 0.5), \min(1, 1)]}{4} + \frac{\min(0.7, 1)}{8}\right.\right\} \\
&= \left\{\frac{0.2}{1} + \frac{0.5}{2} + \frac{1}{4} + \frac{0.7}{8}\right\}.
\end{aligned}$$

In this case, the mapping involves two ways to produce a 2 (1×2 and 2×1) and two ways to produce a 4 (4×1 and 2×2); hence, the maximum operation expressed in Equation (12.17) is necessary.

The extension principle can also be useful in propagating fuzziness through generalized relations that are discrete mappings of ordered pairs of elements from input universes to ordered pairs of elements in an output universe.

Example 12.10

We want to map ordered pairs from input universes $X_1 = \{a, b\}$ and $X_2 = \{1, 2, 3\}$ to an output universe, $Y = \{x, y, z\}$. The mapping is given by the crisp relation, R,

$$\mathrm{R} = \begin{array}{c} \\ a \\ b \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} x & z & x \\ x & y & z \end{bmatrix} \end{array}.$$

We note that this relation represents a mapping, and it does not contain membership values. We define a fuzzy set $\underset{\sim}{A}$ on universe X_1 and a fuzzy set $\underset{\sim}{B}$ on universe X_2 as

$$\underset{\sim}{A} = \left\{ \frac{0.6}{a} + \frac{1}{b} \right\} \text{ and } \underset{\sim}{B} = \left\{ \frac{0.2}{1} + \frac{0.8}{2} + \frac{0.4}{3} \right\}.$$

We wish to determine the membership function of the output, $\underset{\sim}{C} = f(\underset{\sim}{A}, \underset{\sim}{B})$ whose relational mapping, f, is described by R. This is accomplished with the extension principle, Equation (12.17), as follows:

$$\begin{aligned}
\mu_{\underset{\sim}{C}}(x) &= \max[\min(0.2, 0.6), \min(0.2, 1), \min(0.4, 0.6)] = 0.4, \\
\mu_{\underset{\sim}{C}}(y) &= \max[\min(0.8, 1)] = 0.8, \\
\mu_{\underset{\sim}{C}}(z) &= \max[\min(0.8, 0.6), \min(0.4, 1)] = 0.6.
\end{aligned}$$

Hence,

$$\underset{\sim}{C} = \left\{ \frac{0.4}{x} + \frac{0.8}{y} + \frac{0.6}{z} \right\}.$$

The extension principle is also useful in mapping fuzzy inputs through continuous-valued functions. The process is the same as for a discrete-valued function, but the effort involved in the computations is more rigorous.

Fuzzy Algebra

Example 12.11 (Wong and Ross, 1985).
Suppose we have a nonlinear system given by the harmonic function $\underset{\sim}{x} = \cos(\underset{\sim}{\omega} t)$, where the frequency of excitation, $\underset{\sim}{\omega}$, is a fuzzy variable described by the membership function shown in Figure 12.18a. The output variable, $\underset{\sim}{x}$, will be fuzzy because of the fuzziness provided in the mapping from the input variable, $\underset{\sim}{\omega}$. This function represents a one-to-one mapping in two stages, $\underset{\sim}{\omega} \rightarrow \underset{\sim}{\omega} t \rightarrow \underset{\sim}{x}$. The membership function of $\underset{\sim}{x}$ will be determined through the use of the extension principle, which for this example will take on the following form:

$$\mu_{\underset{\sim}{x}}(x) = \bigvee_{x = \cos(\omega t)} [\mu_{\underset{\sim}{\omega}}(\omega)].$$

To show the development of this expression, we will take several time points, such as $t = 0$, $1, \ldots$. For $t = 0$, all values of $\underset{\sim}{\omega}$ map into a single point in the $\underset{\sim}{\omega} t$ domain, that is, $\underset{\sim}{\omega} t = 0$ and into a

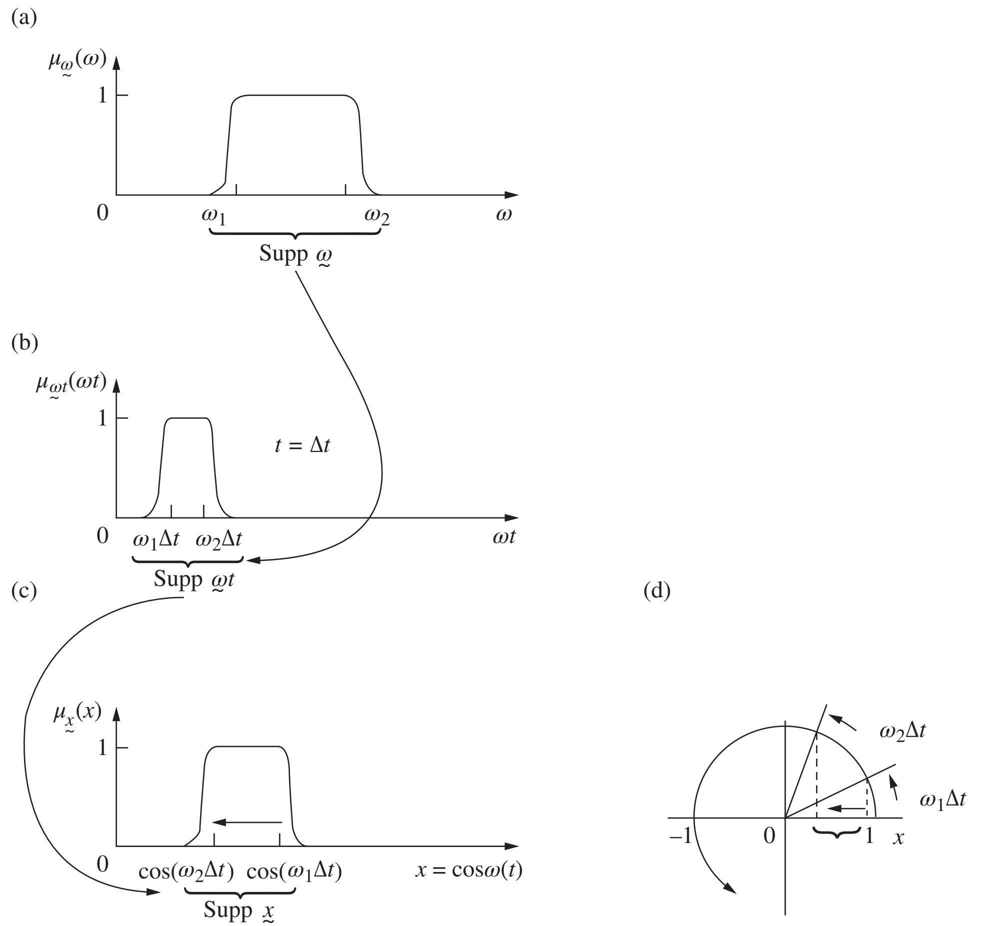

Figure 12.18 Extension principle applied to $\underset{\sim}{x} = \cos(\underset{\sim}{\omega}t)$, at $t = \Delta t$; (a) fuzzy M.F. for $\underset{\sim}{\omega}$, (b) product of $\underset{\sim}{\omega}$ with Δt, (c) fuzzy M.F. for x, (d) phase diagram equivalent of part (c).

single point in the x universe, that is, $x = 1$. Hence, the membership of $\underset{\sim}{x}$ is simply a singleton at $x = 1$, that is,

$$\mu_{\underset{\sim}{x}}(x) = \begin{cases} 1, & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

For a nonzero but small t, say $t = \Delta t$, the support of $\underset{\sim}{\omega}$, denoted in Figure 12.18a as $\underset{\sim}{\omega}$, is mapped into a small but finite support of $\underset{\sim}{x}$, denoted in Figure 12.18c as $\underset{\sim}{x}$ (the support of a fuzzy set is defined in Chapter 4 as the interval corresponding to a λ-cut of $\lambda = 0^+$). The membership value for each $\underset{\sim}{x}$ in this interval is determined directly from the membership of ω in a one-to-one mapping. As can be seen in Figure 12.18, as t increases, the support of $\underset{\sim}{x}$ increases, and the fuzziness in the response spreads with time. Eventually, there will be a value of t when the support of $\underset{\sim}{x}$ folds partly onto itself, that is, we have multi-ω-to-single-x mapping. In this event, the maximum of all candidate membership values of $\underset{\sim}{\omega}$ is used as the membership value of x according to the extension principle, Equation (12.15), as shown in Figure 12.19.

When t is of such magnitude that the support of x occupies the interval [−1, 1] completely, the largest support possible, the membership $\mu_{\underset{\sim}{x}}(x)$ will be unity for all x within this interval. This is the state of complete fuzziness, as illustrated in Figure 12.20. In the equation $\underset{\sim}{x} = \cos(\underset{\sim}{\omega} t)$, the output can have any value in the interval [−1, 1] with equal and complete membership. Once this state is reached, the system remains there for all future time (see similar problem in Chapter 1).

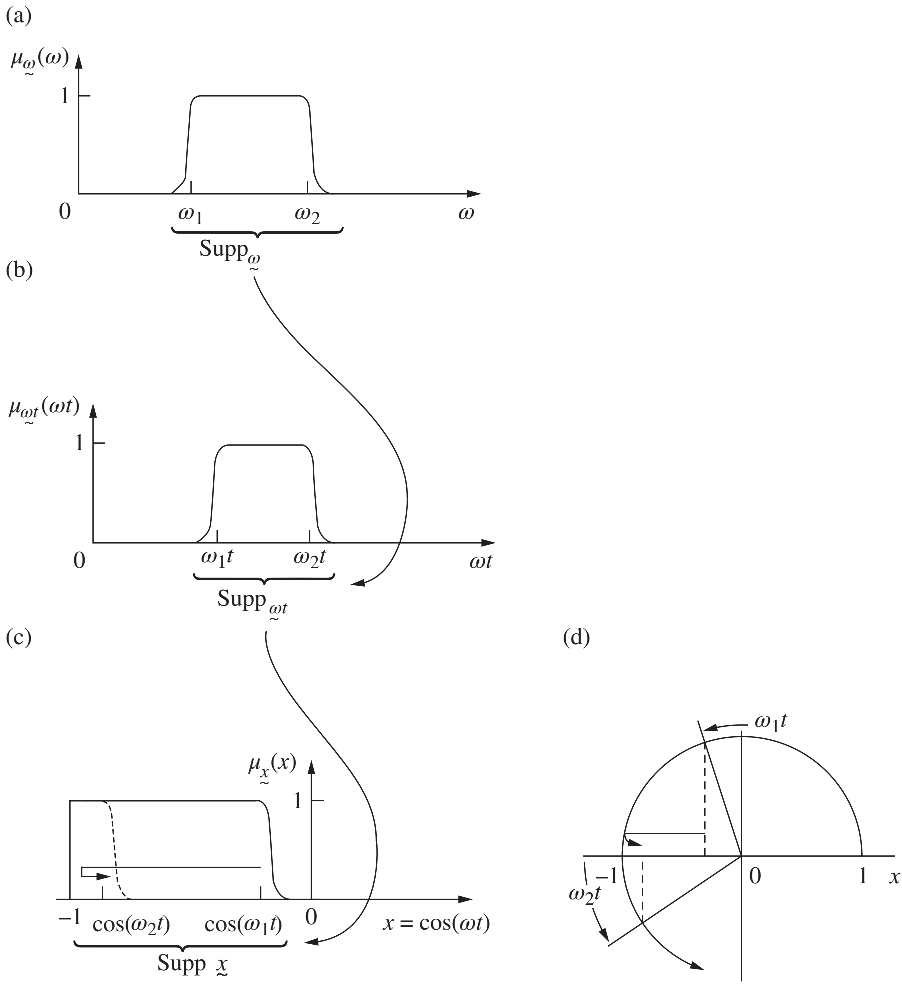

Figure 12.19 Extension principle applied to $\underset{\sim}{x} = \cos(\underset{\sim}{\omega} t)$ showing (a) uncertainty in w, (b) uncertainty in wt, (c) the overlap in the support of x as t increases and (d) phase diagram equivalent of part (c).

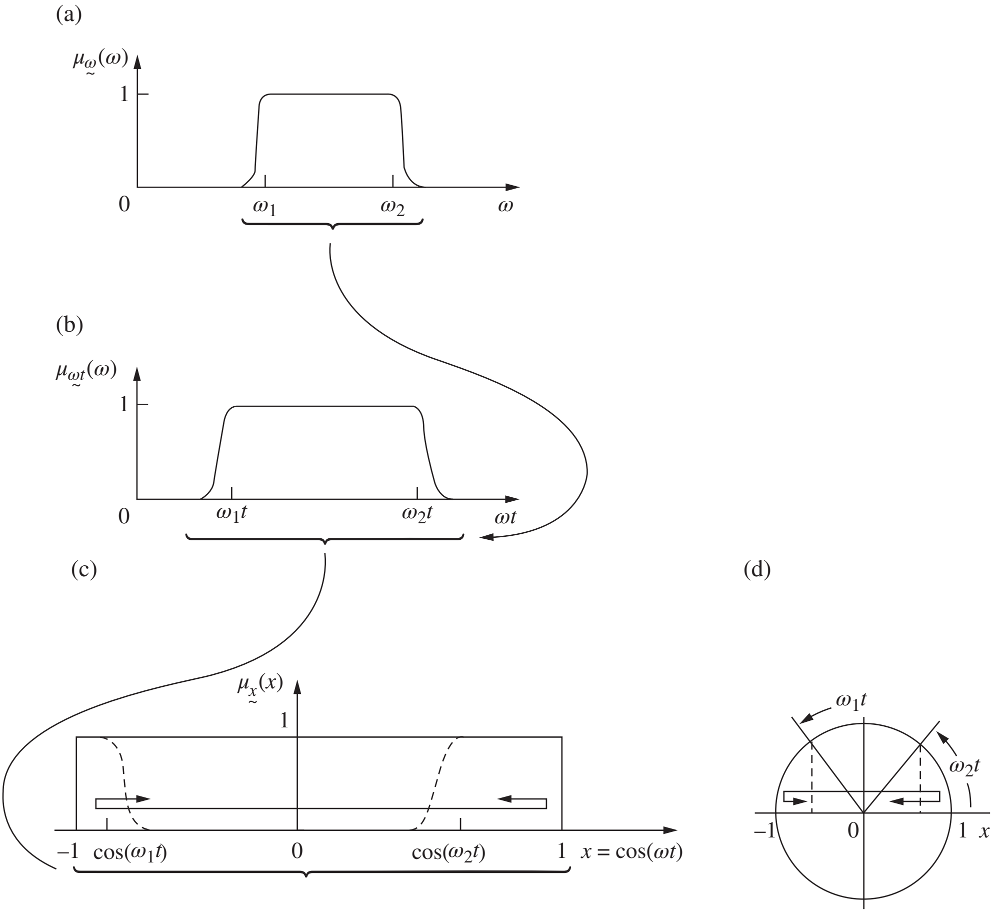

Figure 12.20 Extension principle applied to $\underset{\sim}{x} = \cos(\underset{\sim}{\omega} t)$ when t causes complete fuzziness.

Fuzzy Arithmetic

Chapter 4 defines a fuzzy number as being described by a normal, convex membership function on the real line; fuzzy numbers *usually* also have symmetric membership functions. In this chapter, we wish to use the extension principle to perform algebraic operations on fuzzy numbers (as illustrated in previous examples in this chapter). We define a normal, convex fuzzy set on the real line to be a fuzzy number, and denote it $\underset{\sim}{\mathrm{I}}$.

Let $\underset{\sim}{\mathrm{I}}$ and $\underset{\sim}{\mathrm{J}}$ be two fuzzy numbers, with $\underset{\sim}{\mathrm{I}}$ defined on the real line in universe X and J defined on the real line in universe Y, and let the symbol $*$ denote a general arithmetic operation, that is, $* \equiv \{+, -, \times, \div\}$. An arithmetic operation (mapping) between these two number, denoted $\underset{\sim}{\mathrm{I}} * \underset{\sim}{\mathrm{J}}$, will be defined on universe Z, and can be accomplished using the extension principle, which can take the form:

$$\mu_{\underset{\sim}{\mathrm{I}} * \underset{\sim}{\mathrm{J}}}(z) = \bigvee_{x * y = z} \left(\mu_{\underset{\sim}{\mathrm{I}}}(x) \wedge \mu_{\underset{\sim}{\mathrm{J}}}(y)\right).$$

This operation ($\underset{\sim}{I} * \underset{\sim}{J}$) results in another fuzzy set, the fuzzy number resulting from the arithmetic operation on fuzzy numbers $\underset{\sim}{I}$ and $\underset{\sim}{J}$.

Example 12.12

We want to perform a simple addition ($* \equiv +$) of two fuzzy numbers. Define a fuzzy one by the normal, convex membership function defined on the integers,

$$\underset{\sim}{1} = \left\{ \frac{0.2}{0} + \frac{1}{1} + \frac{0.2}{2} \right\}.$$

Now, we want to add "fuzzy one" plus "fuzzy one," using the extension principle, to get

$$\begin{aligned}
\underset{\sim}{1} + \underset{\sim}{1} = \underset{\sim}{2} &= \left(\frac{0.2}{0} + \frac{1}{1} + \frac{0.2}{2} \right) + \left(\frac{0.2}{0} + \frac{1}{1} + \frac{0.2}{2} \right) \\
&= \left\{ \frac{\min(0.2, 0.2)}{0} + \frac{\max[\min(0.2, 1), \min(1, 0.2)]}{1} \right. \\
&\quad + \frac{\max[\min(0.2, 0.2)\min(1, 1), \min(0.2, 0.2)]}{2} \\
&\quad \left. + \frac{\max[\min(1, 0.2), \min(0.2, 1)]}{3} + \frac{\min(0.2, 0.2)}{4} \right\} \\
&= \left\{ \frac{0.2}{0} + \frac{0.2}{1} + \frac{1}{2} + \frac{0.2}{3} + \frac{0.2}{4} \right\}.
\end{aligned}$$

Note that there are two ways to get the resulting membership value for a 1 (0 + 1 and 1+ 0), three ways to get a 2 (0 + 2, 1 + 1, 2 + 0), and two ways to get a 3 (1 + 2 and 2+ 1). These are accounted for in the implementation of the extension principle.

Data Fusion

Data fusion, or multisensor data fusion is the process of combining or integrating (fusing) measured data originating from different active or passive sensors to produce a more specific, unified dataset or system model about an entity (or object), or event of interest, that has been observed. Data fusion is not simply an additive process because the signals coming from the various sensors are different (e.g., an image is different from a smell). A fusion process is meant to reduce the uncertainty of predicting the state of, or identifying, the observed object. Humans do data fusion constantly. We get information from our fives sensors, our eyes, ears, skin, nose and mouth, and then we "fuse" the data in our brains in a way (Chialvo, 2006) that allows us to reason about a situation. Information from just one of our sensors is not likely to be as accurate as a combination (fusion) of information from all our sensors, and from remembering a similar situation from the past. Our brains make a pattern from past observations, and combine this with the current sensed data to make decisions and assessments. As a simple example, suppose we see a stove burner that is red. We might sense caution, but if we touch the burner we might feel pain. Events like this one, if it happened to us as a child, would modulate our behavior and we might approach the burner in a more cautious way. We desire to make automated systems that can do data fusion in various application domains, such as disciplines like weather forecasting (satellite imaging and barometric sensors), transportation management (radar and voice sensors), and satellite tracking and classification (infrared, optical, electromagnetic, and radar sensors). In all of these domains there is also the presence of "noise" (background signal disturbances like static on radios) in the sensing mechanisms that might be due to weather or other interferences. A human is able to filter out the noise from their senses, in ways we don't understand, but this process is a bit more difficult in automated sensor fusion systems where we have to build directly into the models various schemes to deal with noise. One of the first methods to perform data fusion that deals with filtering noise is the Kalman Filter (KF), first developed by Rudy Kalman for signal noise processing (Kalman, 1960), but lately seeing significant use in many applications in data fusion along with the development of fuzzy KF (see Raol, 2010, and Klein, 2004).

There are developments in data fusion in many different domains. For example the work by Noureldin, El-Shafie, and Reda Taha (2006) in vehicular navigation systems uses a fuzzy-neural module for data fusion. Subramanian, Burks, and Dixon (2009) used a fuzzy KF for autonomous guidance of unmanned vehicles in citrus groves. Yager (1997) developed a general approach for the fusion of imprecise information using a combinability relationship with fuzzy measures, including fuzzy arithmetic. This general structure can be used in any discipline where sensors provide ambiguous and vague information and where the information is associated with different levels of confidence.

There are several models in use today for data fusion: statistical inference, Bayesian inference, Dempster-Shafer evidence theory, artificial neural networks, Boolean voting algorithms, situational awareness, and fuzzy logic approaches (Klein, 2004). And of course, the purpose of data fusion is to provide information to the human, animal, or computer to assess situations, identify dangers or threats, to estimate uncertainties (or confidence), and ultimately to make decisions. The latter method, using a fuzzy Kalman filter, is the subject of the next example.

Kalman Filter in Data Fusion

In the material presented here, the standard KF is illustrated with a simple one-dimensional problem (Esme, 2009). The standard KF is then extended to model the same example using a fuzzy KF. The main reason the KF is so useful in data fusion is that it has intuitively appealing state-space formulation and a predictor-corrector estimation and recursive-filtering structure. KF is a mathematical model-based approach to data fusion that uses the recursive formulation: new estimate of the system = previous estimate + "Kalman gain" times the residuals of the estimation. This is explained.

The KF is a recursive method that involves matrices as the coefficients of the variables; the matrices allow for the modeling of multidimensional states of the system. However, in the simple example here these matrices are reduced to simple scalar values, which are used to model only one state of the system. In particular, in this example a voltage reading is obtained from a single sensor (e.g., a voltmeter; this reading is regarded as a scalar random constant and it is compared to the predicted value of the voltage from the Kalman model). The equations for a one-dimensional problem are much simpler than the full matrix-version equations (see Klein, 2004). In these equations there are two groups: a time update group (sometimes called the prediction) and a measurement update group (sometimes called the *correction*), as detailed in the following example.

