Cyclic Probabilistic Reasoning Networks

— Some Exactly Solvable Iterative Error-control Structures

Wai-shing Lee

A Thesis Submitted in Partial Fulfillment
of the Requirements for the degree of
Master of Philosophy
in
Information Engineering

March 2001

Contents

List of Figures

List of Tables

Abstract

The primary objective of this thesis is to raise the possibility to a new class of probabilistic reasoning networks, namely *cyclic* probabilistic reasoning networks. Traditionally most probabilistic reasoning networks are constructed on an acyclic setting, whether it is the directed acyclic graphs for Bayesian networks or the acyclic joint dependency condition for semi-lattices or cluster graphs. The distinguishing endeavour underlying our proposal is to lift this acyclic constraint. However, by doing so a lot of the nice properties usually associated with the classical models are lost; for instance, our cyclic probabilistic reasoning networks do not permit the existence of a global probability. Thus it becomes a matter of gain vs loss to determines whether cyclic reasoning networks are going to conduce anything useful. By resorting to the emerging theme of complexity in modern science, we propose a possible communication scheme on cyclic reasoning networks. Further, to take our proposal into concrete actions, we devise a series of cyclic reasoning networks directed towards error-control applications. Other than exhibiting certain error correcting capabilities, our models carry very nice algebraic properties which permits effective analytic treatments which in turn enhances understanding of the intrinsic properties of the decoding actions.

摘要

概率推論結構（probabilistic reasoning networks）乃一類運用於不確定情況下之推論模式。典型概率推論結構有其構造上的條件，最基本的莫過如一類整體無環（global acyclicity）限制；然而，這個限制本身的考慮是純數學性的，當中並不涉及推論問題之根本。故此，本論文的目的乃希望撇除這個不自然的限制，嘗試另辦途徑並建立一個更基本的概率推論模式。特別地，我們將集中注意力於一類帶環概率推論結構（cyclic probabilistic reasoning networks）上。誠然，撇除整體無環的限制將使我們喪失部份便於運算處理的數學性質；但是，我們將借助於一個現代科學的前沿命題 —複雜性（Complexity）— 的方法論，提出另一套可供參考的計算模式。我們的討論將起始於一般理論結構，再進一步實踐於信道傳輸中糾錯碼的解碼研究。我們將發現建基於帶環概率推論結構上的糾錯碼解碼法，能比建基於典型概率推論結構上的擁有更良好的代數性質，從而它們不單只能達到有效解碼的目的，更能