Example 12.13

In its simplest form, without all the state matrices the standard KF approach uses the following equation,

$$\hat{x}_k = K_k z_k + (1 - K_k)\hat{x}_{k-1} \tag{12.22}$$

where $\hat{x}$ is the current model estimate of the parameter of interest, K_k is the Kalman gain, z_k is the measured value of the parameter of interest and $\hat{x}_{k-1}$ is the previous estimate of the parameter of interest. Equation 12.22 shows that the value $\hat{x}$ is a weighted sum of the measurement value, z_k, and the previous estimated value, and the Kalman gain is used to determine the weights in this linear combination. Hence, the Kalman gain is an important quantity. If it is large, more weight is placed on the measurement, z_k, and if it is small the most weight is on the previous estimate, $\hat{x}_{k-1}$. Equation 12.22 is recursive, meaning that the values of the process at time step k are dependent on the values from the previous time step k−1.

In a standard KF approach the following equation shows that the signal value of interest at a time step k, x_k, is the linear combination of it previous value at time step k-1, $x_{k\text{-}1}$, a control signal u_k, and a process noise, w_{k-1}.

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \tag{12.23}$$

The second equation, below, indicates that any measurement value, z_k, is a linear combination of the signal value, x_k, and the measurement noise, v_k.

$$z_k = Hx_k + v_k \tag{12.24}$$

The solution process proceeds as follows. There are two groups of equations: a Time Update (prediction) and a Measurement Update (correction). The Time Update equations are:

$$\hat{x}_{\bar{k}} = A\hat{x}_{k-1} + Bu_k \tag{12.25a}$$

$$P_{\bar{k}} = AP_{k-1}A^T + Q \tag{12.25b}$$

where $P_{\bar{k}}$ is the previous error covariance (the error is the difference between the predicted model value at two subsequent time steps). $P_{\bar{k}}$ is a necessary parameter because it is used in determining the Kalman gain, K_k, as will be shown. Q is a “waste basket” for unknown modeling errors, and will not be needed in the example here.

The Measurement Update equations are:

$$K_k = P_{\bar{k}}H^T\left(HP_{\bar{k}}H^T + R\right)^{-1} \tag{12.26a}$$

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k\left(z_k - H\hat{x}_{\bar{k}}\right) \tag{12.26b}$$

$$P_k = (1 - K_k H)P_{\bar{k}} \tag{12.26c}$$

where R is a parameter to model the standard deviation in the experimental noise (the noise in the measurement, z). Equation (12.26b) is found by replacing the measurement noise, v_k, by the equality in Equation (12.24).

The process noise and the measurement noise are presumed to be statistically independent. The matrices A, B, and H are used to model multidimensional problems. Because this example is a simple one-dimensional problem being used for prediction (not for control) the following simplifications are realized.

$$\mathrm{A} = \mathrm{B} = \mathrm{H} = 1,\ Q = 0 \text{ and } u_k = 0,\ w_{k-1} = 0. \tag{12.27}$$

The Time Update (prediction) equations are simplified based on the assumptions of identities in Equation (12.27) to the following. Equations (12.25a) and (12.25b) become

$$\hat{x}_{\bar{k}} = \hat{x}_{k-1} \tag{12.28a}$$

$$P_{\bar{k}} = P_{k-1} \tag{12.28b}$$

The Measurement Update (correction) Equations (12.26) are simplified, as shown as follows, on the assumptions of identities in Equation (12.27); the expression inside the parentheses of Equation (12.26b) is the value of v_k in Equation (12.24).

$$K_k = \frac{P_{\bar{k}}}{P_{\bar{k}} + R} \tag{12.29a}$$

$$\hat{x}_k = \hat{x}_{\bar{k}} + K_k\left(z_k - \hat{x}_{\bar{k}}\right) \tag{12.29b}$$

$$P_k = (1 - K_k)P_{\bar{k}} \tag{12.29c}$$

For the voltage example here, the value of R will be given a constant value of 0.1 volts. The uncertainty in R will be assumed to be Gaussian. The calculations will start with the initial conditions of $P_0 = 1$ and $x_0 = 0$ at time step $k = 1$. The measurement data, from a voltmeter for the time steps ($k = 1, 2, \ldots, 10$) in milliseconds (ms) and the values of the voltmeter readings, in volts (V), are shown in Table 12.2.

Equations (12.28) then become, as shown in the fifth column of Table 12.3 (Time Update),

$$\hat{x}_{\bar{k}} = \hat{x}_{k-1} = 0$$

$$P_{\bar{k}} = P_{k-1} = 1$$

and the Measurement Update values are shown in the sixth column of Table 12.3

Table 12.2 Experimental readings for example problem. Adapted from Esme, 2009.

TIME (ms)	1	2	3	4	5	6	7	8	9	10
VALUE (V)	0.39	0.50	0.48	0.29	0.25	0.32	0.34	0.48	0.41	0.45

Table 12.3 One-parameter KF simulation for a voltage reading. Adapted from Esme, 2009.

k	Z_k	$\hat{x}_{k-1}$	$P_{\bar{k}}$	Time Update	Measurement Update	$\hat{x}_k$	P_k
1	0.39	0	1	$\hat{x}_{\bar{k}} = \hat{x}_{k-1} = 0$ $P_{\bar{k}} = P_{k-1} = 1$	$K_k = 1/(1+0.1) = 0.909$ $\hat{x}_{\bar{k}} = 0 + 0.909(0.390 - 0) = 0.355$ $P_k = (1-.909)(1) = 0.091$	0.355	0.091
2	0.50	0.355	0.091	$\hat{x}_{\bar{k}} = 0.355$ $P_{\bar{k}} = .091$	$K_k = .091/(.091+0.1) = 0.476$ $\hat{x}_{\bar{k}} = 0.355 + 0.476(0.500 - 0.355)$ $= 0.424$ $P_k = (1-.476)(0.091) = 0.048$	0.424	0.048
3	0.48	0.424	0.048			0.442	0.032
4	0.29	0.442	0.032			0.405	0.024
5	0.25	0.405	0.024			0.375	0.02
6	0.32	0.375	0.020			0.366	0.016
7	0.34	0.366	0.016			0.362	0.014
8	0.48	0.362	0.014			0.377	0.012
9	0.41	0.377	0.012			0.38	0.011
10	0.45	0.38	0.011			0.387	0.010

$$K_k = 1/(1+0.1) = 0.909$$

$$\hat{x}_{\bar{k}} = 0 + 0.909(0.390 - 0) = 0.355$$

$$P_k = (1-.909).1 = 0.091$$

These values are also placed in the third and fourth columns for time step k = 2. The measurement update values for k = 2 are given in the calculations that follow,

$$K_k = .091/(.091+0.1) = 0.476$$

$$\hat{x}_{\bar{k}} = 0.355 + 0.476(0.500 - 0.355) = 0.424$$

$$P_k = (1-.476)\,.\,0.091 = 0.048$$

For the remaining time steps (k = 3 to 10) the values are shown in Table 12.3.

The predicted voltages of this simple 10-step simulation are shown in the seventh column of Table 12.3 and also in Figure 12.21. In Figure 12.21 the convergence of the voltage to about 0.38 volts seems obvious. Now, we will see how this same fusion method would work using a fuzzy KF approach.

The expression $\left(z_k - H\hat{x}_{\bar{k}}\right)$ in Equation 12.26b is sometimes referred to as the residual (Raol, 2010), and it is multiplied by the Kalman gain, K_k. The expression $\left(HP_{\bar{k}}H^T + R\right)$ in Equation 12.26a is referred to as the covariance of the residuals. Both of these quantities are expressed in Example 12.13 in their simpler forms of $\left(z_k - \hat{x}_{\bar{k}}\right)$ in Equation (12.29b) and $P_{\bar{k}} + R$ in Equation 12.29a, because H = 1 for the example. The residuals are an important quantity in the KF process because when the residuals gradually decrease the solution of the KF approaches the converged, or true, value of the state variable. The expression $\left(z_k - \hat{x}_{\bar{k}}\right)$ is also equivalent to the measurement noise (see Equation 12.24) and when this noise reaches a

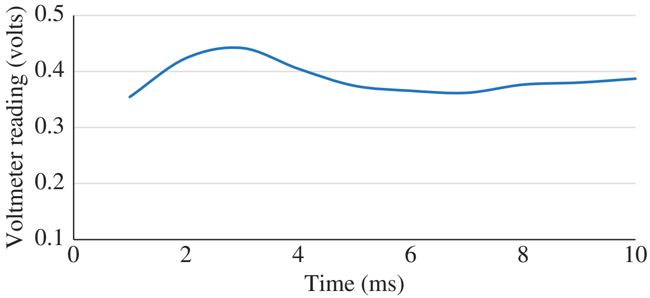

Figure 12.21 KF results converge after 10 time steps for voltmeter example. Adapted from Esme, 2009 with permission.

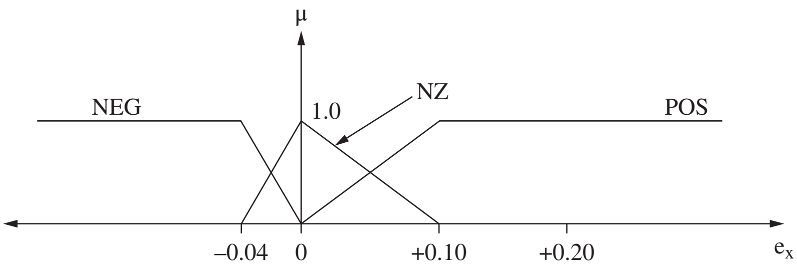

Figure 12.22 Input fuzzy membership functions for e_x.

minimum the solution to the KF has converged. The residuals are the focus of the fuzzy KF illustrated here.

Example 12.14

Fuzzy IF–THEN rules are used to manage the residuals that are important in achieving accurate data fusion results. For the fuzzy KF example we will presume a direct relationship between the differences of successive values (errors) of the parameter of interest, $\hat{x}_k$, and the residuals $\left(z_k - \hat{x}_{\bar{k}}\right)$. So, we define the following variables as the errors input, e_x, and the residuals output, b_x, to be included in the rule base fuzzy KF system:

$$e_x = \hat{x}_k - \hat{x}_{k-1} \tag{12.30}$$

$$b_x = \left(z_k - \hat{x}_{\bar{k}}\right) \tag{12.31}$$

For this system there are three simple rules:

1. IF e_x is NEG, THEN b_x Is NEG
2. IF e_x is NZ, THEN b_x Is POS
3. IF e_x is POS, THEN b_x Is POS

The membership functions for the input errors (Equation 12.30) are shown in Figure 12.22 and the membership functions for the residuals output (b_x) are shown in Figure 12.23.

Table 12.4 Fuzzy inputs (e_x) and resulting defuzzified outputs (b_x) for Example 12.14.

e_x	b_x
−0.04	−0.198
−0.03	−0.1
−0.02	0
−0.01	0.1
0	0.198
0.03	0.185
0.06	0.181
0.1	0.198
0.3	0.198

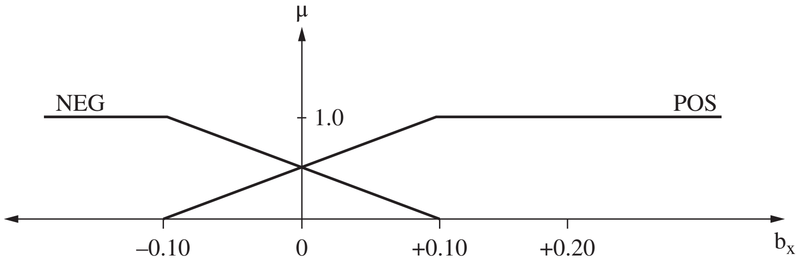

Figure 12.23 Output fuzzy membership functions for b_x.

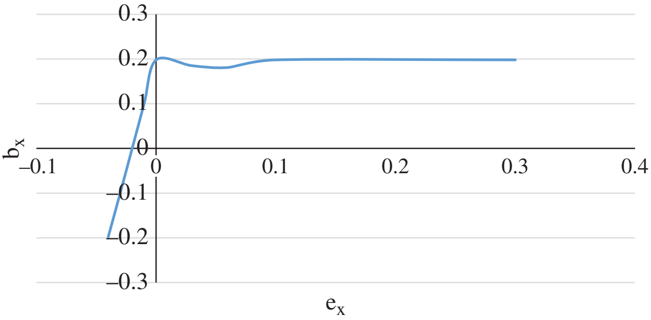

Figure 12.24 Plot of input errors (e_x) vs. output residuals (b_x) for Example 12.14 using a fuzzy system approach.

A simulation, using the MATLAB fuzzy toolbox, and a centroidal defuzzification, for the nine specific values for the inputs (e_x) listed in the first column of Table 12.4, produces the resulting residuals (b_x) that are shown in the second column of Table 12.4, and also in Figure 12.24. The values of the residuals from the fuzzy simulation could then be used to replace the quantity, $\left(z_k - \hat{x}_{\bar{k}}\right)$, in Equation 12.29b and the analysis can be continued as in the previous Example 12.13.

Summary

This chapter summarizes fuzzy logic applications in the areas of optimization, cognitive mapping, fuzzy arithmetic using the extension principle, data fusion, and agent-based models. This begins to scratch only the surface of the plethora of applications being developed in the rapidly expanding field of fuzzy logic. The reader is referred to the literature for many other applications projects, which are summarized in works or collected bibliographies such as Schmucker (1984), Klir and Folger (1988), McNeill and Freiberger (1993), Kosko (1993), Dubois and Prade (1980), Ross, Booker, and Parkinson (2003), and Cox (1994) or discussed in some of the active international research journals focusing on fuzzy applications such as *Intelligent and Fuzzy Systems*, *Fuzzy Sets and Systems*, and *IEEE Transactions on Fuzzy Systems*. In a previous edition of this book (Ross, 2004) additional areas of system identification, linear regression, and rule-base reductions were also covered.

The extension principle is one of the most basic ideas in fuzzy set theory. It provides a general method for extending crisp mathematical concepts to address fuzzy quantities, such as real algebraic operations on fuzzy numbers. These operations are computationally effective generalizations of interval analysis. Fuzzy cognitive maps are growing in popularity because they allow for the modeling of subjective information that is linguistic in nature, and for a more effective assessment of casual affects between concepts within the map that are also very subjective and uncertain. Fuzzy agent-based models will see increasing use because they allow for the extension of deductive reasoning into the realm of inducting reasoning.

Finally, a new area called *situational awareness* (referred to earlier in data fusion) is growing rapidly in many disciplines, from battlefield strategy, weather forecasting, political policies, and satellite control and positional awareness (Raol, 2010). In situational awareness, there are various elements being identified by a classification method, and then the elements are combined to assess a possible situation. For example, suppose you are tracking unknown aircraft in a battlefield scenario, and you have sensors that tell you the identity of the aircraft as being either a friend or foe, and you have a sensor that tells you the direction of flight (departing from you or approaching you), and you want to assess whether the behavior of the aircraft is friendly or hostile based on the information from these two sensors. The information could be noisy, or it could come with low confidence. Fuzzy logic is very useful in these kinds of problems.

References

Akai, T. (1994). *Applied numerical methods for engineers* (New York: John Wiley & Sons).

Axelrod, R. (1976). *Structure of Decision* (Princeton, NJ: Princeton University Press).

Axelrod, R. (2003). Advancing the art of simulation in the social sciences. *Jap. J. Math. Inf. Syst.*, **12** (3), 1–19.

Axtell, R. (2000). "*Why agents?*" The Brookings Institution, Center on Social and Economic Dynamics, Washington, DC, 22 p (CSED Working Paper, n. 17).

Ball, P. (2004). *Critical Mass: How One Thing Leads to Another* (New York: Farrar, Straus, and Giroux).

Bankes, S. C. (2002). Tools and techniques for developing policies for complex and uncertain systems. *Proc. Nat. Acad. Sci. USA*, **99**, 2763–2766.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Nat. Acad. Sci. USA*, **99**, 7280–7287.

Carroll, L. (1865). *Alice in Wonderland* (London, UK: Macmillan).

Chialvo, D. (2006). "Psychophsics: Are our senses critical?" *Nature Physics*, **2**, 301–302.

Cox, E. (1994). *The Fuzzy Systems Handbook* (Cambridge, MA: Academic Press Professional).

deNeufville, R. (1990). *Applied Systems Analysis: Engineering Planning and Technology Management* (New York: McGraw-Hill).

Dubois, D., and Prade, H. (1980). *Fuzzy Sets and Systems: Theory and Applications* (New York: Academic Press).

Epstein, J., and Axtell, R. (1996). *Growing Artificial Societies: Social Science from Bottom Up* (Washington, DC/ Cambridge, MA: Brookings Press and MIT Press).

Esme, B. (2009), Kalman Filter for Dummies, retrieved from bilgin.esme.org/BitsBytes/KalmanFilterforDummies.aspx.

Federal Energy Regulatory Commission (FERC). (2007). A 2007 assessment of demand response and advanced metering.

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (Reading, MA: Addison-Wesley).

Harp, D. (2009). Inclusion of irrational decision making in agent-based modeling using fuzzy logic. Unpublished research report to T. Ross.

Kalman, R. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engineering*, **82** (1), 35–45.

Kardaras, D., and Karakostas, B. (1999). The use of fuzzy cognitive maps to simulate the information systems strategic planning process. *Inf. Softw. Technol.*, **41**, 197–210.

Klein, L. (2004). *Sensor and Data Fusion* (Bellingham, WA: SPIE Press).

Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).

Kosko, B. (1986). Fuzzy cognitive maps. *Int. J. Man Machine Stud.*, **24**, 65–75.

Kosko, B. (1992). *Neural Networks and Fuzzy Systems* (Englewood Cliffs, NJ: Prentice-Hall).

Kosko, B. (1993). *Fuzzy Thinking* (New York: Hyperion Press).

Kosko, B. (1997). *Fuzzy Engineering* (Upper Saddle River, NJ: Prentice Hall).

Lincoln, D., and Ross, T. (2012). Evaluation of commercial facility demand response using genetically evolved fuzzy cognitive mapping. *Int. J. Distributed Energy Resources*, **8** (4), 315–342.

Macy, M. W., and Flache, A. (2002). Leaning dynamics in social dilemmas. *Ann. Rev. Sociol.*, **28**, 7280–7287.

Macy, M., and Willer, R. (2002). From factors to actors: Computational sociology and agent-based modelling. *Ann. Rev. Sociol.*, **28**, 143–166.

Mateou, N. H., and Andreou, A. S. (2008). A framework for developing intelligent decision support systems using evolutionary fuzzy cognitive maps. *J. Intellig. Fuzzy Syst.*, **19**, 151–170.

McNeill, D., and Freiberger, P. (1993). *Fuzzy Logic* (New York: Simon & Schuster, New York).

Nourdeldin, A., El-Shafie, A., and Reda Taha, M. (2007). Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation. *Eng. Appl. Artificial Intell.*, **20**, 49–61.

Pelaez, C. E., and Bowles, J. B. (1995). Applying fuzzy cognitive-maps knowledge-representation to failure modes effects analysis. *Proceedings Annual Reliability and Maintainability Symposium* (New York: IEEE Press).

Pelaez, C. E., and Bowles, J. B. (1996). Using fuzzy cognitive maps as a system model for failure modes and effects analysis. *Inf. Sci.*, **88**, 177–199.

Raol, J. (2010). *Multi-Sensor Data Fusion with MATLAB*. (Boca Raton, FL: CRC Press).

Ross, T. (2004). *Fuzzy Logic with Engineering Applications*, 2nd ed. (Chichester, UK: John Wiley & Sons, Ltd.).

Ross, T. (2010). *Fuzzy Logic with Engineering Applications*, 3rd ed. (Chichester, UK: John Wiley & Sons, Ltd.).

Ross, T., Booker, J., and Parkinson, J. (2003). *Fuzzy Logic and Probability Applications: Bridging the Gap* (Philadelphia, PA: Society for Industrial and Applied Mathematics).

Sakawa, M. (1993). *Fuzzy Sets and Interactive Multiobjective Optimization* (New York: Plenum Press).

Schmucker, K. (1984). *Fuzzy Sets, Natural Language Computations, and Risk Analysis* (Rockville, MD: Computer Science Press).

Subramanian, V., Burks, T., and Dixon, W. (2009). Sensor fusion using fuzzy logic enhanced Kalman filter for autonomous vehicle guidance in Citrus groves. *Tran. Am. Soc. Agricul. and Biol. Eng.*, **52** (5), 1411–1422.

Taber, R. (1994). Fuzzy cognitive maps model social systems. *AI Exp.*, July 19–23.

Tsadiras, A. K., and Margaritis, K. G. (1996). Using certainty neurons in fuzzy cognitive maps. *Neural Netw. World*, **4**, 719–728.

Vanderplaats, G. (1984). *Numerical Optimization Techniques for Engineering Design with Applications* (New York: McGraw-Hill).

Vasantha, W. B., and Smarandache, F. (2003). *Fuzzy Cognitive Maps and Neutrosophic Cognitive Maps* (Phoenix, AZ: Xiquan).

Wong, F., and Ross, T. (1985) Treatment of uncertainties in structural dynamics models, in F. Deyi and L. Xihui, eds., *Proceedings of the International Symposium on Fuzzy Mathematics in Earthquake Research* (Bejing: Seismological Press).
Yager, R. R. (1997). A general approach to the fusion of imprecise information. *Int. J. Intell. Sys.*, **12** (1), 1–29.
Yager, R. R. (1986). A characterization of the extension principle. *Fuzzy Sets Syst.*, **18**, 205–217.
Zadeh, L. (1972). *On Fuzzy Algorithms*. Memo UCB/ERL M-325. (Berkeley: University of California).
Zadeh, L. (1975). The concept of a linguistic variable and its application to approximate reasoning, Part I. *Inf. Sci.*, **8**, 199–249.

Problems

Fuzzy Optimization

12.1 The feed-forward transfer function for a unit-feedback control system is $1/(s+1)$, as shown in Figure P12.1. A unit step signal is input to the system. Determine the minimum error of the system response by using a fuzzy optimization method for the time period, $0 < t < 10$ s, and a fuzzy constraint given as

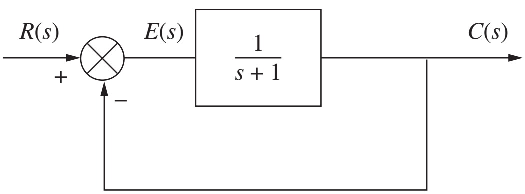

Figure P12.1

$$\mu_c(t) = \begin{cases} 1, & 0 \le t \le 1, \\ e^{1-t}, & t > 1. \end{cases}$$

12.2 A beam structure is forced by an axial load P (Figure P12.2). When P is increased to its *critical* value, the beam will buckle. Prove that the critical force P to cause buckling can be expressed by a function

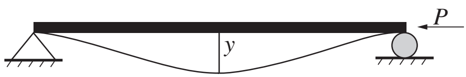

Figure P12.2

$$P = \frac{n^2 \pi^2 EI}{L},$$

where

EI = stiffness of the beam
L = span length of the beam

n = number of sine waves the beam shape takes when it buckles (assume it to be continuous).

If $0 \leq n \leq 2$, assume that n is constrained by the fuzzy member function

$$\mu_c(n) = \begin{cases} 1-n, & 1 \leq n \leq 2, \\ 0, & n < 1. \end{cases}$$

12.3 Suppose that the beam structure in Problem 12.2 also has a transverse load P applied at the middle of the beam. Then, the maximum bending stress can be calculated by the equation $\sigma_b = Pl/4w_z$, where w_z, with units cubic meters, is a coefficient based on the shape and size of the cross section of the beam, and l is in meters. If $0 \leq \sigma_b \leq$ 60 MPa, and the fuzzy constraint function for σ_b is

$$\mu_c(\sigma_b) = \begin{cases} \dfrac{1}{(x-1)^2}, & 0 \leq x \leq 1, \\ 0, & x > 1, \end{cases}$$

where

$$x = \frac{\sigma}{60\,\text{MPa}},$$

combine the conditions given in Problem 12.2 to find the optimum load P, in newtons. *Hint: This problem involves multiple constraints.*

12.4 In the metallurgical industry, the working principle for a cold rolling mill is to extrude a steel strip through two rows of working rollers, as shown in Figure P12.4. The size of the roller is important. The stress between the roller and the strip can be expressed by the following function:

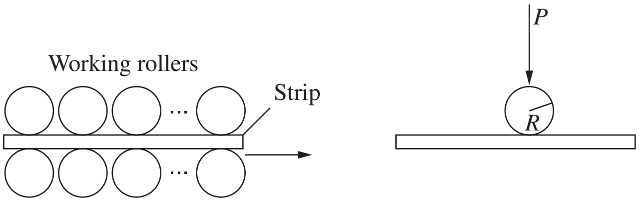

Figure P12.4

$$\sigma_H = 0.564\sqrt{\frac{PE}{LR}},$$

where

E = Young's modulus (kN cm^{-2})
P = loading force (N)

L = contact length between roll and strip (cm)
R = radius of a roller (cm).

If $\sigma_H = 2.5$ kN cm^{-2} and $10 < R < 20$, find the minimum R in which σ_H has a maximum value. The radius R has a fuzzy constraint of

$$\mu_c(R) = \begin{cases} 1, & 10 \le R \le 15, \\ \dfrac{20-R}{5}, & 15 < R \le 20. \end{cases}$$

Fuzzy Cognitive Mapping

12.5 For the information in Example 12.3, find the stabilized state vector corresponding to an initial state vector of [0, 1, 0, 0, 0].

12.6 For the information in Example 12.3, find the stabilized state vector corresponding to a fuzzy linguistic effect on the path from C_1 to C_2 that is "Much."

12.7 For the information in Example 12.3, find the stabilized state vector corresponding to an initial state vector of [0, 0, 1, 0, 0] and a fuzzy linguistic effect on the path from C_2 to C_3 that is "Some."

12.8 An engineer develops a five-concept FCM for the design of a building. The concepts are C_1, design requirements; C_2, amount of material required for the job; C_3, cost of the building; C_4, aesthetics; C_5, owner satisfaction with the building. The FCM has two paths from C_1 to C_5, with the following effect levels: $I_1 = \{C_1 - C_2 - C_3 - C_5\} = \{+1$ (a lot), $+1$ (a lot), -1(much)$\}$ and $I_2 = \{C_1 - C_4 - C_5\} = \{+1$(some), $+1$(much)$\}$. Using an initial vector of [1, 0, 0, 0, 0], $P = \{$none < some < much < a lot$\}$, and activation levels of $\{-1, 0, +1\}$, draw the FCM diagram, find the adjacency matrix, and conduct an analysis to stabilization to find the output state vector and the linguistic total effect. Discuss how a conventional CM would have produced an *indeterminate* situation.

12.9 An environmental engineer wants to find the effects of fertilizers on growing green mass in an ecosystem. There are five concepts: C_1, amount of fertilizer; C_2, amount of weeds in the ecosystem; C_3, resulting levels of phosphorus and nitrogen in runoff water; C_4, fines that need to be paid for fertilizer use; C_5, amount of green mass grown. The FCM has two paths from C_1 to C_5, with the following effect levels: $I_1 = \{C_1 - C_3 - C_4 - C_5\} = \{+1$(much), $+1$(much), -1(some)$\}$ and $I_2 = \{C_1 - C_2 - C_5\} = \{-1$(some), $+1$ (much)$\}$. Using an initial vector of [1, 0, 0, 0, 0], $P = \{$none < some < much < a lot$\}$, and activation levels of $\{-1, 0, +1\}$, draw the FCM diagram, find the adjacency matrix, and conduct an analysis to stabilization to find the output state vector and the linguistic total effect.

12.10 An industrial engineer wants to find the profitability of a steel manufacturing plant. There are five concepts: C_1, steel manufacturing process; C_2, amount of workers to produce the steel; C_3, resulting levels of pollution from the steel plant; C_4, amount of pollution reduction machinery to reduce toxic emissions; C_5, profitability of the plant. The FCM has two paths from C_1 to C_5, with the following effect levels:

$I_1 = \{C_1 - C_2 - C_3 - C_4 - C_5\} = \{-1(\text{a lot}), +1\ (\text{a lot}), +1(\text{much}), -1(\text{some})\}$ and $I_2 = \{C_1 - C_2 - C_5\} = \{-1(\text{a lot}), -1(\text{a lot})\}$. Using an initial vector of [1, 0, 0, 0, 0], $P = \{\text{none} < \text{some} < \text{much} < \text{a lot}\}$, and activation levels of $\{-1, 0, +1\}$, draw the FCM diagram, find the adjacency matrix, and conduct an analysis to stabilization to find the output state vector and the linguistic total effect.

12.11 The following FCM shows a problem in developing a new industry in the use of bamboo poles for engineered structures. The FCM has fuzzy weights on the connective paths as shown in Figure P12.11. Using the following concepts,

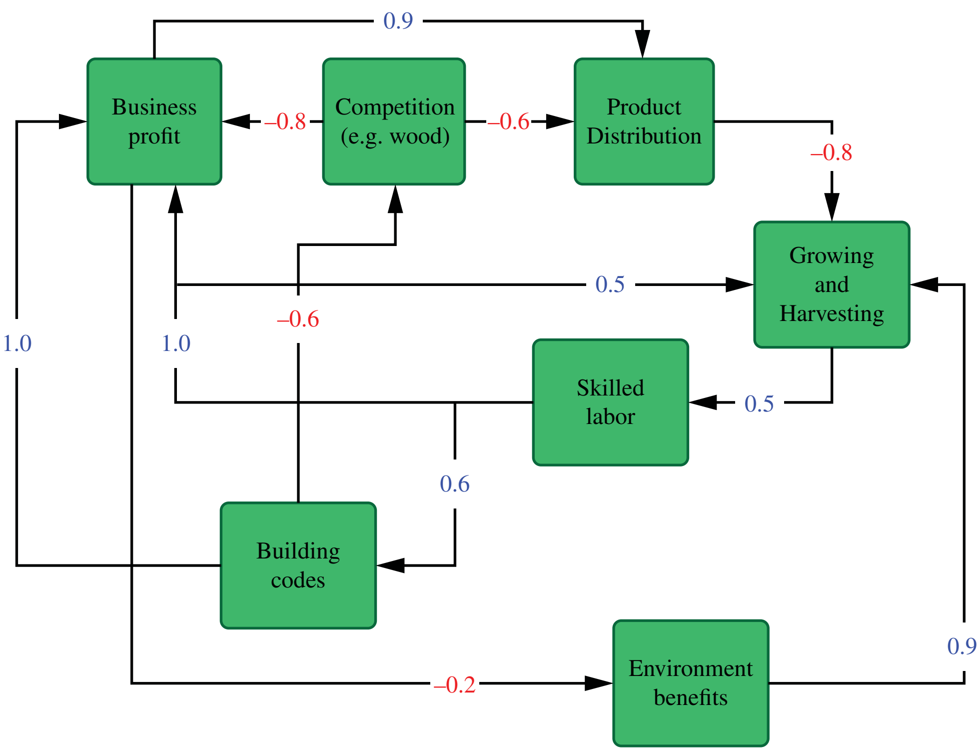

Figure P12.11

C1 = Business profits
C2 = Availability of Skilled Labor
C3 = Competition from other products (e.g., wood)
C4 = Bamboo Distribution
C5 = Growing and Harvesting Conditions
C6 = Environmental Benefits
C7 = Building Codes containing bamboo specifications

conduct an FCM based on your own assumptions for the activation levels (initial input vector) for each of the concepts where the goals might be to maximize the business profits, or to maximize environmental benefits, etc. Because this will be a new industry, many of these concepts contain sketchy and fuzzy information.

Fuzzy Agent-Based Models

12.12 In a construction operation in a remote area, the performance of truck tires is essential. Suppose there are two variables: amount of air in the tires and the amount of remaining tread on the tires. The performance of the tires will increase with greater air pressure and greater tread. The engineer wants to find out what the *expected performance* will be if the amount of air in the tires is *medium* and the amount of tread on the tires is *high*. For this problem, develop an ABM and use three-membership functions for the air in the tires and for the tread on the tires, both on the input domain [0, 1], use three-membership functions for the performance of the air and the tread, both on the output domain [0, 1], and use the following rules:

IF air is low, THEN air performance is weak
IF air is medium, THEN air performance is OK
IF air is high, THEN air performance is strong
IF tread is low, THEN tread performance is weak
IF tread is medium, THEN tread performance is OK
IF tread is high, THEN tread performance is Strong

Using a parameter $wf = 0.5$, determine the linguistic value of the total performance.

12.13 An engineer is creating a model to classify footings of building foundations. He wants to classify the capacity of the footings according to the inputs of vertical loads on the columns (low, medium, high) and bearing capacity of the soil (weak, moderate, good). Using a simple fuzzy ABM approach, with assumed membership functions and rules and a value for *wf*, conduct an analysis for this problem.

Fuzzy Kalman Filter

12.14 Using the fuzzy KF illustrated in Example 12.14, conduct a simulation using different rules and membership functions from those specified in the example.

13

Monotone Measures

Belief, Plausibility, Probability, and Possibility

Only those who attempt the absurd can achieve the impossible.
Albert Einstein, theoretical physicist philosopher and Nobel Prize Winner

A bag contains 2 counters, as to which nothing is known except that each is either black or white. Ascertain their colours without taking them out of the bag.
Lewis Carroll, author and mathematician, *Pillow Problems*, 1893

I still believe in the possibility of a model of reality, that is to say, of a theory, which represents things themselves and not merely the probability of their occurrence.
Albert Einstein, commenting on quantum physics after his Herbert Spencer Lecture, 1933

Most of this text has dealt with the quantification of various forms of nonnumeric uncertainty. Two prevalent forms of uncertainty are those arising from vagueness and from imprecision. How do vagueness and imprecision differ as forms of uncertainty? Often, vagueness and imprecision are used synonymously, but they can differ in the following sense. Vagueness can be used to describe certain kinds of uncertainty associated with linguistic information or intuitive information. Examples of vague information are that the image quality is “good,” or that the transparency of an optical element is “acceptable.” Imprecision can be associated with quantitative or countable data as well as uncountable data. As an example of the latter, one might say the length of a bridge span is “long.” An example of countable imprecision would be to report the length to be 300 m. If we take a measuring device and measure the length of the bridge 100 times we likely will come up with 100 different values; the differences in the numbers will no doubt be on the order of the precision of the measuring device. Measurements using a 10-m chain will be less precise than those developed from a laser theodolite. If we plot the bridge lengths on some sort of probit paper and develop a Gaussian distribution to describe the length of this bridge, we could state the imprecision in probabilistic terms. In this case, the length of the bridge is uncertain to some degree of precision that is quantified in the language of statistics. Because we are not able to make this measurement an infinite number of times, there is also uncertainty in the statistics describing the bridge length. Hence, imprecision can be used to quantify random variability in quantitative uncertainty, and it can also be used to describe a lack of knowledge for descriptive entities (e.g., acceptable transparency and good image quality). Vagueness is usually related to nonmeasurable issues.

This chapter develops the relationships between probability theory and evidence theory; to a limited extent, it also shows the relationship between a possibility theory, founded on crisp sets, and a fuzzy set theory. All of these theories are related under an umbrella theory termed *monotone measures* (see Klir and Smith, 2001), which was termed *fuzzy measure theory* for a couple of decades despite the confusion this generates when we try to distinguish other theories from fuzzy set theory; all of these theories have been used to characterize and model various forms of uncertainty. That they are all related mathematically is an especially crucial advantage in their use in quantifying the uncertainty spectrum because, as more information about a problem becomes available, the mathematical description of uncertainty can easily transform from one theory to the next in the characterization of the uncertainty. This chapter begins by developing monotone measures as an overarching framework for the other theories used to characterize various forms of uncertainty. The development continues with specific forms of monotone measures such as belief, plausibility, possibility, and probability. The chapter illustrates a new method in developing possibility distributions from empirical data, and it briefly describes a special kind of relationship between a possibility distribution and a fuzzy set. Examples are provided to illustrate the various theories.

Monotone Measures

A monotone measure describes the vagueness or imprecision in the assignment of an element *a* to two or more crisp sets. Figure 13.1 shows this idea. In the figure the universe of discourse

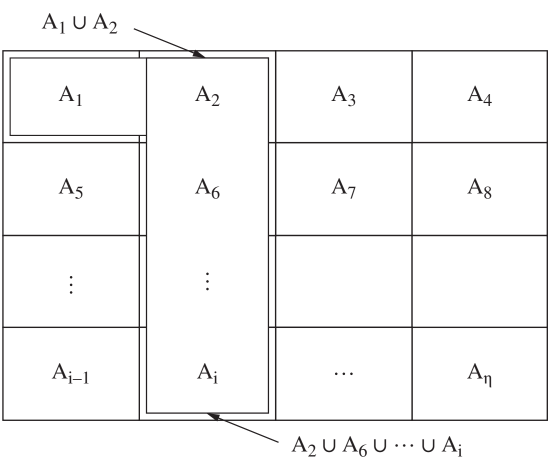

Figure 13.1 A monotone measure.

comprises a collection of sets and subsets, or the power set. In a monotone measure what we are trying to describe is the vagueness or imprecision in assigning this element *a* to any of the crisp sets on the power set. This notion is not random; the crisp sets have no uncertainty about them. The uncertainty is about the assignment. This uncertainty is usually associated with evidence to establish an assignment. The evidence can be completely lacking—the case of total ignorance—or the evidence can be complete—the case of a probability assignment. Hence, the difference between a monotone measure and a fuzzy set on a universe of elements is that, in the former, the imprecision is in the assignment of an element to one of two or more crisp sets, and in the latter the imprecision is in the prescription of the boundaries of a set.

Belief and Plausibility

There are special forms of monotone measures. A form associated with preconceived notions is called a *belief measure*. A form associated with information that is possible, or plausible, is called a *plausibility measure*. Specific forms of belief measures and plausibility measures are known as *certainty* (necessity) and *possibility measures*, respectively. The intersection of belief measures and plausibility measures (i.e., where belief equals plausibility) will be shown to be a probability. Monotone measures are defined by weaker axioms than probability theory, thus subsuming probability measures as specific forms of monotone measures.

Basically, a belief measure is a quantity, denoted bel(A), that expresses the degree of support, or evidence, for a collection of elements defined by one or more of the crisp sets existing on the power set of a universe. The plausibility measure of this collection A is defined as the "complement of the Belief of the complement of A," or

$$\mathrm{pl}(\mathrm{A}) = 1 - \mathrm{bel}\left(\overline{\mathrm{A}}\right). \tag{13.1}$$

Because belief measures are quantities that measure the degree of support for a collection of elements or crisp sets in a universe, it is entirely possible that the belief measure of some set A plus the belief measure of $\overline{\mathrm{A}}$ will not be equal to unity (the total belief, or evidence, for all elements or sets on a universe is equal to 1, by convention). When this sum equals 1, we have the condition where the belief measure is a probability; that is, the evidence supporting set A can be described probabilistically. The difference between the sum of these two quantities $\left(\mathrm{bel}(\mathrm{A}) + \mathrm{bel}\left(\overline{\mathrm{A}}\right)\right)$ and 1 is called the *ignorance*, that is, ignorance $= 1 - \left[\mathrm{bel}(\mathrm{A}) + \mathrm{bel}\left(\overline{\mathrm{A}}\right)\right]$. When the ignorance equals 0, we have the case where the evidence can be described by probability measures.

Say we have evidence about a certain prospect in our universe of discourse, evidence of some set occurring or some set being realized, and we have no evidence (zero evidence) of the complement of that event. In probability theory we must assume, because of the excluded middle axioms, that if we know the probability of A then the probability of $\bar{\mathrm{A}}$ is also known, because we have in all cases involving probability measures, $\mathrm{prob}(\mathrm{A}) + \mathrm{prob}\left(\overline{\mathrm{A}}\right) = 1$. This constraint of the excluded middle axioms is not a requirement in evidence theory. The probability of $\bar{\mathrm{A}}$ also has to be supported with some sort of evidence. If there is no evidence (zero degree of support) for $\bar{\mathrm{A}}$ then the degree of ignorance is large. This distinction between evidence theory and probability theory is important. It will also be shown that this is an important distinction between fuzzy set theory and probability theory (Gaines, 1978).

Monotone measures are useful in quantifying uncertainty that is difficult to measure or that is linguistic in nature. For example, in assessing structural damage in buildings and bridges after an earthquake or hurricane, evidence theory has proven quite successful because what we have are nonquantitative estimates from experts; the information concerning damage is not about how many inches of displacement or microinches per inch of strain the structure might have undergone, but rather is about expert judgment concerning the suitability of the structure for habitation or its intended function. These kinds of judgments are not quantitative; they are qualitative.

The mathematical development for monotone measures follows (see, for example, Klir and Folger, 1988). We begin by assigning a value of membership to each crisp set existing in the power set of a universe, signifying the degree of evidence or belief that a particular element from the universe, say x, belongs in any of the crisp sets on the power set. We will label this membership $g(\mathrm{A})$, where it is a mapping between the power set and the unit interval,

$$g : \mathrm{P(X)} \to [0,1], \tag{13.2}$$

and where $\mathrm{P}(X)$ is the power set of all crisp subsets on the universe, X (Chapter 2). So, the membership value $g(\mathrm{A})$ represents the degree of available evidence of the belief that a given element x belongs to a crisp subset A.

The collection of these degrees of belief represents the *fuzziness* associated with several *crisp* alternatives. This type of uncertainty, which we call a monotone measure, is *different* from the uncertainty associated with the boundaries of a single set, which we call a *fuzzy set*. Monotone measures are defined for a *finite* universal set by at least three axioms, two of which are given here (a third axiom is required for an infinite universal set):

$$\begin{aligned}
&1.\ g(\varnothing)=0, \quad g(X)=1,\\
&2.\ g(\mathrm{A}) \le g(\mathrm{B}) \text{ for } \mathrm{A},\mathrm{B} \in \mathrm{P(X)},\ \mathrm{A} \subseteq \mathrm{B}.
\end{aligned} \tag{13.3}$$

The first axiom represents the boundary conditions for the monotone measure, $g(\mathrm{A})$. It says that there is no evidence for the null set and there is complete (i.e., unity) membership for the universe. The second axiom represents monotonicity by simply stating that if one set A is completely contained in another set B, then the evidence supporting B is at least as great as the evidence supporting the subset A.

A belief measure also represents a mapping from the crisp power set of a universe to the unit interval representing evidence, denoted

$$\mathrm{bel} : \mathrm{P(X)} \to [0,1]. \tag{13.4}$$

Belief measures can be defined by adding a third axiom to those represented in Equation (13.3), which is given as

$$\begin{aligned}
\mathrm{bel}(\mathrm{A}_1 \cup \mathrm{A}_2 \cup \cdots \cup \mathrm{A}_n) \ge {} & \sum_i \mathrm{bel}(\mathrm{A}_i) - \sum_{i<j} \mathrm{bel}(\mathrm{A}_i \cap \mathrm{A}_j) + \cdots \\
& + (-1)^{n+1}\mathrm{bel}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \cdots \cap \mathrm{A}_n),
\end{aligned} \tag{13.5}$$

where there are n crisp subsets on the universe X. For each crisp set $A \in P(X)$, bel(A) is the degree of belief in set A based on the available evidence. When the sets A_i in Equation (13.5) are pairwise disjoint, that is, where $A_i \cap A_j = \emptyset$, then Equation (13.5) becomes

$$\mathrm{bel}(A_1 \cup A_2 \cup \cdots \cup A_n) \geq \mathrm{bel}(A_1) + \mathrm{bel}(A_2) + \cdots + \mathrm{bel}(A_n). \tag{13.6}$$

For the special case where $n = 2$, we have two disjoint sets A and $\bar{A}$, and Equation (13.6) becomes

$$\mathrm{bel}(A) + \mathrm{bel}(\overline{A}) \leq 1. \tag{13.7}$$

A plausibility measure is also a mapping on the unit interval characterizing the total evidence, that is,

$$\mathrm{pl} : P(X) \to [0, 1]. \tag{13.8}$$

Plausibility measures satisfy the basic axioms of monotone measures, Equation (13.3), and one additional axiom (different from Equation (13.5) for beliefs),

$$\begin{aligned} \mathrm{pl}(A_1 \cap A_2 \cap \cdots \cap A_n) \leq \sum_i \mathrm{pl}(A_i) - \sum_{i<j} \mathrm{pl}(A_i \cup A_j) + \cdots \\ + (-1)^{n+1} \mathrm{pl}(A_1 \cup A_2 \cup \cdots \cup A_n). \end{aligned} \tag{13.9}$$

From Equation (13.1), we have a mutually dual system between plausibility and belief (see Shafer, 1976),

$$\begin{aligned} \mathrm{pl}(A) &= 1 - \mathrm{bel}(\overline{A}) \\ \mathrm{bel}(A) &= 1 - \mathrm{pl}(\overline{A}) \end{aligned}. \tag{13.10}$$

For the specific case of $n = 2$, that is, for two disjoint sets A and $\bar{A}$, Equation (13.10) produces

$$\mathrm{pl}(A) + \mathrm{pl}(\overline{A}) \geq 1. \tag{13.11}$$

By combining Equations (13.7) and (13.10), it can be shown that

$$\mathrm{pl}(A) \geq \mathrm{bel}(A). \tag{13.12}$$

Equation (13.12) simply states that for whatever evidence supports set A, its plausibility measure is always at least as great as its belief measure.

We now define another function on the crisp sets ($A \in P(X)$) of a universe, denoted $m(A)$, which can be used to express and determine both belief and plausibility measures. This measure is also a mapping from the power set to the unit interval,

$$m : P(X) \to [0, 1]. \tag{13.13}$$

This measure, called a *basic evidence assignment* (bea), has been termed a *basic probability assignment* (bpa) before in the literature, and has boundary conditions

$$m(\emptyset) = 0, \tag{13.14}$$

$$\sum_{A \in P(X)} m(A) = 1. \tag{13.15}$$

The measure $m(A)$ is the degree of belief that a specific element, x, of the universe X belongs to the set A, *but not to any specific subset of* A. In this way $m(A)$ differs from both beliefs and plausibility. It is important to remark here, to avoid confusion with probability theory, that there is a distinct difference between a bea and a probability density function (pdf). The former are defined on sets of the power set of a universe (i.e., on $A \in P(X)$), whereas the latter are defined on the singletons of the universe (i.e., on $x \in P(X)$). This difference will be reinforced through some examples in this chapter. To add to the jargon of the literature, the second boundary condition, Equation (13.15), provides for a *normal* bea.

The bea is used to determine a belief measure by

$$\text{bel}(A) = \sum_{B \subseteq A} m(B). \tag{13.16}$$

In Equation (13.16), note that $m(A)$ is the degree of evidence in set A *alone*, whereas bel(A) is the total evidence in set A *and* all subsets (B) of A. The measure $m(A)$ is used to determine a plausibility measure by

$$\text{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{13.17}$$

Equation (13.17) shows that the plausibility of an event A is the total evidence in set A plus the evidence in all sets of the universe that intersect with A (including those sets that are also subsets of A). Hence, the plausibility measure in set A contains all the evidence contained in a belief measure (bel(A)) plus the evidence in sets that intersect with set A. Hence, Equation (13.12) is verified.

Example 13.1

A certain class of short-range jet aircraft has had, for the shorter fuselage versions, a history of an oscillatory behavior described as *vertical bounce*. This is as a result of the in-flight flexing of the fuselage about two body-bending modes. Vertical bounce is most noticeable at the most forward and aft locations in the aircraft. An acceptable acceleration threshold of $\pm 0.1\,g$ has been set as the point at which aft lower-body vortex generators should be used to correct this behavior. To avoid the cost of instrumented flight tests, expert engineers often decide whether vertical bounce is present in the aircraft. Suppose an expert engineer is asked to assess the evidence in a particular plane for the following two conditions:

1. Are oscillations caused by other phenomena? (O)
2. Are oscillations characteristic of the vertical bounce? (B)

Table 13.1 Measures of evidence for aircraft bounce.

	Expert		
Focal element, A_i	$m(A_i)$	bel(A_i)	pl(A_i)
$\emptyset$	0	0	0
O	0.4	0.4	0.8
B	0.2	0.2	0.6
$O \cup B$	0.4	1	1

This universe is a simple one, consisting of the singleton elements O and B. The non-null (Equation [13.14] reminds us that the null set contains no evidence, that is, $m(\emptyset) = 0$) power set then consists simply of the two singletons and the union of these two, $O \cup B$; including the null set there are $2^2 = 4$ elements in the power set. All the elements in the power set are called *focal elements*. Suppose the expert provides the measures of evidence shown in Table 13.1 for each of the focal elements (i.e., the expert gives $m(A_i)$, for $i = 1, \ldots, 4$). Note that the sum of the evidences in the $m(A)$ column equals unity, as required by Equation (13.15). We now want to calculate the degrees of belief and plausibility for this evidence set. Using Equation (13.16), we find

$$\text{bel}(O) = m(O) = 0.4 \quad \text{and} \quad \text{bel}(B) = m(B) = 0.2,$$

as seen in Table 13.1. The singletons O and B have no other subsets in them. Using Equation (13.16) we find

$$\text{bel}(O \cup B) = m(O) + m(B) + m(O \cup B) = 0.4 + 0.2 + 0.4 = 1,$$

as seen in Table 13.1. Using Equation (13.17), we find

$$\text{pl}(O) = m(O \cap O) + m(O \cap (O \cup B)) = 0.4 + 0.4 = 0.8$$

and

$$\text{pl}(B) = m(B \cap B) + m(B \cap (O \cup B)) = 0.2 + 0.4 = 0.6$$

because sets O and B both intersect with the set $O \cup B$; and finally,

$$\begin{aligned}\text{pl}(O \cup B) = {} & m((O \cup B) \cap O) + m((O \cup B) \cap B) \\ & + m((O \cup B) \cap (O \cup B)) = 0.4 + 0.2 + 0.4 = 1\end{aligned}$$

because all sets in the power set intersect with $O \cup B$. These quantities are included in the fourth column of Table 13.1. Thus, the engineer *believes* the evidence supporting set O (other oscillations) is at least 0.4 and *possibly* as high as 0.8 (plausibility), and *believes* the evidence supporting set B (vertical bounce) is at least 0.2 and *possibly* as high as 0.6 (plausibility). Finally, the evidence supporting either of these sets ($O \cup B$) is full, or complete (i.e., bel = pl = 1).

Evidence Theory

The material presented in the preceding section now sets the stage for a more complete assessment of evidence, called *evidence theory* (Shafer, 1976). Suppose the evidence for certain monotone measures comes from more than one source, say two experts. Evidence obtained in the same context (e.g., for sets A_i on a universe X) from two independent sources (e.g., two experts) and expressed by two beas (e.g., m_1 and m_2) on some power set P(X) can be combined to obtain a joint bea, denoted m_{12}, using Dempster's rule of combined evidence (Dempster, 1967). The procedure to combine evidence is given here in Equations (13.18) and (13.19):

$$m_{12}(\mathrm{A}) = \frac{\sum_{\mathrm{B}\cap\mathrm{C}=\mathrm{A}} m_1(\mathrm{B})\cdot m_2(\mathrm{C})}{1-K}, \quad \text{for } \mathrm{A} \neq \varnothing, \tag{13.18}$$

where the denominator is a normalizing factor such that

$$K = \sum_{\mathrm{B}\cap\mathrm{C}=\varnothing} m_1(\mathrm{B})\cdot m_2(\mathrm{C}). \tag{13.19}$$

Dempster's rule of combination combines evidence in a manner analogous to the way in which joint pdfs in probability theory are calculated from two independent marginal pdfs. We can define a *body of evidence*, then, as a pair (A, m) where A are sets with available evidence m(A).

Example 13.2

If a generator is to run untended, the external characteristic of the shunt machine may be unsatisfactory, and that of a series even more so, because a source of constant potential difference supplying a varying load current is usually required. The situation is even less satisfactory if the load is supplied via a feeder with appreciable resistance because this will introduce an additional drop in potential at the load end of the feeder. What is required is a generator with rising external characteristics because this would counteract the effect of feeder resistance. Such a characteristic may be obtained from a compound generator. In a compound generator, we can get variable-induced electromotive force (emf) with increase of load current by arranging the field magnetomotive force (mmf). So, shunt and field windings are used in the generator as shown in Figure 13.2. In this figure, R_a, R_c, R_f, and R_s are the armature, compound, field, and series resistance, respectively; I_f and I_l are the field and load current, respectively; E_a and V_t are the induced armature and terminal voltages, respectively; and N_f and N_s are the number of turns in the series and field windings, respectively. By arranging the field winding in different combinations (i.e., varying the difference between N_s and N_f), we can get different combinations of compound generators; in particular, we can get (1) overcompounded (OC), (2) flatcompounded (FC), and (3) undercompounded (UC) generators. Each type of compounded generator has its own external and internal characteristics.

We can say that these three types comprise a universal set of generators, X. Let us consider two experts, E_1 an electrical engineer and E_2 a marketing manager, called to evaluate the efficiency and performance of a compound generator. We can come to some conclusion that, say for a particular outdoor lighting situation, the machine chosen by the two experts may be different for various reasons. On the one hand, the electrical engineer may think about performance

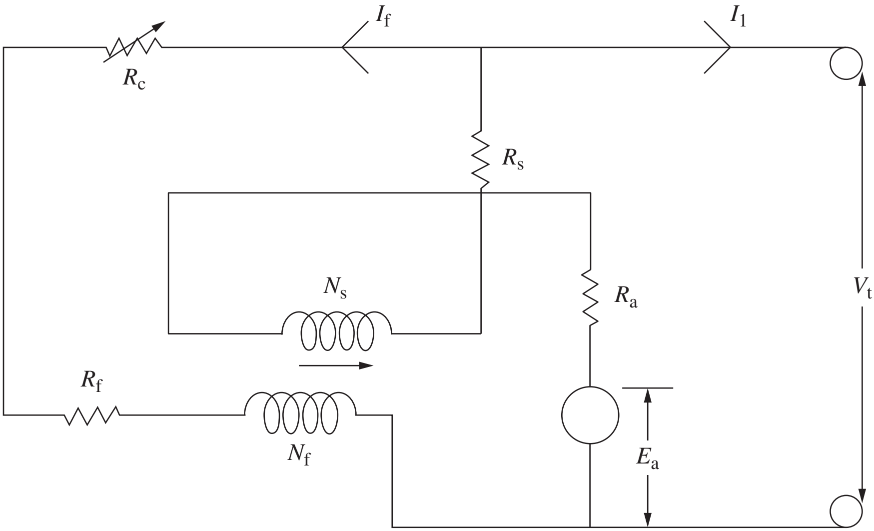

Figure 13.2 Electrical diagram of a compound generator.

issues like minimizing the error, maintaining constant voltage, and other electrical problems. On the other hand, the marketing manager may be concerned only with issues like minimizing the cost of running or minimizing maintenance and depreciation costs. In reality, both experts may have valid reasons for the selection of a specific machine required for final installation.

Suppose that a company hires these two individuals to help it decide on a specific kind of generator to buy. Each expert is allowed to conduct tests or surveys to collect information (evidence) about the value of each of the three generators. The universe showing the individual sets of the power set is illustrated in Figure 13.3. The focal elements of the universe in this figure are OC, FC, UC, OC ∪ FC, OC ∪ UC, FC ∪ UC, and OC ∪ FC ∪ UC (hereafter we ignore the null set in determining evidence since this set contains no evidence by definition, Equation (13.14)), as listed in the first column of Table 13.2. Note that there are $2^3 - 1 = 7$ non-null elements. Suppose that the two experts, E_1 and E_2, give their information (evidence measures) about each focal element A_i, where $i = 1, 2, \ldots, 7$; that is, they provide $m_1(A_i)$ and $m^2(A_i)$, respectively (the second and fourth columns in Table 13.2). Note that the sum of the entries in the second and fourth columns equals unity, again guaranteeing Equation (13.15).

Using this information, we can calculate the belief measures for each expert. For example,

$$\mathrm{bel}_1(\mathrm{OC} \cup \mathrm{FC}) = \sum_{\mathrm{B} \subseteq (\mathrm{OC} \cup \mathrm{FC})} m_1(\mathrm{B}) = m_1(\mathrm{OC} \cup \mathrm{FC}) + m_1(\mathrm{OC}) + m_1(\mathrm{FC})$$

$$= 0.15 + 0 + 0.05 = 0.20.$$

$$\mathrm{bel}_2(\mathrm{FC} \cup \mathrm{UC}) = \sum_{\mathrm{B} \subseteq (\mathrm{FC} \cup \mathrm{UC})} m_2(\mathrm{B}) = m_2(\mathrm{FC} \cup \mathrm{UC}) + m_2(\mathrm{FC}) + m_2(\mathrm{UC})$$

$$= 0.20 + 0.15 + 0.05 = 0.40.$$

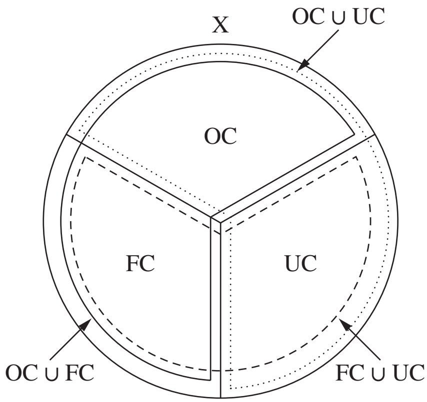

Figure 13.3 Universe X of compound generators.

Table 13.2 Focal elements and evidence for compound generators.

	Expert 1		Expert 2		Combined evidence	
Focal elements, A_i	m_1 (A_i)	bel_1 (A_i)	m_2 (A_i)	bel_2 (A_i)	m_{12} (A_i)	bel_{12} (A_i)
OC	0	0	0	0	0.01	0.01
FC	0.05	0.05	0.15	0.15	0.21	0.21
UC	0.05	0.05	0.05	0.05	0.09	0.09
OC ∪ FC	0.15	0.20	0.05	0.20	0.12	0.34
OC ∪ UC	0.05	0.10	0.05	0.10	0.06	0.16
FC ∪ UC	0.10	0.20	0.20	0.40	0.20	0.50
OC ∪ FC ∪ UC	0.60	1	0.50	1	0.30	1

The remaining calculated belief values for the two experts are shown in the third and fifth columns of Table 13.2.

Using the Dempster rule of combination (Equations (13.18) and (13.19)), we can calculate the combined evidence (m_{12}) measures (calculations will be to two significant figures). First, we must calculate the normalizing factor, K, using Equation (13.19). In calculating this expression we need to sum the multiplicative measures of all those focal elements whose intersection is the null set, that is, all those focal elements that are disjoint:

$$\begin{aligned}K = {} & m_1(\mathrm{FC})m_2(\mathrm{OC}) + m_1(\mathrm{FC})m_2(\mathrm{UC}) + m_1(\mathrm{FC})m_2(\mathrm{OC}\cup\mathrm{UC}) + m_1(\mathrm{OC})m_2(\mathrm{FC}) \\ & + m_1(\mathrm{OC})m_2(\mathrm{UC}) + m_1(\mathrm{OC})m_2(\mathrm{FC}\cup\mathrm{UC}) + m_1(\mathrm{UC})m_2(\mathrm{FC}) \\ & + m_1(\mathrm{UC})m_2(\mathrm{OC}) + m_1(\mathrm{UC})m_2(\mathrm{FC}\cup\mathrm{OC}) + m_1(\mathrm{FC}\cup\mathrm{OC})m_2(\mathrm{UC}) \\ & + m_1(\mathrm{FC}\cup\mathrm{UC})m_2(\mathrm{OC}) + m_1(\mathrm{OC}\cup\mathrm{UC})m_2(\mathrm{FC}) = 0.03.\end{aligned}$$

Hence $1 - K = 0.97$. So, for example, combined evidence on the set FC can be calculated using Equation (13.18):

$$
\begin{aligned}
m_{12}(\mathrm{FC}) = \{&m_1(\mathrm{FC})m_2(\mathrm{FC}) + m_1(\mathrm{FC})m_2(\mathrm{OC}\cup\mathrm{FC}) + m_1(\mathrm{FC})m_2(\mathrm{FC}\cup\mathrm{UC}) \\
&+ m_1(\mathrm{FC})m_2(\mathrm{FC}\cup\mathrm{UC}\cup\mathrm{OC}) + m_1(\mathrm{OC}\cup\mathrm{FC})m_2(\mathrm{FC}) \\
&+ m_1(\mathrm{OC}\cup\mathrm{FC})m_2(\mathrm{FC}\cup\mathrm{UC}) + m_1(\mathrm{FC}\cup\mathrm{UC})m_2(\mathrm{FC}) \\
&+ m_1(\mathrm{FC}\cup\mathrm{UC})m_2(\mathrm{OC}\cup\mathrm{FC}) + m_1(\mathrm{FC}\cup\mathrm{UC}\cup\mathrm{OC})m_2(\mathrm{FC})\}/0.97 \\
= \{&0.05\times0.15 + 0.05\times0.05 + 0.05\times0.2 + 0.05\times0.5 + 0.15\times0.15 \\
&+ 0.15\times0.2 + 0.1\times0.15 + 0.1\times0.05 + 0.6\times0.15\}/0.97 = 0.21.
\end{aligned}
$$

Similarly, for the combined event FC ∪ UC, we get

$$
\begin{aligned}
m_{12}(\mathrm{FC}\cup\mathrm{UC}) = \{&m_1(\mathrm{FC}\cup\mathrm{UC})m_2(\mathrm{FC}\cup\mathrm{UC}) + m_1(\mathrm{FC}\cup\mathrm{UC})m_2(\mathrm{FC}\cup\mathrm{OC}\cup\mathrm{UC}) \\
&+ m_1(\mathrm{FC}\cup\mathrm{OC}\cup\mathrm{UC})m_2(\mathrm{FC}\cup\mathrm{UC})\}/0.97 \\
= \{&0.1\times0.2 + 0.1\times0.5 + 0.6\times0.2\}/0.97 = 0.20.
\end{aligned}
$$

Finally, using the combined evidence measures, m_{12}, we can calculate the combined belief measures (bel_{12}). For example, for OC ∪ FC we have

$$
\mathrm{bel}_{12}(\mathrm{OC}\cup\mathrm{FC}) = m_{12}(\mathrm{OC}\cup\mathrm{FC}) + m_{12}(\mathrm{OC}) + m_{12}(\mathrm{FC}) = 0.12 + 0.01 + 0.21 = 0.34.
$$

The remaining calculated values are shown in Table 13.2.

Probability Measures

When the additional belief axiom (Equation (13.5)) is replaced with a stronger axiom (illustrated for only two sets, A and B),

$$
\mathrm{bel}(\mathrm{A}\cup\mathrm{B}) = \mathrm{bel}(\mathrm{A}) + \mathrm{bel}(\mathrm{B}),\ \ \mathrm{A}\cap\mathrm{B} = \varnothing \tag{13.20}
$$

we get a *probability measure*. Let us now introduce a formal definition for a probability measure in the context of an evidence theory.

If we have a bea for a singleton, x, denoted $m(x) = \mathrm{bel}(x)$, and we have $m(\mathrm{A}) = 0$ for all subsets A of the power set, P(X), that are *not* singletons, then $m(x)$ is a probability measure. A probability measure is also a mapping of some function, say $p(x)$, to the unit interval, that is,

$$
p : x \rightarrow [0,\ 1]. \tag{13.21}
$$

To conform to the literature, we will let $m(x) = p(x)$ to denote $p(x)$ as a probability measure. The mapping $p(x)$ then maps evidence only on singletons to the unit interval. The key distinction between a probability measure and either a belief or plausibility measure, as can be seen from Equation (13.21), is that a probability measure arises when all the evidence is only on singletons, that is, only on elements x, whereas when we have some evidence on subsets that are not singletons, we cannot have a probability measure and will have only belief and plausibility measures (both, because they are duals, see Equation (13.10)). If we have a probability measure, we will then have

$$\mathrm{bel}(\mathrm{A}) = \mathrm{pl}(\mathrm{A}) = p(\mathrm{A}) = p(\mathrm{A}) = \sum_{x \in \mathrm{A}} p(x), \qquad \text{for all } \mathrm{A} \in \mathrm{P}(\mathrm{X}), \tag{13.22}$$

where set A is simply a collection of singletons; this would define the probability of set A. Equation (13.22) reveals that the belief, plausibility, and probability of a set A are all equal for a situation involving probability measures. Moreover, Equations (13.7) and (13.11) become a manifestation of the excluded middle axioms (Chapter 2) for a probability measure:

$$\mathrm{pl}(\mathrm{A}) = p(\mathrm{A}) = \mathrm{bel}(\mathrm{A}) \rightarrow p(\mathrm{A}) + p(\overline{\mathrm{A}}) = 1. \tag{13.23}$$

Example 13.3

Two quality control experts from PrintLaser, Inc., are trying to determine the source of scratches on the media that exit the sheet feeder of a new laser printer already in production. One possible source is the upper arm and the other source is media sliding on top of other media (e.g., paper on paper). We shall denote the following focal elements:

W denotes scratches from wiper arm.
M denotes scratches from other media.

The experts provide their assessments of evidence supporting each of the focal elements as follows:

Focal elements	Expert 1, m_1	Expert 2, m_2
W	0.6	0.3
M	0.4	0.7
W ∪ M	0	0

We want to determine the beliefs, plausibilities, and probabilities for each non-null focal element. We can see that evidence is available only on the singletons, W and M. We find the following relationships for the first expert:

$$\mathrm{bel}_1(\mathrm{W}) = m_1(\mathrm{W}) = 0.6,$$

$$\mathrm{bel}_1(\mathrm{M}) = m_1(\mathrm{M}) = 0.4,$$

$$\mathrm{bel}_1(\mathrm{W} \cup \mathrm{M}) = m_1(\mathrm{W}) + m_1(\mathrm{M}) + m_1(\mathrm{W} \cup \mathrm{M}) = 0.6 + 0.4 + 0 = 1,$$

$$\mathrm{pl}_1(\mathrm{W}) = m_1(\mathrm{W}) + m_1(\mathrm{W} \cup \mathrm{M}) = 0.6 + 0 = 0.6,$$

$$\mathrm{pl}_1(\mathrm{M}) = m_1(\mathrm{M}) + m_1(\mathrm{W} \cup \mathrm{M}) = 0.4 + 0 = 0.4,$$

$$\mathrm{pl}_1(\mathrm{W} \cup \mathrm{M}) = m_1(\mathrm{W}) + m_1(\mathrm{M}) + m_1(\mathrm{W} \cup \mathrm{M}) = 0.6 + 0.4 + 0 = 1.$$

We note that bel_1 (W) = pl_1 (W), bel_1 (M) = pl_1 (M), and bel_1 (W ∪ M) = pl_1 (W ∪ M). From Equation (13.23), these are all probabilities. Hence, p_1 (W) = 0.6, p_1 (M) = 0.4, and p_1 (W ∪ M) = p(W) + p(M) = 0.6 + 0.4 = 1 (this also follows from the fact that the probability of the union of disjoint events is the sum of their respective probabilities). In a similar fashion for the second expert we find

$$p_2(\mathrm{W}) = 0.3, \quad p_2(\mathrm{M}) = 0.7 \text{ and } p_2(\mathrm{W} \cup \mathrm{M}) = 0.3 + 0.7 = 1.$$

Possibility and Necessity Measures

Suppose we have a collection of some or all of the subsets on the power set of a universe, which have the property $\mathrm{A}_1 \subset \mathrm{A}_2 \subset \mathrm{A}_3 \subset \cdots \subset \mathrm{A}_n$. With this property, these sets are said to be *nested* (Shafer, 1976). When the elements of a set, or universe, having evidence are nested, we say that the belief measures, bel(A_i), and the plausibility measures, pl(A_i), represent a *consonant* body of evidence. By consonant we mean that the evidence allocated to the various elements of the set (subsets on the universe) does *not* conflict, that is, the evidence is free of dissonance.

For a consonant body of evidence, we have the following relationships (Klir and Folger, 1988) for two different sets on the power set of a universe, that is, for A, B ∈ P(X):

$$\mathrm{bel}(\mathrm{A} \cap \mathrm{B}) = \min[\mathrm{bel}(\mathrm{A}), \mathrm{bel}(\mathrm{B})], \tag{13.24}$$

$$\mathrm{pl}(\mathrm{A} \cup \mathrm{B}) = \max[\mathrm{pl}(\mathrm{A}), \mathrm{pl}(\mathrm{B})]. \tag{13.25}$$

The expressions in Equations (13.24) and (13.25) indicate that the belief measure of the intersection of two sets is the smaller of the belief measures of the two sets and the plausibility measure of the union of these two sets is the larger of the plausibility measures of the two sets.

In the literature consonant belief and plausibility measures are referred to as *necessity or certainty* (denoted η) and *possibility* (denoted π) measures, respectively. Equations (13.24) and (13.25) become, respectively, for all A, B ∈ P(X),

$$\eta(\mathrm{A} \cap \mathrm{B}) = \min[\eta(\mathrm{A}), \eta(\mathrm{B})], \tag{13.26}$$

$$\pi(\mathrm{A} \cup \mathrm{B}) = \max[\pi(\mathrm{A}), \pi(\mathrm{B})]. \tag{13.27}$$

For a consonant body of evidence, the dual relationships expressed in Equation (13.10) then take the forms,

$$\begin{aligned} \pi(\mathrm{A}) &= 1 - \eta(\overline{\mathrm{A}}), \\ \eta(\mathrm{A}) &= 1 - \pi(\overline{\mathrm{A}}). \end{aligned} \tag{13.28}$$

Because the necessity (certainty) and possibility measures are dual relationships, the discussion to follow focuses only on one of these, possibility. If necessity measures are desired, they can always be derived with the expressions in Equation (13.28).

We now define a possibility distribution function as a mapping of the singleton elements, x, in the universe, X, to the unit interval, that is,

$$r : \mathrm{X} \rightarrow [0,1]. \tag{13.29}$$

This mapping will be related to the possibility measure, $\pi(\mathrm{A})$, through the relationship

$$\pi(\mathrm{A}) = \max_{x \in \mathrm{A}} r(x), \tag{13.30}$$

for each $\mathrm{A} \in \mathrm{P(X)}$ (see Klir and Folger, 1988, for a proof). Now, a possibility distribution can be defined as an ordered sequence of values,

$$\mathbf{r} = (\rho_1, \rho_2, \rho_3, \ldots, \rho_n), \tag{13.31}$$

where $\rho_i = r(x_i)$ and where $\rho_i \geq \rho_j$ for $i < j$. The *length* of the ordered possibility distribution given in Equation (13.31) is the number n. Also, every possibility measure can be characterized by the n-tuple, denoted as a basic distribution (Klir and Folger, 1988),

$$\mathbf{m} = (\mu_1, \mu_2, \mu_3, \ldots, \mu_n), \tag{13.32a}$$

$$\sum_{i=1}^{n} \mu_i = 1, \tag{13.32b}$$

where $\mu_i \in [0, 1]$ and $\mu_i = m(\mathrm{A}_i)$. Of course, the sets A_i are nested as is required of all consonant bodies of evidence. From Equation (13.17) and the relationship

$$\rho_i = r(x_i) = \pi(x_i) = \mathrm{pl}(x_i), \tag{13.33}$$

it can be shown (Klir and Folger, 1988) that

$$\rho_i = \sum_{k=i}^{n} \mu_k = \sum_{k=i}^{n} m(\mathrm{A}_k) \tag{13.34}$$

or, in a recursive form,

$$\mu_i = \rho_i - \rho_{i+1}, \tag{13.35}$$

where $\rho_{n+1} = 0$ by convention. Equation (13.35) produces a set of equations of the form

$$\begin{array}{rlccccccccc}
\rho_1 & = \mu_1 & + & \mu_2 & + & \mu_3 & + & \cdots & + & \mu_n, \\
\rho_2 & = & & \mu_2 & + & \mu_3 & + & \cdots & + & \mu_n, \\
\rho_3 & = & & & & \mu_3 & + & \cdots & + & \mu_n, \\
 & = & & & & & & \cdots & & \\
\rho_n & = & & & & & & & & \mu_n.
\end{array} \tag{13.36}$$

Nesting of focal elements can be an important physical attribute of a body of evidence. Consider the following example where physical nesting is an important feature of an engineering system.

Example 13.4

Suppose there are seven nodes in a communication network X, labeled $x_1 - x_7$ and represented by boxes. Of these seven nodes, one is causing a problem. The company network expert is asked for an opinion on which node is causing the communications problem. The network expert aggregates these nodes into sets, as given in the accompanying table. The third column in the table represents the expert's basic distribution (beas) and the last column is the possibility distribution found from Equation (13.34).

Set A	Aggregation of focal elements	$\mu_n = \mu\ (A_n)$	ρ_i
A_1	x_1	0.4	1
A_2	$x_1 \cup x_2$	0.2	0.6
A_3	$x_1 \cup x_2 \cup x_3$	0	0.4
A_4	$x_1 \cup x_2 \cup x_3 \cup x_4$	0.1	0.4
A_5	$x_1 \cup x_2 \cup x_3 \cup x_4 \cup x_5$	0	0.3
A_6	$x_1 \cup x_2 \cup x_3 \cup x_4 \cup x_5 \cup x_6$	0.2	0.3
A_7	$x_1 \cup x_2 \cup x_3 \cup x_4 \cup x_5 \cup x_6 \cup x_7$	0.1	0.1

The physical significance of this nesting (shown in Figure 13.4) can be described as follows. In the network expert's belief, node x_1 is causing the problem. This node has new hardware and is an experimental CPU. For these reasons, the network expert places the highest belief on this set (set A_1). The next set with nonzero belief (supporting evidence), A_2, comprises the union, x_1 or x_2. The network expert has less belief that node x_1 or x_2 is causing the problem. Node x_2 has new hardware as well, but has a trusted CPU. The next set with nonzero belief, A_4, is nodes x_1 or x_2 or x_3 or x_4. The network expert has even less belief that the problem is caused by this set. The expert reasons that x_3 or x_4 has trusted hardware and CPUs. The next set with nonzero belief, A_6, is nodes x_1 or x_2 or x_3 or x_4 or x_5 or x_6. The network expert has slightly more belief that this set is the problem than set A_4, but much less than the initial set A_1, the reasons being that there are two new programmers using these nodes for testing communications software. The final set with evidence is the union of all seven nodes. The expert has little belief that this set is the problem because node x_7 is usually turned off.

Note that the first element of any ordered possibility distribution, ρ_1, is always equal to unity, that is, $\rho_1 = 1$. This fact is guaranteed by Equation 13.32b. The smallest possibility distribution of length n has the form $\mathbf{r} = (1, 0, 0, 0, \ldots, 0)$, where there are $(n - 1)$ zeros after a value of unity in the distribution. The associated basic distribution would have the form $\mathbf{m} = (1, 0, 0, 0, \ldots, 0)$. In this case there would be only one focal element with evidence, and it would have all the

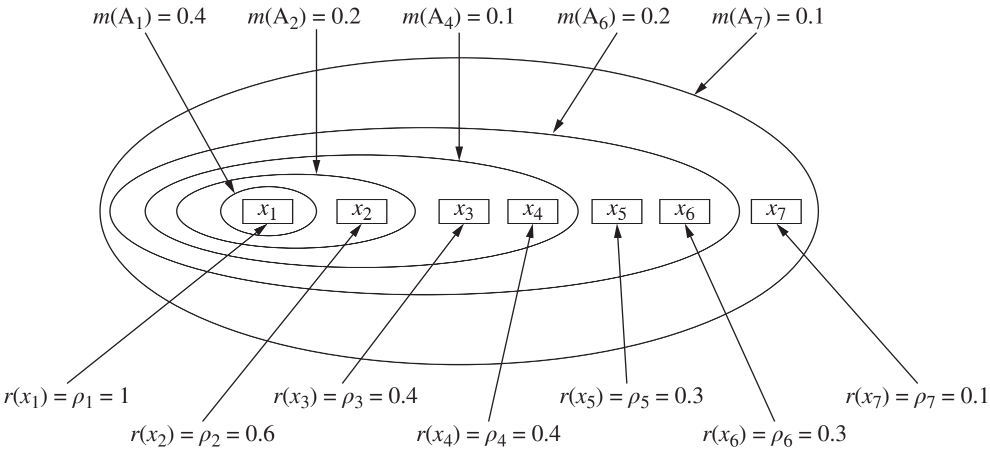

Figure 13.4 Nesting diagram, Example 13.4.

evidence. This situation represents *perfect evidence*; there is no uncertainty involved in this case.

Alternatively, the largest possibility distribution of length n has the form $\mathbf{r} = (1, 1, 1, 1, \ldots, 1)$, where all values are unity in the distribution. The associated basic distribution would have the form $\mathbf{m} = (0, 0, 0, 0, \ldots, 1)$. In this case all the evidence is on the focal element comprising the entire universe, that is, $A_n = x_1 \cup x_2 \cup \cdots \cup x_n$; hence, we know nothing about any specific focal element in the universe except the universal set. This situation is called *total ignorance*. In general, the larger the possibility distribution, the less specific the evidence and the more ignorant we are of making any conclusions.

Because possibility measures are special cases of plausibility measures and necessity measures are special cases of belief measures, we can relate possibility measures and necessity measures to probability measures. Equation (13.23) shows that, when all the evidence in a universe resides solely on the singletons of the universe, the belief and plausibility measures become probability measures. In a similar fashion it can be shown that the plausibility measure approaches the probability measure from an upper bound and that the belief measure approaches the probability measure from a lower bound; the result is a range around the probability measure (Yager and Filev, 1994),

$$\text{bel}(A) \le p(A) \le \text{pl}(A). \tag{13.37}$$

Example 13.5

Probabilities can be determined by finding point-valued quantities and then determining the relative frequency of occurrence of these quantities. In determining the salvage value of older computers, the age of the computer is a key variable. This variable is also important in assessing depreciation costs for the equipment. Sometimes there is uncertainty in determining the age of computers if their purchase records are lost or if the equipment was acquired through secondary acquisitions or trade. Suppose the age of five computers is known, and we have no uncertainty; here the ages are point-valued quantities.

Computer	Age (months)
1	26
2	21
3	33
4	24
5	30

With the information provided in the table we could answer the following question: What percentage of the computers have an age in the range of 20–25 months, that is, what percentage of the ages fall in the interval [20, 25]? This is a countable answer of $\frac{2}{5}$, or 40%.

Now suppose that the age of the computers is not known precisely, but rather each age is assessed as an interval. Now the ages are set-valued quantities, as follows:

Computer	Age (months)
1	[22, 26]
2	[20, 22]
3	[30, 35]
4	[20, 24]
5	[28, 30]

With this information we can only assess possible solutions to the question just posed: What percentage of the computers possibly fall in the age range of [20, 25] months? Because the ages of the computers are expressed in terms of ranges (or sets on the input space), the solution space of percentages will also have to be expressed in terms of ranges (or sets on the solution space).

To approach the solution we denote the query range as Q, that is, Q = [20, 25] months. We denote the age range of the ith computer as D_i. Now, we can determine the certainty and possibility ranges using the following rules:

1. Age (i) is certain if $D_i \subset \mathrm{Q}$.
2. Age (i) is possible if $D_i \cap \mathrm{Q} \neq \emptyset$.
3. Age (i) is not possible if $D_i \cap \mathrm{Q} = \emptyset$.

The first rule simply states that the age is certainly in the query range if the age range of the ith computer is completely contained within the query range. The second rule states that the age is possibly in the query range if the age range of the ith computer and the query range have a non-null intersection, that is, if they intersect at any age. The third rule states that the age is not possible if the age range of the ith computer and the query range have no age in common, that is, their intersection is null. We should note here that a solution that is certain is necessarily possible (certainty implies possibility), but the converse is not always true (things that are possible are not always certain). Hence, the set of certain quantities is a subset of the set of possible quantities. In looking at the five computers and using the three rules already given, we determine the following relationships:

Computer	η or π
1	Possible
2	Certain
3	Not possible
4	Certain
5	Not possible

In the table we see that, of the five computers, two have age ranges that are certainly (denoted η (Q)) in the query interval and three have age ranges that are possibly (denoted π (Q)) in the query interval (one possible and two certain). We will denote the solution as the response to the query, or resp(Q). This will be an interval-valued quantity as indicated previously, or

$$\text{resp}(\text{Q}) = [\eta(Q),\ \pi(Q)] = \left[\frac{2}{5}, \frac{3}{5}\right].$$

Hence, we can say that the answer to the query is "Certainly 40% and possibly as high as 60%." We can also see that the range represented by resp(Q) represents a lower bound and an upper bound to the actual point-valued probability (which was determined to be 40%) as indicated in Equation (13.37).

In Example 13.5, all computer ranges were used in determining the percentages for possibilities and certainties (i.e., all five). In evidence theory, null values are *not* counted in the determination of the percentages as seen in the normalization constant, K, expressed in Equation (13.19). This characteristic can lead to fallacious responses, as illustrated in the following example.

Example 13.6

Suppose again we wish to determine the age range of computers, this time expressed in units of years. In this case, we ask people to tell us the age of their own computer (PC). These responses are provided in the accompanying table. In the table, the null symbol, Ø, indicates that the person queried has no computer.

Person	Age of PC (years)
1	[3, 4]
2	Ø
3	[2, 3]
4	Ø
5	Ø

Let us now ask the following question: What percentage of the computers has an age in the range Q = [2, 4] years? Using the rules given in Example 13.5, we see that we have two certainties (hence, we have two possibilities) and three null values. If we include the null values in our count, the solution is

$$\mathrm{resp}(\mathrm{Q}) = [\eta(Q), \pi(Q)] = \left[\frac{2}{5}, \frac{2}{5}\right] = \frac{2}{5}.$$

In this case, we have not used a normalization process because we have counted the null values. If we decide to neglect the null values (hence, we normalize as Dempster's rule of combination suggests), then the solution is

$$\mathrm{resp}(\mathrm{Q}) = [\eta(Q), \pi(Q)] = \left[\frac{2}{2}, \frac{2}{2}\right] = 1.$$

We see a decidedly different result when normalization is used.

A graphical interpretation of the evidence theory developed by Dempster and Shafer is provided in what is called the *ball–box analogy* (Zadeh, 1986).

Example 13.7 (Zadeh 1984).
Suppose the king of country X believes a submarine, S, is in the territorial waters of X. The king summons n experts, $E_1, E_2, \ldots, E_n$, to give him advice on the location of the submarine, S. The n experts each provide their assessment of the location of S; call these possible locations $L_1, L_2, \ldots, L_m, \ldots, L_n$, where $m \le n$. Here, L_i are subsets of the territorial waters, X. To be more specific, experts $E_1, E_2, \ldots, E_m$ say that S is in $L_1, L_2, \ldots, L_m$ and experts $E_{m+1}, \ldots, E_{n-1}, E_n$ say that S is *not* located in the territorial waters of X, that is, $L_{m+1} = L_{m+2} = \cdots = L_n = \emptyset$. So, there are $(m - n)$ experts who say that S is not in the territorial waters. Now the king asks, "Is S in a *subset* A of our territorial waters?" Figure 13.5 shows possible location regions, L_i, and the query region of interest, region A.

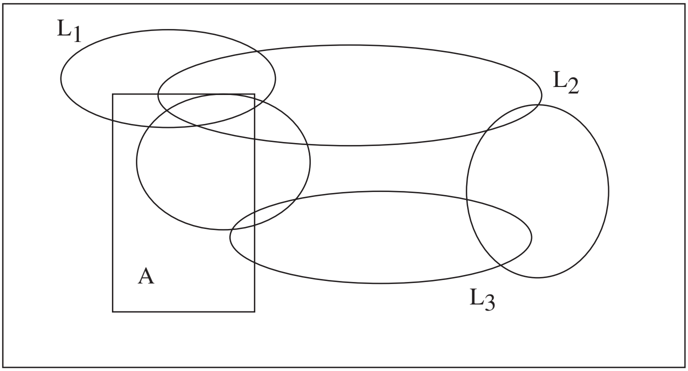

Figure 13.5 Ball–box analogy of the Dempster–Shafer evidence theory.

If we denote T_i as the location proffered by the ith expert, we have the following two rules:

1. $T_i \subset A$ implies that it is certain that $S \in A$.
2. $T_i \cap A \neq \emptyset$ implies that it is possible that $S \in A$.

We note again that a certainty is contained in the set of possibilities, that is, certainty implies possibility. We further assume that the king aggregates the opinions of his experts by averaging. Thus, if k out of n experts vote for rule number 1, then the average certainty = k/n; if l out of n (where $l \geq k$) experts vote for rule number 2, then the average possibility = l/n. Finally, if the judgment of those experts who think there is no submarine anywhere in the territorial waters is ignored, the average certainty and possibility will be k/m and l/m, respectively (where $m \leq n$). Ignoring the opinion of those experts whose L_i is the null set corresponds to the normalization (Equation (13.19)) in the Dempster–Shafer evidence theory.

Dempster's rule of combination may lead to counterintuitive results because of the normalization issue. The reason for this (Zadeh, 1984) is that normalization throws out evidence that asserts that the object under consideration does not exist, that is, is null or empty (∅). The following example from the medical sciences illustrates this idea effectively.

Example 13.8 (Zadeh, 1984).
A patient complaining of a severe headache is examined by two doctors (doctor 1 and doctor 2). The diagnosis of doctor 1 is that the patient has either meningitis (M) with probability 0.99 or a brain tumor (BT) with probability 0.01. Doctor 2 agrees with doctor 1 that the probability of a brain tumor (BT) is 0.01, but disagrees with doctor 1 on the meningitis; instead doctor 2 feels that there is a probability of 0.99 that the patient just has a concussion (C). The following table shows the evidence for each of the focal elements in this universe for each of the doctors (m_1 and m_2) as well as the calculated values for the combined evidence measures (m_{12}). In the table there is evidence only on the singletons M, BT, and C; hence, all the measures are probability measures (the doctors provided their opinions in terms of probability).

Focal element	m_1	m_2	m_{12}
M	0.99	0	0
BT	0.01	0.01	1
C	0	0.99	0
M ∪ BT	0	0	
M ∪ C	0	0	
BT ∪ C	0	0	
M ∪ BT ∪ C	0	0	

The combined evidence measures are calculated as follows. First, by using Equation (13.19) the normalization constant, K, is calculated. Then use of Equation (13.18) produces values for m_{12}. For example, m_{12} for a brain tumor is found as

$$m_{12}(\mathrm{BT}) = \frac{0.01(0.01)}{1 - \{0.99[0.01 + 0.99] + 0.01[0 + 0.99] + 0[0 + 0.01]\}} = \frac{0.0001}{1 - 0.9999} = 1.$$

One can readily see that the combined measures m_{12}(C) and m_{12}(M) will be zero from the fact that

$$m_1(\mathrm{C}) = m_2(\mathrm{M}) = 0.$$

The table reveals that using the Dempster–Shafer rule of combination (in Shafer, 1976, null values are not allowed in the definition of belief functions but do enter in the rule of combination of evidence) results in a combined probability of 1 that the patient has a brain tumor when, in fact, both doctors agreed individually that it was only one chance in a hundred! What is even more confusing is that the same conclusion (i.e., $m_{12}(\mathrm{BT}) = 1$) results regardless of the probabilities associated with the other possible diagnoses.

In Example 13.8, it appears that the normalization process *suppressed* expert opinion; but, is this omission mathematically allowable? This question leads to the conjecture that the rule of combination cannot be used until it is ascertained that the bodies of evidence are conflict free; that is, at least one parent relation exists that is absent of conflict. In particular, under this assertion, it is not permissible to combine distinctly different bodies of evidence. In the medical example, the opinions of both doctors reveal some missing information about alternative diagnoses. A possibility theory might suggest an alternative approach to this problem in which the incompleteness of information in the knowledge base propagates to the conclusion and results in an interval-valued, possibilistic answer. This approach addresses rather than finesses (like excluding null values) the problem of incomplete information.

Possibility Distributions as Fuzzy Sets

Belief structures that are nested are called *consonant*. A fundamental property of consonant belief structures is that their plausibility measures are possibility measures. As suggested by Dubois and Prade (1988), possibility measures can be seen to be formally equivalent to fuzzy sets. In this equivalence, the membership grade of an element x corresponds to the plausibility of the singleton consisting of that element, x; that is, a consonant belief structure is equivalent to a fuzzy set F of X where $\mathrm{F}(x) = \mathrm{pl}(\{x\})$.

A problem in equating consonant belief structures with fuzzy sets is that the combination of two consonant belief functions using Dempster's rule of combination in general does not necessarily lead to a consonant result (Yager, 1993). Hence, because Dempster's rule is essentially a conjunction operation, the intersection of two fuzzy sets interpreted as consonant belief structures may not result in a valid fuzzy set (i.e., a consonant structure).

Example 13.9 (Yager, 1993).
Suppose we have a universe comprising five singletons, that is,

$$X = \{x_1,\ x_2,\ x_3,\ x_4,\ x_5\},$$

and we have evidence provided by two experts. The accompanying table provides the experts' degrees of belief about specific subsets of the universe, X.

Focal elements	Expert 1, m_1 (A_i)	Expert 2, m_2 (B_i)
$A_1 = \{x_1, x_2, x_3\}$	0.7	
$A_2 = X$	0.3	
$B_1 = \{x_3, x_4, x_5\}$		0.8
$B_2 = X$		0.2

Because $A_1 \subset A_2$ and $B_1 \subset B_2$, we have two consonant (nested) belief structures represented by A and B. Using Dempster's rule of combination and applying Equations (13.18) and (13.19), we have for any set D on the universe X

$$m(D) = \frac{1}{1-k} \sum_{A_i \cap B_j = D} m_1(A_i) \cdot m_2(B_j)$$

and

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) \cdot m_2(B_j).$$

Because there are two focal elements in each experts' belief structures, we will have $2^2 = 4$ belief structures in the combined evidence case, which we will denote as m. We note for these data that we get a value of $k = 0$, because there are no intersections between the focal elements of A and B that result in the null set. For example, the intersection between A_1 and B_1 is the singleton, x_3. Then, we get

$$D_1 = A_1 \cap B_1 = \{x_3\} \qquad m(D_1) = 0.56 \quad \text{(i.e., } 0.7 \times 0.8\text{)},$$

$$D_2 = A_1 \cap B_2 = \{x_1, x_2, x_3\} \quad m(D_2) = 0.14 \quad \text{(i.e., } 0.7 \times 0.2\text{)},$$

$$D_3 = A_2 \cap B_1 = \{x_3, x_4, x_5\} \quad m(D_3) = 0.24 \quad \text{(i.e., } 0.3 \times 0.8\text{)},$$

$$D_4 = A_2 \cap B_2 = X \qquad m(D_4) = 0.06 \quad \text{(i.e., } 0.3 \times 0.2\text{)}.$$

For the focal elements D_i, we note that $D_1 \subset D_2 \subset D_4$ and $D_1 \subset D_3 \subset D_4$, but we do not have $D_2 \subset D_3$ or $D_3 \subset D_2$. Hence, the combined case is not consonant (i.e., not completely nested).

Yager (1993) has developed a procedure to prevent the situation illustrated in Example 13.9 from occurring; that is, a method is available to combine consonant possibility measures where the result is also a consonant possibility measure. However, this procedure is lengthy to describe and is beyond the scope of this text; the reader is referred to the literature (Yager, 1993) to learn this method.

Another interpretation of a possibility distribution as a fuzzy set was proposed by Zadeh (1978). He defined a possibility distribution as a fuzzy restriction that acts as an elastic constraint on the values that may be assigned to a variable. In this case the possibility distribution represents the degrees of membership for some linguistic variable, but the membership values are strictly monotonic as they are for an ordered possibility distribution. For example, let $\underset{\sim}{A}$ be a fuzzy set on a universe X, and let the membership value, μ, be a variable on X that assigns a "possibility" that an element of x is in $\underset{\sim}{A}$. So, we get

$$\pi(x) = \mu_{\underset{\sim}{A}}(x). \tag{13.38}$$

Zadeh points out that the possibility distribution is nonprobabilistic and is used primarily in natural language applications. There is a loose relationship, however, between the two through a possibility/probability consistency principle (Zadeh, 1978). In sum, what is *possible* may not be *probable*, but what is *impossible* is inevitably *improbable* (see the discussion on this issue later in this chapter).

Example 13.10

Let $\underset{\sim}{A}$ be a fuzzy set defined on the universe of columns needed to support a building. Suppose there are 10 columns altogether, and we start taking columns away until the building collapses; we record the number of columns at the time the building collapses. Let $\underset{\sim}{A}$ be the fuzzy set defined by the number of columns "possibly" needed, out of 10 total, just before the structure fails. The structure most certainly needs at least three columns to stand (imagine a stool). After that the number of columns required for the building to stand is a fuzzy issue (because of the geometric layout of the columns, the weight distribution, etc.), but there is a possibility it may need more than three. The following fuzzy set may represent this possibility:

$$\underset{\sim}{A} = \left\{\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{0.9}{4} + \frac{0.6}{5} + \frac{0.3}{6} + \frac{0.1}{7} + \frac{0}{8} + \frac{0}{9} + \frac{0}{10}\right\}.$$

A probability distribution on the same universe may look something like the following table, where u is the number of columns prior to collapse and $p(u)$ is the probability that u is the number of columns at collapse:

u	1	2	3	4	5	6	7	8	9	10
$p(u)$	0	0	0.1	0.5	0.3	0.1	0	0	0	0

As seen, although it is *possible* that one column will sustain the building, it is not *probable*. Hence, neither does a high degree of possibility imply a high probability nor does a low degree of probability imply a low degree of possibility.

Belief measures and plausibility measures overlap when they both become probability measures. However, possibility, necessity, and probability measures do not overlap with one another except for one special measure: the measure of one focal element that is a singleton. These three measures become equal when one element of the universal set is assigned a value of unity, and all other elements in the universe are assigned a value of zero. This measure represents *perfect evidence* (Klir and Folger, 1988).

Possibility Distributions Derived from Empirical Intervals

Analyzing empirical data is an important exercise in any experimental analysis. In practice, it is common to use probabilistic tools to analyze data from experimental studies. However, there are a number of occasions when it is more appropriate to conduct possibilistic analysis than a probabilistic analysis. For example, in the determination of the residual strength of an existing bridge, one might have only subjective estimates from visual inspections of the bridge or limited information of the strength from nondestructive evaluation. In such cases, data are usually available as a range of numbers or intervals (Figure 13.6), and an analyst is required to determine the best possible estimate from such data. When data are available as a set of intervals, possibilistic analysis captures the true uncertainty in the interval without relying on predetermined distributions and by not requiring an analyst to come up with specific data (i.e., singletons). Possibility distributions capture the imprecision resulting from nonspecificity of the intervals by considering the entire length of the interval. This section describes an approach for developing possibility distributions from such empirical measurements.

As described in the previous section, when the intervals are nested (consonant intervals), possibility measures are plausibility measures as defined by evidence theory. Thus, for nested intervals, one can derive a possibility distribution by tracing a contour over all the nested intervals. In practice, however, intervals are seldom nested and are usually available as sets of overlapping and nonoverlapping portions. In such cases, special methods are required to transform the nonconsonant intervals into consonant intervals in accordance with the available evidence.

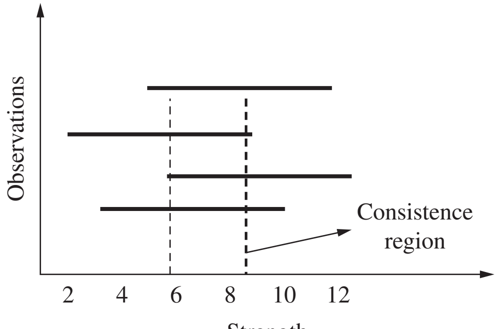

Figure 13.6 Residual strength estimates acquired as “consistent” intervals.

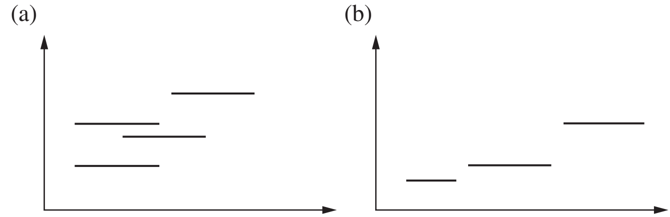

Figure 13.7 (a) Overlapping but inconsistent intervals; (b) completely disjoint intervals.

The nonconsonant intervals can be consistent, where at least one common interval exists among all the measurements (Figure 13.6); partially overlapping, where some intervals overlap (Figure 13.7a); or disjoint, where there is no overlap among the intervals (Figure 13.7b). The next section describes a method for developing possibility distributions for more common cases where the intervals are consistent and partially overlapping.

Deriving Possibility Distributions from Overlapping Intervals

Let us consider a system that can be described within a domain $X = \{x_1, x_2, \ldots, x_n\}$; the behavior of which is described by evidence obtained as observations over a collection of sets, F. The set X is called the *domain* of the system. For example, the domain of the system can be all the possible residual strengths of the bridge under consideration for maintenance. Now, let $F = \{\langle A_j,\ w_j\rangle\}$ represent the original intervals along with their weights. If M measurements are observed, then the weights w_j of each observation A_j is calculated by frequency analysis as (Donald, 2003)

$$w_j = \frac{n(A_j)}{M}, \tag{13.39}$$

where $n\ (A_j)$ is the frequency count of interval A_j, and $\sum_j w_j = 1,\ \sum_j n(A_j) = M.$

In conventional probability analysis, the observations A_1 and A_2 are disjoint such that $A_1 \cap A_2 = \emptyset$, and the frequency of occurrence of any event A_j is simply the ratio of the count of a particular event to the total number of occurrences of all the events. However, such disjoint measurements are uncommon and measurements are usually such that F consists of overlapping intervals A_j. In deriving a set of consonant intervals, it is assumed that the underlying evidence is coherent such that the observations obtained should reveal intervals that are nested within each other. This is accomplished such that the smallest interval (having the least nonspecificity) is selected as the interval that has the most intersections with the observations and the next smallest one as that having the second-most intersections, and so forth. This process yields Q unique countable intersections and a set $G = \{\langle B_k,\ \nu(B_k)\rangle\}$, $k = 1, 2, \ldots, Q$ of intersections, where B_k is the interval obtained by the kth intersection of original intervals A_i and A_j, and $\nu(B_k)$ is the weight assigned to the corresponding B_k. Union of all the sets, $S = \cap_{A_j \in F} A_j$, where $w\ (A_j) \neq 0$, forms the support interval and is derived as

$$S=\left[\min_{A_i\in F,\ x\in A_i}(x),\ \max_{A_i\in F,\ x\in A_i}(x)\right]. \tag{13.40}$$

Weights ν_k assigned to each of the Q elements of G are determined by using any of the t-norms (Chapter 2) for conjunctions as $\nu(B_k) = \mathrm{T}(A_i, A_j)_k$, where $\mathrm{T}(A_i, A_j)_k$ is any conjunctive triangular norm operating on the sets A_i and A_j that form the kth intersection. In this section, the standard fuzzy intersection (t-norm, minimum operator) is used to determine weights $\nu(B_k)$ of the conjunction of sets A_i and A_j, and therefore the weights are obtained as

$$v(B_k)=\min\left(w_i,w_j\right)_k. \tag{13.41}$$

Because a min t-norm is based on a weaker axiom of nonadditivity, the weights do not necessarily add to 1, and hence the weights derived are normalized as

$$\eta(B_k)=\frac{v(B_k)}{\sum_{k=1}^{Q},v(B_k)}, \tag{13.42}$$

such that

$$\sum_k \eta(B_k)=1.$$

In deriving the elements of G, however, some elements by virtue of their origin from the intersections from parent sets (sets from which the intervals were derived) tend to be consonant with the parent set while they are not necessarily consonant with other parent sets. Therefore, if Q represents the total number of focal elements (intervals with nonzero weights) in the focal set G derived from the intersections of original measurements, there are Q_{H} elements in the consonant set H, and Q_I elements in the nonconsonant set I, such that

$$\sum_{B_k\in Q_{\mathrm{H}}}\eta(B_k)+\sum_{B_k\in Q_{\mathrm{I}}}\eta(B_k)=1.0. \tag{13.43}$$

Consonant intervals can be extracted from the intersection set G as a combination of frequency analysis and expert judgments. The most possible interval is the one that occurs most frequently, and in which an expert has the most confidence. Once the task of selecting consonant intervals is accomplished, the weight from the remaining nonconsonant intervals is redistributed to the consonant intervals in such a manner that the total information from underlying evidence is preserved.

Redistributing Weight from Nonconsonant to Consonant Intervals

Weights are redistributed from nonconsonant to consonant sets according to the dissonance between individual intersecting sets (the conflict between two sets). The logic used here is that the higher the similarity between two sets (or the lower the conflict), the greater the weight that can be transferred between the two sets. Parameters that are useful in the redistribution of weight are identified as the cardinality of each set, $|H_i|$, and the number of common elements between the sets $|H_i \cap I_j|$. In the case of continuous intervals, the cardinality $|.|$ can be replaced by the length l of the interval defined over a real line, and the set of real numbers comprising the intersecting interval can be determined as the length that is common to both the intervals. The similarity, β, of two sets is given as

$$\beta_{ij} = \frac{|H_i \cap I_j|}{|H_i|} \tag{13.44}$$

or, for a set of real numbers,

$$\beta_{ij} = \frac{l\left[\min \omega_i | \omega_i \in H_i \cap I_j \neq \emptyset, \max \omega_i | \omega_i \in H_i \cap I_j \neq \emptyset\right]}{l\,[H_i]}, \tag{13.45}$$

where $l[.]$ denotes the length of the interval and is simply given as $l[a, b] = b - a$, with $l[.]$ equal to 1 for singletons and $\beta_{ij} = 1$ when H_i is completely included in I_j.

A redistribution factor k is then computed as

$$\kappa_{ij} = \frac{\beta_{ij}}{\sum_{i=1}^{Q_H} \beta_{ij}}, \tag{13.46}$$

such that for any j,

$$\sum_{i=1}^{Q_H} \kappa_{ij} = 1.0.$$

The redistribution factor can be viewed as the fraction of the weight that is transferred from the nonconsonant to the consonant interval. The redistribution weight ρ is then calculated by determining the weight of the nonconsonant interval I_j that is transferred to the corresponding consonant interval H_i. Therefore, for the ith consonant interval and jth nonconsonant interval,

$$\rho_{ij} = \kappa_{ij}^{*} \eta(I_j). \tag{13.47}$$

From Equation (13.45), it is clear that when I_j intersects with H_i and no other set $\beta = 1$, thus assigning the entire weight of I to H, and also, when H does not intersect with any portion of I, $\beta = 0$. The final weights of the initial consonant sets as determined after the redistribution is given over the entire set of nonconsonant intervals as

$$m(H_i) = \eta(H_i) + \sum_{j=1}^{Q_I} \rho_{ij} = \eta(H_i) + \sum_{j=1}^{Q_I} \kappa_{ij} \eta(I_j). \tag{13.48}$$

The process indicates that the total weight is preserved among the consonant data intervals. It can be proved that $\sum_{i=1}^{Q_{\mathrm{H}}} m(\mathrm{H}_i) = 1.0$. From Equation (13.48) and with the constraint $\sum_{i=1}^{Q_{\mathrm{H}}} \kappa_{ij} = 1.0$, we have

$$\begin{aligned}\sum_{i=1}^{Q_{\mathrm{H}}} m(\mathrm{H}_i) &= \sum_{i=1}^{Q_{\mathrm{H}}} \eta(\mathrm{H}_i) + \sum_{i=1}^{Q_{\mathrm{H}}}\sum_{j=1}^{Q_{\mathrm{I}}} \kappa_{ij}\eta(\mathrm{I}_j) \\ &= \sum_{i=1}^{Q_{\mathrm{H}}} \eta(\mathrm{H}_i) + \sum_{j=1}^{Q_{\mathrm{I}}}\left(\eta(\mathrm{I}_j)\sum_{i=1}^{Q_{\mathrm{H}}} \kappa_{ij}\right) \\ &= \sum_{i=1}^{Q_{\mathrm{H}}} \eta(\mathrm{H}_i) + \sum_{j=1}^{Q_{\mathrm{I}}} \eta(\mathrm{I}_j) \\ &= 1.0.\end{aligned}$$

The possibility distribution from the weights is then obtained as follows:

$$\pi(x) = \sum_{x \in \mathrm{H}_i} m(\mathrm{H}_i). \tag{13.49}$$

Example 13.11

Suppose it is required to determine the strength of a wooden bridge that is subject to heavy foot traffic and an expert is hired to estimate the residual strength of the bridge. The expert uses various nondestructive techniques at various points of the bridge and offers estimates for the strength as shown in Table 13.3. On the basis of these estimates, an analyst needs to determine the possible strength of the bridge to decide on the appropriate action.

The first step in the solution is to determine the support of the distribution and the most possible interval based on the intersections of all the measurements. The support is calculated from Equation (13.40) as $S = [1000, 5000]$, and the intervals obtained by intersections of all the measurements are shown in Table 13.4.

From this table, it can be seen that the interval [3000, 4000] is included in all the intervals and, hence, is selected as the most possible interval. The selection of this most possible interval can also be based on other criteria that the expert chooses. The set of *consonant intervals* is then

Table 13.3 Original data intervals.

Estimate	Residual strength (lb in.$^{-2}$)	Weight
1	[1000, 4000]	0.2
2	[2000, 4000]	0.4
3	[3000, 5000]	0.2
4	[2000, 5000]	0.2

Table 13.4 Intervals obtained by intersections of original data intervals in Table 13.3.

Interval	Weight	Normalized weights
[3000, 4000]	0.2	0.143
[1000, 4000]	0.2	0.143
[2000, 4000]	0.4	0.285
[3000, 5000]	0.2	0.143
[2000, 5000]	0.2	0.143
[1000, 5000]	0.2	0.143

Table 13.5 Redistribution of weights between the nonconsonant and consonant intervals.

Nonconsonant intervals	Consonant intervals	β	k	ρ
[3000, 5000]	[3000, 4000]	1	0.316	0.045
	[2000, 4000]	1	0.316	0.045
	[2000, 5000]	0.66	0.209	0.03
	[1000, 5000]	0.5	0.159	0.023
[1000, 4000]	[3000, 4000]	1	0.267	0.038
	[2000, 4000]	1	0.267	0.038
	[2000, 5000]	1	0.267	0.038
	[1000, 5000]	0.75	0.199	0.029

determined from this most possible interval by choosing intervals in the order of increasing cardinality, which include the most possible interval such that the selected intervals are nested. Any intervals that are not part of this nested structure then form the elements for the nonconsonant set. The set of *consonant intervals* forms the intervals for the possibility distribution and the total weight of the evidence is then preserved by transferring the weight from the nonconsonant set to the consonant set (Equations (13.44) to (13.46)). Table 13.5 shows this redistribution of weights.

Redistribution parameters β, k, and ρ are determined using Equations (13.44) to (13.47) as

$$\beta_{11} = \frac{[3000, 5000] \cap [3000, 4000]}{[3000, 4000]} = 1,$$

$$\kappa_{11} = \frac{\beta_{11}}{\beta_{11} + \beta_{21} + \beta_{31} + \beta_{41}} = \frac{1}{3.16} = 0.316,$$

$$\rho_{11} = \kappa_{11}^{*} \eta(\mathrm{I}_1) = 0.316^{*}0.143 = 0.045.$$

The final weights are then determined using Equation (13.48) as

$$\begin{aligned} m[3000, 4000] &= 0.143 + \rho_{11} + \rho_{12} \\ &= 0.143 + 0.045 + 0.038 = 0.226. \end{aligned}$$

Table 13.6 Possibility intervals and final weights.

Interval	Final weight	Possibilistic weight
[3000, 4000]	0.226	1.0
[2000, 4000]	0.368	0.774
[2000, 5000]	0.211	0.406
[1000, 5000]	0.195	0.195

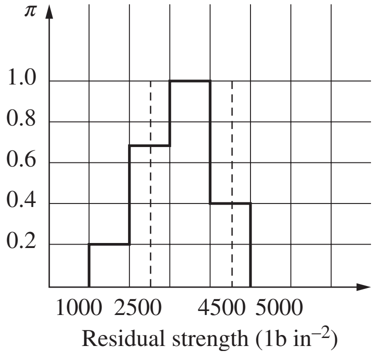

Figure 13.8 Possibility distribution for the residual strength of the bridge.

Similarly, weights for other interval measurements are determined and are shown in Table 13.6.

Finally, a possibility distribution is traced over the consonant intervals from Table 13.6 with the corresponding new weights, as shown in Figure 13.8.

There is significant information in any possibility distribution that is not available when one chooses to represent the same information probabilistically. For example, the possibility distribution in Figure 13.8 has possibilistic information on residual strength of a wooden bridge. Suppose we choose to focus on a specific interval of strengths, say $A = [2500, 4500]$ lb in.$^{-2}$ (as shown by the broken lines in Figure 13.8). The possibility of the actual strength of the bridge being in the interval [2500, 4500] lb in.$^{-2}$ is unity. This is because the actual interval in Figure 13.8 with $\pi([3000, 4000]) = 1.0$ is fully contained within the interval of interest, that is, within [2500, 4500] lb in.$^{-2}$. Said another way, the possibility of the actual strength not being in the interval [2500, 4500] lb in.$^{-2}$ is 0.774, that is, $\pi(\text{not A}) = \max(\text{any possibility value for values outside the interval } A) = \max(0.774, 0.406, 0.195)$. Finally, but most importantly, the necessity (Equation (13.28)) is $\eta = 1 - \pi(\text{not } A) = 0.226$. Hence, we can say that the actual strength of the bridge being in the interval [2500, 4500] lb in.$^{-2}$ is *certainly* 0.226 but could be *possibly* 1.0. The interval of [3000, 4000] lb in.$^{-2}$ would contain the *most possible* value. A probabilistic assessment of this same question would provide only a confidence level about the interval A, which does not truly represent the kind of evidence available; that is, the information is ambiguous and imprecise and not subject to random variability. In a probabilistic assessment of these data, we would get a 95% confidence interval around the *most probable* value, which usually overestimates the interval that would contain the actual strength.

Moreover, to get such a confidence interval, we have to make assumptions about the data (e.g., the underlying probability distribution) for which we might not have information.

As explained, possibility theory is based on two dual functions, necessity measures (η) and possibility measures (π). The two functions, whose range is [0, 1], can be converted to a single function, C, whose range is [−1, 1], as described below (Klir and Yuan, 1995):

$$C(A) = \eta(A) + \pi(A) - 1. \tag{13.50}$$

Positive values of $C(A)$ indicate the *degree of confirmation* of A by the available evidence, and negative values of $C(A)$ express the *degree of disconfirmation* of A by the evidence. Such a metric adds value to the use of possibility theory in its use in characterizing the forms of uncertainty due to ambiguity, nonspecificity, and imprecision. For instance, in Example 13.11, we compute the *degree of confirmation* for the interval A to be 0.226.

Example 13.12

In the previous example, the intervals were consistent, and hence it was relatively straightforward to calculate the most possible interval and subsequent nested intervals. This example illustrates the application of Donald's (2003) method when the intervals are not consistent, that is, when there is no common interval that spans across all the data intervals. Consider a set of data intervals as shown in Table 13.7.

As there is no interval that is included in all the original intervals, experts can choose the maximally possible interval by relying on their experiences, or they can choose the interval that intersects with the most original data intervals. Owing to its objectivity, the latter approach is used in this example. Once the maximally possible interval is identified, the process of choosing the nested intervals is dependent on the degree to which each interval includes a subsequent interval. The following steps illustrate the entire process of generating a possibility distribution. Table 13.8 shows all the intervals produced by taking the intersections of the original data.

From Table 13.8, it can be seen that the interval [11, 13] intersects with the most intervals and thus forms the maximally possible interval. Given this interval, the rest of the consonant intervals are selected according to the nesting structure of the subsequent intervals. Table 13.9 shows one of the series of consonant and nonconsonant intervals that was selected for this example. Once the nesting structure is determined, the weights are then redistributed from the nonconsonant to the consonant intervals as explained in the previous example. Table 13.10 shows the final weights assigned to each interval after this redistribution, and Figure 13.9 shows the possibility distribution plotted according to the weights on the intervals shown in Table 13.10.

Table 13.7 Original data intervals.

Observation	Intervals	Weight
1	[1, 6]	0.25
2	[2, 14]	0.25
3	[7, 13]	0.25
4	[11, 17]	0.25

Table 13.8 Intervals produced by taking the intersections of all the original data intervals.

Interval	Weight	Normalized weight
[1, 6]	0.25	0.125
[2, 6]	0.25	0.125
[2, 14]	0.25	0.125
[7, 13]	0.25	0.125
[11, 17]	0.25	0.125
[11, 13]	0.25	0.125
[11, 14]	0.25	0.125
[1, 17]	0.25	0.125

Table 13.9 Redistribution of weights from the nonconsonant to the consonant intervals.

Nonconsonant interval	Consonant interval	β	K	ρ
[1, 6]	[11, 13]	0	0	0
	[7, 13]	0	0	0
	[2, 14]	0.3333	0.516	0.0645
	[1, 17]	0.3125	0.484	0.0605
[2, 6]	[11, 13]	0	0	0
	[7, 13]	0	0	0
	[2, 14]	0.3333	0.571	0.0714
	[1, 17]	0.2500	0.429	0.0536
[11, 17]	[11, 13]	1	0.512	0.0640
	[7, 13]	0.3333	0.170	0.0213
	[2, 14]	0.2500	0.128	0.0160
	[1, 17]	0.3750	0.190	0.0238
[11, 14]	[11, 13]	1	0.566	0.0708
	[7, 13]	0.3333	0.187	0.0233
	[2, 14]	0.2500	0.141	0.0176
	[1, 17]	0.1875	0.106	0.0133

Table 13.10 Final possibilistic weights after the transfer of weight from the nonconsonant to the consonant intervals.

Interval	Final weight	Possibilistic weight
[11, 13]	0.26	1.00
[7, 13]	0.17	0.74
[2, 14]	0.29	0.57
[1, 17]	0.28	0.28

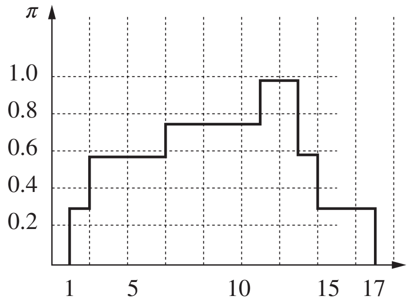

Figure 13.9 Possibility distribution generated from nonconsistent intervals.

Table 13.11 Experimental intervals of concrete cracking strength.

Evidence source	Interval (MPa)	Weight	Reference
Direct tension test	[1.0, 1.6]	1/9	I-1
	[1.4, 2.2]	1/9	I-2
	[2.0, 2.4]	1/9	I-3
Splitting tensile strength test	[2.8, 3.8]	1/9	I-4
	[3.0, 3.6]	1/9	I-5
	[3.2, 4.0]	1/9	I-6
Flexural strength test	[4.4, 4.8]	1/9	I-7
	[5.0, 6.0]	1/9	I-8
	[5.4, 5.8]	1/9	I-9

Example 13.13

We now present an example where we might have a collection of different sets of information which relate to the same problem. For example, suppose we want to determine the most possible interval that would characterize the cracking strength of concrete from a single mix-design. To do this we conduct three different types of cracking strength tests: (1) flexural strength test, (2) splitting tensile strength test, and (c) direct tension test. From each type of experimental test, three intervals of cracking strength were determined from three different batches, as shown in Table 13.11. The table shows the type of test, the cracking strength interval, the evidence weight we assign to each interval (we assume equal weights), and a reference number for the test (tests 1 through 9). We want to generate a possibility distribution of concrete cracking strength and to determine the most possible cracking strength interval.

In our solution process we first want to classify the nine intervals by type of interval; we will call each type of interval an *interval set*. Second, for each interval set, we will develop a specific possibility distribution. Third, for all of the interval sets we will combine the information into a "global" possibility distribution by using a procedure called *anchoring* Kim (2009).

1. Classification: The nine intervals in Table 13.11 can be classified by their connectivity into four interval sets: (a) consonant intervals, (b) independent interval, (c) consistent intervals, and (d) overlapping intervals, these are all shown in Figure 13.10.

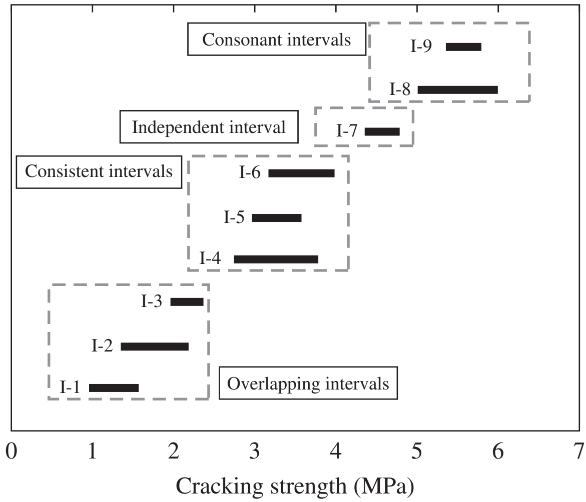

Figure 13.10 Classification of nine intervals into four "interval sets".

Table 13.12 Classified "intervals sets" and normalized weights.

Classification	Reference	Local weight	Global weight
Overlapping intervals	I-1	1/3	1/3
	I-2	1/3	
	I-3	1/3	
Consistent intervals	I-4	1/3	1/3
	I-5	1/3	
	I-6	1/3	
Independent interval	I-7	1	1/9
Consonant intervals	I-8	1/2	2/9
	I-9	1/2	

2. Generate possibility distribution for each interval set: For each interval set, global weights of intervals are normalized as shown in Table 13.12. We now go through the process of developing a possibility distribution for each interval set.

   a. *Consonant intervals*: For intervals, I-8 and I-9 (Figure 13.10), a possibility distribution can be generated by using Equation (13.49), and repeated here,

$$\pi(x) = \sum_{x \in \mathrm{H}_i} m(\mathrm{H}_i). \tag{13.51}$$

   Because there are only two intervals in this set, and each has equal weight, the possibility distribution is easy to generate, and is shown in Table 13.13.

Table 13.13 Possibility distribution for the consonant "interval set".

Reference	Interval	Local possibility (%)
I-8	[5.0, 6.0]	50
I-9	[5.4, 5.8]	100

Table 13.14 Possibility distribution for independent "interval set".

Reference	Interval	Local possibility distribution (%)
I-7	[4.4, 4.8]	100

Table 13.15 All possible intervals from I-4, I-5, and I-6.

Reference	Interval	Local weight	Normalized local weight
I-4	[2.8, 3.8]	1/3	1/9
I-5	[3.0, 3.6]	1/3	1/9
I-6	[3.2, 4.0]	1/3	1/9
New	[2.8, 3.6]	1/3	1/9
New∗	[2.8, 4.0]	1/3	1/9
New∗	[3.0, 3.8]	1/3	1/9
New	[3.0, 4.0]	1/3	1/9
New∗	[3.2, 3.6]	1/3	1/9
New	[3.2, 3.8]	1/3	1/9

b. *Independent interval*: Interval I-7 is independent (has no overlap with any others) from all the other intervals (Figure 13.10); the possibility distribution will be 100% as shown in Table 13.14.

c. *Consistent intervals*: For intervals I-4, I-5, and I-6 (Figure 13.10), all possible intervals are generated and local weights are normalized, as shown in Table 13.15. In this table, the references labeled "New" are those intervals that result from combinations of the lower bounds and upper bounds (intersections) of the three reference intervals given on the first three lines of Table 13.15.

Because there is only one interval, [3.2, 3.6], that is "consistent" (an interval which intersects all original intervals) with all the others in Table 13.15, a consonant body of intervals can be constructed directly from this consistent interval and these are shown with an asterisk in Table 13.15.

In Table 13.15, there are six other intervals that are not consonant with the interval [3.2, 3.6], and these are called *nonconsonant intervals*. As in the previous example (Example 13.12), we must redistribute weights to these intervals. Hence, the weights of the nonconsonant intervals are redistributed to the consonant intervals based on the similarity between two intervals. The similarity of two intervals β is defined in Equation (13.44), and is repeated here,

$$\beta_{ij} = \frac{|H_i \cap I_j|}{|H_i|}. \tag{13.52}$$

A redistribution factor k is then computed from Equation (13.46), which is repeated here, as

$$\kappa_{ij} = \frac{\beta_{ij}}{\sum_{i=1}^{n_H} \beta_{ij}}. \tag{13.53}$$

Finally, weights of consonant intervals are determined by an expression that is similar to Equation (13.48),

$$m(H_i) = w(H_i) + \sum_{j=1}^{n_I} \kappa_{ij} w(I_j). \tag{13.54}$$

The calculated similarity β and redistribution κ factors for the six nonconsonant intervals of Table 13.15 are presented in Table 13.16.

As an illustration of the determination of some of the values in Table 13.16, the similarity and redistribution factors between nonconsonant interval [2.8, 3.6] and consonant intervals are calculated using Equations (13.52) and (13.53) as

Table 13.16 Similarity and redistribution factors for weight redistribution.

Nonconsonant intervals	Consonant intervals	β	k
[2.8, 3.6]	[3.2, 3.6]	1	0.375
	[3.0, 3.8]	1	0.375
	[2.8, 4.0]	0.667	0.25
[2.8, 3.8]	[3.2, 3.6]	1	0.3529
	[3.0, 3.8]	1	0.3529
	[2.8, 4.0]	0.833	0.2941
[3.0, 3.6]	[3.2, 3.6]	1	0.4444
	[3.0, 3.8]	0.75	0.3333
	[2.8, 4.0]	0.5	0.2222
[3.0, 4.0]	[3.2, 3.6]	1	0.3529
	[3.0, 3.8]	1	0.3529
	[2.8, 4.0]	0.833	0.2941
[3.2, 3.8]	[3.2, 3.6]	1	0.4444
	[3.0, 3.8]	0.75	0.3333
	[2.8, 4.0]	0.5	0.2222
[3.2, 4.0]	[3.2, 3.6]	1	0.375
	[3.0, 3.8]	1	0.375
	[2.8, 4.0]	0.667	0.25

$$\beta_{11} = \frac{|[2.8, 3.6] \cap [3.2, 3.6]|}{|[3.2, 3.6]|} = \frac{0.4}{0.4} = 1,$$

$$\beta_{21} = \frac{|[2.8, 3.6] \cap [3.0, 3.8]|}{|[3.0, 3.8]|} = \frac{0.8}{0.8} = 1,$$

$$\beta_{31} = \frac{|[2.8, 3.6] \cap [2.8, 4.0]|}{|[2.8, 4.0]|} = \frac{0.8}{1.2} = 0.667, \text{ and}$$

$$\kappa_{11} = \frac{\beta_{11}}{\beta_{11} + \beta_{21} + \beta_{31}} = \frac{1}{1 + 1 + 0.667} = 0.375, \quad \kappa_{21} = \frac{1}{2.667} = 0.375,$$

$$\kappa_{31} = \frac{0.667}{2.667} = 0.25.$$

Finally, local weights of the consonant intervals are determined using Equation (13.48)

$$m(\mathrm{H}_1) = w(\mathrm{H}_1) + \sum_{j=1}^{6} \kappa_{1j} w(\mathrm{I}_j)$$

$$= 1/9 + (0.375 + 0.3529 + 0.4444 + 0.3529 + 0.4444 + 0.375)/9 = 0.3716.$$

The remaining final local weights are determined in a similar manner, and the possibility distribution for the consistent "interval set" is presented in Table 13.17.

d. *Overlapping intervals:* For the overlapping interval set (Figure 13.10), I-1, I-2, and I-3, we determine all possible intervals (combinations of the upper and lower bounds of the intervals I-1, I-2, and I-3) along with their local and normalized weights, as shown in Table 13.18.

Table 13.18 shows that there are two consistent intervals, that is, [1.4, 1.6] and [2.0, 2.2], so we need to determine the consistent interval that is the "best" in some sense to construct a consonant body of intervals. Therefore, we make use of an expression called the *expected interval* for the original overlapping intervals (I-1, I-2, and I-3), as given by Kim (2009),

$$E(\mathrm{O}) = \left[\sum_{i=1}^{n} m(\mathrm{O}_i)\mathrm{O}_i^l, \ \sum_{i=1}^{n} m(\mathrm{O}_i)\mathrm{O}_i^u \right]. \tag{13.55}$$

The calculated expected interval is then calculated to be,

$$E(\mathrm{O}) = [(1.0 + 1.4 + 2.0)/3, (1.6 + 2.2 + 2.4)/3] = [1.467, 2.067]$$

Table 13.17 Final local weights and possibility distribution for consistent "interval set".

Interval	Normalized local weight	Local possibility distribution (%)
[3.2, 3.6]	0.3716	100
[3.0, 3.8]	0.347	62.84
[2.8, 4.0]	0.2814	28.14

Table 13.18 All possible intervals from I-1, I-2, and I-3.

Reference	Interval	Local weight	Normalized local weight
I-1	[1.0, 1.6]	1/3	1/8
I-2	[1.4, 2.2]	1/3	1/8
I-3	[2.0, 2.4]	1/3	1/8
New	[1.0, 2.2]	1/3	1/8
New	[1.0, 2.4]	1/3	1/8
New	[1.4, 1.6]	1/3	1/8
New	[1.4, 2.4]	1/3	1/8
New	[2.0, 2.2]	1/3	1/8

The expected interval is actually the mean of the lower bounds of the two consistent intervals and the minimum lower bound of all intervals in Table 13.18, and the mean of the upper bounds of the two consistent intervals and the maximum upper bound of all intervals in Table 13.18. Finally, the Euclidean distances between the expected interval and the two consistent intervals are calculated, and the closest consistent interval to the expected interval is determined as the "best" interval; for example,

$$\Delta([1.4, 1.6]-[1.467, 2.067]) = \sqrt{(1.4-1.467)^2+(1.6-2.067)^2} = 0.472,$$

$$\Delta([2.0, 2.2]-[1.467, 2.067]) = \sqrt{(2.0-1.467)^2+(2.2-2.067)^2} = 0.549$$

Therefore, interval [1.4, 1.6] is determined as the "best" (it has the minimum distance of 0.472) interval and a consonant body of intervals is then constructed based on this "best" consistent interval, as shown in Table 13.19.

Following procedures similar to those of the consistent interval set, similarity and redistribution factors are calculated and shown in Table 13.20.

As an example of the calculations in Table 13.20, similarity and redistribution factors between the nonconsonant interval [1.0, 1.6] and the consonant intervals (Table 13.19) are calculated using Equations (13.52) and (13.53) as

$$\beta_{11} = \frac{|[1.0,\ \ 1.6]\cap[1.4,\ \ 1.6]|}{|[1.4,\ \ 1.6]|} = \frac{0.2}{0.2} = 1,$$

$$\beta_{21} = \frac{|[1.0,\ \ 1.6]\cap[1.4,\ \ 2.2]|}{|[1.4,\ \ 2.2]|} = \frac{0.2}{0.8} = 0.25,$$

$$\beta_{31} = \frac{|[1.0,\ \ 1.6]\cap[1.0,\ \ 2.4]|}{|[1.0,\ \ 2.4]|} = \frac{0.6}{1.4} = 0.429, \text{ and}$$

$$\kappa_{11} = \frac{\beta_{11}}{\beta_{11}+\beta_{21}+\beta_{31}} = \frac{1}{1+0.25+0.429} = 0.596, \quad \kappa_{21} = \frac{0.25}{1.679} = 0.149,$$

$$\kappa_{31} = \frac{0.429}{1.679} = 0.255.$$

Table 13.19 Consonant intervals for the overlapping "interval set".

Interval	Normalized local weight
[1.4, 1.6]	1/8
[1.4, 2.2]	1/8
[1.0, 2.4]	1/8

Table 13.20 Similarity and redistribution factors for weight redistribution.

Nonconsonant intervals	Consonant intervals	β	k
[1.0, 1.6]	[1.4, 1.6]	1	0.5957
	[1.4, 2.2]	0.25	0.1489
	[1.0, 2.4]	0.429	0.2553
[1.0, 2.2]	[1.4, 1.6]	1	0.35
	[1.4, 2.2]	1	0.35
	[1.0, 2.4]	0.857	0.3
[1.4, 2.4]	[1.4, 1.6]	1	0.3684
	[1.4, 2.2]	1	0.3684
	[1.0, 2.4]	0.714	0.2632
[2.0, 2.2]	[1.4, 1.6]	0	0
	[1.4, 2.2]	0.25	0.6364
	[1.0, 2.4]	0.143	0.3636
[2.0, 2.4]	[1.4, 1.6]	0	0
	[1.4, 2.2]	0.25	0.4667
	[1.0, 2.4]	0.286	0.5333

Finally, local weights of the consonant intervals are determined using Equation (13.54) as

$$m(\mathrm{H}_1) = w(\mathrm{H}_1) + \sum_{j=1}^{5} \kappa_{1j} w(\mathrm{I}_j)$$

$$= 1/8 + (0.5957 + 0.35 + 0.3684 + 0 + 0)/8 = 0.2893.$$

The calculated final local weights and possibility distributions values for overlapping "interval set" are presented in Table 13.21.

3. Combine disjointed possibility distributions by anchoring method Kim (2009): To combine the four disjointed possibility distributions shown in Tables 13.3, 13.14, 13.17, and 13.21, a weighted anchor is used,

$$A(\mathrm{D}) = \left[\sum_{j=1}^{n} m(\mathrm{D}_j)\mathrm{D}_j^l, \sum_{j=1}^{n} m(\mathrm{D}_j)\mathrm{D}_j^u \right]. \tag{13.56}$$

To use Equation (13.56), we must first calculate the global weights for all the intervals in the four interval sets. The third column in Table 13.22 contains the local weights of the four interval sets (Tables 13.13, 13.14, 13.17, and 13.21). The global weights are determined

Table 13.21 Final local weights and possibility distribution for overlapping "interval set".

Interval	Normalized local weight	Local possibility distribution (%)
[1.4, 1.6]	0.2893	100
[1.4, 2.2]	0.3713	71.07
[1.0, 2.4]	0.3394	33.94

Table 13.22 Global weights for all interval sets.

Classification	Consonant intervals	Local weight	Global weight
Overlapping set	[1.4, 1.6]	0.2893	0.096
	[1.4, 2.2]	0.3713	0.1238
	[1.0, 2.4]	0.3394	0.1131
Consistent set	[3.2, 3.6]	0.3716	0.1239
	[3.0, 3.8]	0.347	0.1157
	[2.8, 4.0]	0.2814	0.0938
Independent set	[4.4, 4.8]	1	0.1111
Consonant set	[5.0, 6.0]	0.5	0.1111
	[5.4, 5.8]	0.5	0.1111

from the local weights in the following manner. The local weights for the overlapping set, for example, had an original weight of 1/9 and three of these intervals are in the overlapping interval set. So, for example, the global weight for consonant interval [1.4, 1.6] is determined as, $0.2893 * (1/9) * 3 = 0.096$. Continuing, the global weight for interval [5.0, 6.0] of the consonant interval set is obtained by $0.5 * (1/9) * 2 = 0.1111$ (since there are two intervals in this interval set). The other global weights are shown in Table 13.22.

Now, using the consonant intervals from the four interval sets and their corresponding global weights from Table 13.22, we determine the weighted anchor as,

$$A(\mathrm{D}) = \begin{bmatrix} 1.4(0.096) + 1.4(0.1238) + 1.0(0.1131) + 3.2(0.1239) + 3.0(0.1157) \\ +2.8(0.0938) + 4.4(0.1111) + 5.0(0.1111) + 5.4(0.1111), \\ 1.6(0.096) + 2.2(0.1238) + 2.4(0.1131) + 3.6(0.1239) + 3.8(0.1157) \\ +4.0(0.0938) + 4.8(0.1111) + 6.0(0.1111) + 5.8(0.1111) \end{bmatrix}$$

$$= [3.08, 3.80]$$

Then, the distances between the weighted anchor and the consonant intervals are calculated and reordered (from lowest distance to highest), as shown in Table 13.23. For example,

$$\Delta([1.4, 1.6] - [3.08, 3.80]) = \sqrt{(1.4-3.08)^2 + (1.6-3.80)^2} = 2.77.$$

Therefore, the most "possible" interval (interval with a possibility value = 1.0) is determined as [3.0, 3.8]. On the basis of the ordered consonant intervals of Table 13.23, global consonant

Table 13.23 Reordered intervals from consonant "interval sets" and distance from anchor.

Reordered intervals	Distance	Global weight
[3.0, 3.8]	0.08	0.1157
[3.2, 3.6]	0.24	0.1239
[2.8, 4.0]	0.34	0.0938
[4.4, 4.8]	1.66	0.1111
[1.4, 2.2]	2.32	0.1238
[1.0, 2.4]	2.50	0.1131
[1.4, 1.6]	2.77	0.096
[5.0, 6.0]	2.92	0.1111
[5.4, 5.8]	3.07	0.1111

Table 13.24 Global consonant intervals and weight.

Consonant intervals	Global weight	Global possibility distribution (%)
[3.0, 3.8]	0.2395	100
[2.8, 4.0]	0.0938	76.04
[2.8, 4.8]	0.1111	66.66
[1.4, 4.8]	0.1238	55.55
[1.0, 4.8]	0.2095	43.17
[1.0, 6.0]	0.2222	22.22

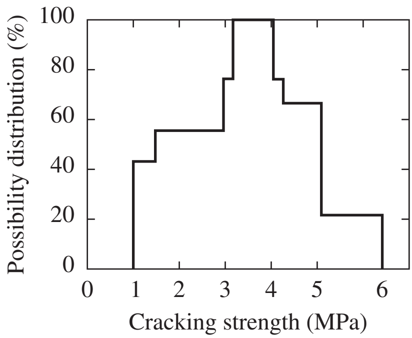

Figure 13.11 Final possibility distribution for Example 13.13.

intervals and their corresponding weights are determined using the following expressions (Kim, 2009) and shown in Table 13.24 and the possibility distribution in Figure 13.11:

$$\mathrm{H}_i = \left[\min\left(\mathrm{D}^l | \mathrm{D}^l \in \mathrm{D}_{\mathrm{New}}\right), \max\left(\mathrm{D}^u | \mathrm{D}^u \in \mathrm{D}_{\mathrm{New}}\right)\right], \tag{13.57}$$

$$m(\mathrm{H}_i) = \sum_{\mathrm{D}_i \in \mathrm{H}_i} m(\mathrm{D}_i) - \sum_{\mathrm{D}_i \in \mathrm{H}_{i+1}} m(\mathrm{D}_i). \tag{13.58}$$

Using Equation (13.57), the most possible interval $H_1 = [3.0, 3.8]$ is compared to the second closest interval [3.2, 3.6] as

$$H_2 = [\min(3.0, 3.2), \max(3.8, 3.6)] = [3.0, 3.8].$$

For the next interval, and so forth, we get,

$$H_3 = [\min(3.0, 3.2, 2.8), \max(3.8, 3.6, 4.0)] = [2.8, 4.0]$$
$$H_4 = [\min(3.0, 3.2, 2.8, 4.4), \max(3.8, 3.6, 4.0, 4.8)] = [2.8, 4.8]$$
$$H_5 = [\min(3.0, 3.2, 2.8, 4.4, 1.4), \max(3.8, 3.6, 4.0, 4.8, 2.2)] = [1.4, 4.8]$$
$$H_6 = \begin{bmatrix} \min(3.0, 3.2, 2.8, 4.4, 1.4, 1.0), \\ \max(3.8, 3.6, 4.0, 4.8, 2.2, 2.4) \end{bmatrix} = [1.0, 4.8]$$
$$H_7 = \begin{bmatrix} \min(3.0, 3.2, 2.8, 4.4, 1.4, 1.0, 1.4), \\ \max(3.8, 3.6, 4.0, 4.8, 2.2, 2.4, 1.6) \end{bmatrix} = [1.0, 4.8]$$
$$H_8 = \begin{bmatrix} \min(3.0, 3.2, 2.8, 4.4, 1.4, 1.0, 1.4, 5.0), \\ \max(3.8, 3.6, 4.0, 4.8, 2.2, 2.4, 1.6, 6.0) \end{bmatrix} = [1.0, 6.0]$$
$$H_9 = \begin{bmatrix} \min(3.0, 3.2, 2.8, 4.4, 1.4, 1.0, 1.4, 5.0, 5.4), \\ \max(3.8, 3.6, 4.0, 4.8, 2.2, 2.4, 1.6, 6.0, 5.8) \end{bmatrix} = [1.0, 6.0]$$

Then, using Equation (13.58), we can determine the final global weights for the combined consonant intervals,

$$m([3.0, 3.8]) = m([3.0, 3.8]) + m([3.2, 3.6]) = 0.1157 + 0.1239 = 0.2396$$
$$\begin{aligned} m([2.8, 4.0]) &= \{m([3.0, 3.8]) + m([3.2, 3.6]) + m([2.8, 4.0])\} \\ &\quad - \{m([3.0, 3.8]) + m([3.2, 3.6])\} \\ &= \{0.1157 + 0.1239 + 0.0938\} - \{0.1157 + 0.1239\} = 0.0938, \end{aligned}$$
$$m([2.8, 4.8]) = 0.4445 - 0.3334 = 0.1111$$
$$m([1.4, 4.8]) = 0.5683 - 0.4445 = 0.1238$$
$$m([1.0, 4.8]) = 0.7778 - 0.5683 = 0.2095$$
$$m([1.0, 6.0]) = 1 - 0.7778 = 0.2222$$

Comparison of Possibility Theory and Probability Theory

Both possibility theory and probability theory are special branches of evidence theory. This chapter has shown how the two theories relate. Both share similar axiomatic foundations, but there are a few, quite distinct, differences. Their normalizations are different: probabilities must sum to unity, and possibilities have one or more maximal elements equal to unity. And the way in which each represents ignorance is different. For probability measures, total ignorance is expressed by the uniform probability distribution function

$$p(x) = \frac{1}{|\mathrm{X}|} = m(\{x\}), \tag{13.59}$$

for all $x \in \mathrm{X}$. This follows from the fact that *bea's* are required to focus only on singletons. This choice is justified on several grounds with probability theory, where it is required that every uncertainty situation be characterized by a single probability distribution. But this reasoning represents a paradox: on purely intuitive grounds if no information is available about a situation, then no distribution is supported by any evidence and, hence, a choice of one over the other is arbitrary. Total ignorance should be represented by all possible distribution functions, but this is not a formulation of the theory.

In possibility theory, ignorance is represented naturally (as discussed in the section on evidence theory), and is expressed as

$$m(\mathrm{X}) = 1. \tag{13.60}$$

Here, we have that all the evidence is allocated only to the full universe, X. Hence, the information is completely nonspecific; there is no evidence supporting any singleton or subset of the universe, that is, $m(A) = 0$ for all $A \neq \mathrm{X}$.

Although interpretations of possibility theory are less developed than their probabilistic counterparts, it is well established that possibility theory provides a link between fuzzy set theory and probability theory. When information regarding some situation is given in both probabilistic and possibilistic terms, the two interpretations should, in some sense, be consistent. That is, the two measures must satisfy some *consistency condition* (this form of consistency should not be confused with the same term used in the previous section on consistent intervals). Although several such conditions have been reported (Klir and Yuan, 1995) the weakest one acceptable on an intuitive basis is stated as "an event that is probable to some degree must be possible to at least that same degree," or, the *weak consistency condition* can be expressed formally as

$$p(A) \leq \pi(A), \tag{13.61}$$

for all $A \in \mathrm{P(X)}$. The *strongest consistency condition* would require, alternatively, that any event with nonzero probability must be fully possible; formally,

$$p(A) > 0 \Rightarrow \pi(A) = 1, \tag{13.62}$$

All other consistency conditions fall between the extremes specified by Equations (13.61) and (13.62).

Summary

This chapter has summarized briefly a few of the various elements of monotone measures: beliefs, plausibilities, possibilities, necessities (certainties), and probabilities. The axiomatic expansion of these measures, along with fuzzy set theory as a special case, is known collectively in the literature as *generalized information theory* (GIT) (Klir and Wierman, 1999). GIT contains fuzzy set theory, but this development is not explored in this text. Suffice it to say that the current research in GIT is expanding and showing that, collectively, these various theories are powerful in representing a large suite of uncertainties: fuzziness, vagueness, unknownness, nonspecificity, strife, discord, conflict, randomness, and ignorance.

The engineering community has mostly relied on probabilistic methods to analyze empirical data. These methods are widely used in experimental analysis for studying the variation in model parameters, for determining structural properties of various materials, and for image processing. The variables under consideration are assumed to be randomly distributed, and hence the analysis depends on methods that satisfy the axioms of probability. However, on many occasions, the data available do not necessarily represent complete knowledge and thus do not support probabilistic analysis. When faced with limited data it might be more appropriate to use possibility distributions such that complete knowledge is not assumed. Possibility distributions can capture the imprecision in data and are thus useful in quantifying uncertainty resulting from incomplete knowledge. This chapter presented some of the recent advances in the area of generation of possibility distributions from empirical data that are imprecise, that is, when data are available as a set of overlapping intervals. In addition to the methods presented in this chapter, other methods for deriving possibility distributions exist; the reader is referred to Donald (2003) and Joslyn (1997) for alternative methods, especially for deriving possibility distributions from disjoint data sets. The method presented in this chapter on the generation of possibility distributions has been successfully used in the area of damage detection in a multistory structure (see Altunok, Reda Taha, and Ross, 2007). In this work, possibility distributions of the damage feature were produced for "healthy" and "damaged" structures. It was shown that damage can be quantified by comparing the two possibility distributions of "healthy" and "damaged" structures.

There exists a formal relationship between probability and fuzzy logics; this relationship, discussed in Chapter 1, illustrates axiomatically that their common features are more substantial than their differences. It should be noted in Gaines (1978) that, whereas the additivity axiom is common to both a probability and a fuzzy logic, it is rejected in the Dempster–Shafer theory of evidence, and it often presents difficulties to humans in their reasoning. For a probability logic, the axiom of the excluded middle (or its dual, the axiom of contradiction, that is, $p(x \wedge \bar{x}) = 0$) *must* apply; for a fuzzy logic it *may* or *may not* apply.

There are at least two reasons why the axiom of the excluded middle might be inappropriate for some problems. First, people may have a high degree of belief about a number of possibilities in a problem. A proposition should not be "crowded out" just because it has a large number of competing possibilities. The difficulties people have in expressing beliefs consistent with the axioms of a probability logic are sometimes manifested in the rigidity of the axiom of the excluded middle (Wallsten and Budescu, 1983). Second, the axiom of the excluded middle results in an inverse relationship between the information content of a proposition and its probability. For example, in a universe of n singletons, as more and more evidence becomes available on each of the singletons, the relative amount of evidence on any one diminishes

(Blockley, 1983). This characteristic makes one of these axioms inappropriate as a measure for modeling uncertainty in many situations.

Finally, rather than debate what is the correct set of axioms to use (i.e., which logic structure) for a given problem, one should look closely at the problem, determine which propositions are vague or imprecise and which ones are statistically independent or mutually exclusive, and use these considerations to apply a proper uncertainty logic, with or without the axiom of the excluded middle. By examining a problem so closely as to determine these relationships, one finds out more about the structure of the problem in the first place. Then, the assumption of a strong truth-functionality (for a fuzzy logic) could be viewed as a computational device that simplifies calculations, and the resulting solutions would be presented as ranges of values that most certainly form bounds around the true answer if the assumption is not reasonable. A choice of whether a fuzzy logic is appropriate is, after all, a question of balancing the model with the nature of the uncertainty contained within it. Problems without an underlying physical model; problems involving a complicated weave of technical, social, political, and economic factors; and problems with incomplete, ill-defined, and inconsistent information where conditional probabilities cannot be supplied or rationally formulated perhaps are candidates for fuzzy logic applications. Perhaps, then, with additional algorithms like fuzzy logic, those in the technical and engineering professions will realize that such difficult issues can now be modeled in their designs and analyses.

References

Altunok, E., Reda Taha, M. M., and Ross, T. J. (2007). A possibilistic approach for damage detection in structural health monitoring. *ASCE J. Struct. Eng.*, **133** (9), 1247–1256.

Blockley, D. (1983) Comments on "Model uncertainty in structural reliability," by Ove Ditlevsen. *J. Struct. Safety*, **1**, 233–235.

Carrol, L. (1893/1958). *The Mathematical Recreations of Lewis Carroll: Pillow Problems and a Tangled Tale* (Mineola, NY: Dover Publications).

Dempster, A. (1967). Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.*, **38**, 325–339.

Donald, S. (2003). Development of empirical possibility distributions in risk analysis. PhD dissertation. University of New Mexico, Department of Civil Engineering, Albuquerque, NM.

Dubois, D., and Prade, H. (1988). *Possibility Theory* (New York: Plenum Press).

Gaines, B. (1978). Fuzzy and probability uncertainty logics. *Inf. Control*, **38**, 154–169.

Joslyn, C. (1997). Measurement of possibilistic histograms from interval data. *Int. J. Gen. Syst.*, **26**, 9–33.

Kim, J. J. (2009). Uncertainty quantification in serviceability of reinforced concrete structures. PhD dissertation. Department of Civil Engineering, University of New Mexico, Albuquerque, NM.

Klir, G., and Folger, T. (1988). *Fuzzy Sets, Uncertainty, and Information* (Englewood Cliffs, NJ: Prentice Hall).

Klir, G., and Smith, R. (2001). On measuring uncertainty and uncertainty-based information: recent developments. *Ann. Math. Artif. Intellig.*, **32**, 5–33.

Klir, G., and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic* (Upper Saddle River, NJ: Prentice Hall).

Klir, G., and Wierman, M. (1999) *Uncertainty Based Information: Elements of Generalized Information Theory* (Heidelberg, Germany: Physica-Verlag/Springer).

Shafer, G. (1976). *A Mathematical Theory of Evidence* (Princeton, NJ: Princeton University Press).

Wallsten, T., and Budescu, D. (1983). *Manage. Sci.*, **29** (2), 167.

Yager, R. (1993). Aggregating fuzzy sets represented by belief structures. *J. Intellig. Fuzzy Syst.*, **1** (3), 215–224.

Yager, R., and Filev, D. (1994). Template-based fuzzy systems modeling. *J. Intellig. Fuzzy Syst.*, **2** (1), 39–54.

Zadeh, L. (1978) .Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.*, **1**, 3–28.

Zadeh, L. (1984). Review of the book *A mathematical theory of evidence*, by Glenn Shafer. *AI Mag.*, 81–83.

Zadeh, L. (1986). Simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination. *AI Mag.*, 85–90.

Problems

13.1 Suppose you have found an old radio (vacuum tube type) in your grandparents' attic and you are interested in determining its age. The make and model of the radio are unknown to you; without this information you cannot find in a collector's guide the year in which the radio was produced. Here, the year of manufacture is assumed to be within a particular decade. You have asked two antique radio collectors for their opinion on the age. The evidence provided by the collectors is fuzzy. Assume the following questions:

1. Was the radio produced in the 1920s?
2. Was the radio produced in the 1930s?
3. Was the radio produced in the 1940s?

Let R, D, and W denote subsets of our universe set P—the set of radio-producing years called the *1920s* (Roaring 20s), the set of radio-producing years called the *1930s* (Depression years), and the set of radio-producing years called the *1940s* (War years), respectively. The radio collectors provide *bea*'s as given in the accompanying table.

	Collector 1			Collector 2			Combined evidence		
Focal elements	m_1	bel_1	pl_1	m_2	bel_2	pl_2	m_{12}	bel_{12}	pl_{12}
R	0.05	0.05	0.8	0.15	0.15	0.85	0.1969		
D	0.1	0.1		0.1	0.1				
W	0	0		0	0				
$R \cup D$	0.2	0.35	1	0.25	0.5	1	0.2677		
$R \cup W$	0.05			0.05					
$D \cup W$	0.1			0.05					
$R \cup D \cup W$	0.5			0.4					

a. Calculate the missing belief values for the two collectors.
b. Calculate the missing plausibility values for the two collectors.
c. Calculate the missing combined evidence values.
d. Calculate the missing combined belief and plausibility values.

13.2 The quality control for welded seams in the hulls of ships is a major problem. Ultrasonic defectoscopy is frequently used to monitor welds, as is X-ray photography. Ultrasonic defectoscopy is faster but less reliable than X-ray photography. Perfect identification of flaws in welds is dependent on the experience of the person reading the signals. An abnormal signal occurs for three possible types of situations; two of these are flaws in welds: a cavity (C) and a cinder inclusion (I); the former is the more dangerous. Another situation is due to a loose contact of the sensor probe (L), which is not a defect in the welding seams but an error in measuring. Suppose we have two experts, each using a different weld monitoring method, who are asked to identify the defects in an important welded seam. Their responses in terms of bea's are given in the table. Calculate the missing portions of the table.

	Expert 1			Expert 2			Combined evidence		
Focal elements	m_1	bel_1	pl_1	m_2	bel_2	pl_2	m_{12}	bel_{12}	pl_{12}
C	0.3	0.3	0.85	0.2	0.2		0.4	0.4	
I	0.05	0.05		0.1	0.1		0.15	0.15	
L	0.05	0.05		0.05	0.05				
C ∪ I	0.2	0.55		0.15	0.45		0.16	0.71	
C ∪ L	0.05	0.4	0.95	0.05	0.3				
I ∪ L	0.05	0.5		0.15	0.3				
C ∪ I ∪ L	0.3	1		0.3	1				

13.3 You are an aerospace engineer who wishes to design a bang–bang control system for a particular spacecraft using thruster jets. You know that it is difficult to get a good feel for the amount of thrust that these jets will yield in space. Gains of the control system depend on the amount of the force the thrusters yield. Thus, you pose a region of three crisp sets that are defined with respect to specific gains. Each set will correspond to a different gain of the control system.

You can use an initial estimate of the force you get from the thrusters, but you can refine it in real time using different gains for the control system. You can get a force estimate and a belief measure for that estimate for a specific set. Suppose you define the following regions for the thrust values, where thrust is in pounds:

A_1 applies to a region $0.8 \le$ thrust value ≤ 0.9.
A_2 applies to a region $0.9 \le$ thrust value ≤ 1.0.
A_3 applies to a region $1.0 \le$ thrust value ≤ 1.1.

Two expert aerospace engineers have been asked to provide evidence measures reflecting their degree of belief for the various force estimates. These *bea*'s along with calculated belief measures are given here. Calculate the combined belief measure for each focal element in the table.

	Expert 1		Expert 2	
Focal elements	m_1	bel_1	m_2	bel_2
A_1	0.1	0.1	0	0
A_2	0.05	0.05	0.05	0.05
A_3	0.05	0.05	0.1	0.1
$A_1 \cup A_2$	0.05	0.2	0.05	0.1
$A_1 \cup A_3$	0.05	0.2	0.15	0.25
$A_2 \cup A_3$	0.1	0.2	0.05	0.2
$A_1 \cup A_2 \cup A_3$	0.6	1	0.6	1

13.4 A test and diagnostics capability is being developed for a motion control subsystem that consists of the following hardware: a motion control IC (integrated circuit), an interconnect between motion control IC, an H-switch current driver, an interconnect between

H-switch current driver, a motor, and an optical encoder. The elements of the motion control subsystem are as follows:

x_1 = motion control IC
x_2 = interconnect 1
x_3 = H – switch current driver
x_4 = interconnect 2
x_5 = motor
x_6 = optical encoder

If a motion control subsystem failure exists, a self-test could describe the failure in the following *bea*: $m = (0.2, 0, 0.3, 0, 0, 0.5)$. This nested structure is based on the level of hardware isolation of the diagnostic software. This isolation is hierarchical in nature. You first identify a motion control subsystem failure m (A_6) that includes a possibility of any component failure (x_1, x_2, x_3, x_4, x_5, x_6). The test then continues and, due to isolation limitations, a determination can be made of the failure possibility consisting of m (A_3), subset (x_1, x_2, x_3), followed by the ability to isolate to an x_1 failure if x_1 is at fault. *Bea*'s are constructed from empirical data and experience.

Find the associated possibility distribution and draw the nesting diagram.

13.5 Design of a geometric traffic route can be described by four roadway features: a corner, a curve, a U-turn, and a circle. The traffic engineer can use four different evaluation criteria (expert guidance) to use in the design process:

m_1 = criteria: fairly fast, short distance, arterial road, low slope points.
m_2 = criteria: slow, short distance, local road, low slope points.
m_3 = criteria: fast, long distance, ramp—type road, medium slope points.
m_4 = criteria: very fast, medium distance, highway, medium slope points.

Corner	Curve	U-turn	Circle	m_1	m_2	m_3	m_4
0	0	0	1	0	0	0.2	0.1
0	0	1	0	0.1	0.1	0	0
0	0	1	1	0	0	0	0
0	1	0	0	0.3	0.2	0.2	0.4
0	1	0	1	0	0	0.5	0.3
0	1	1	0	0.1	0.1	0	0
0	1	1	1	0	0	0	0
1	0	0	0	0.2	0.3	0.1	0
1	0	0	1	0	0	0	0
1	0	1	0	0.1	0.1	0	0.2
1	0	1	1	0	0	0	0
1	1	0	0	0.1	0.1	0	0
1	1	0	1	0	0	0	0
1	1	1	0	0.1	0.1	0	0
1	1	1	1	0	0	0	0

Using the 15 $(2^4 - 1)$ focal elements shown in the accompanying table, determine which, if any, of the four evidence measures ($m_1 - m_4$) results in an ordered possibility distribution.

13.6 There are a number of hazardous waste sites across the country that pose significant health risk to humans. However, because of high costs involved in exposure analysis only a limited amount of information can be collected from each site to determine the extent of contamination. Suppose it is determined that one of the sites is contaminated by a new carcinogenic chemical identified as *Tox*. The table shows the results from the chemical analysis of the groundwater samples collected from one of the sites. Given these sparse data, determine the possibility distribution of exposure concentrations for the chemical *Tox*. Assume the observation weights are identical.

Observation	Concentration ($mg\ l^{-1}$)
1	[0.01, 0.12]
2	[0.03, 0.24]
3	[0.03, 0.15]
4	[0.008, 0.06]

13.7 Because of their excellent self-healing properties, rock salt caverns are used to store nuclear waste from various nuclear plants. One of the properties useful in determining the suitability of a cavern for nuclear waste storage is the creep rate of salt; salt creeps very slowly with time. This creep rate determines the strength of the cavern and the duration that the cavern can be accessible to human operations. The table shows the strain rate results from creep tests conducted on rock salt cores from four locations of the waste repository. Given these data, determine the strain rate interval that is 80% possible (possibility weight = 0.8). Assume the observation weights are identical. Also, find the *degree of confirmation*.

Observation	Strain rate (s^{-1})
1	[6.0E-10, 9.0E-10]
2	[8.0E-10, 1.2E-9]
3	[9.0E-10, 3.0E-9]
4	[5.0E-10, 10.0E-10]

13.8 Predicting interest rates is critical for financial portfolio management and other investment decisions. Based on historical variations and other factors, the following interest rates are predicted for the next two months. Assume the observation weights are identical.

Observation	Interest rate
1	[0.75, 1.5]
2	[1.0, 1.25]
3	[0.75, 1.25]
4	[1.5, 2.0]
5	[1.75, 2.25]

a. What is the possibility that the interest rates will be higher than 2%?
b. Give the reason for your choice of consonant intervals.
c. Find the *degree of confirmation*.

Index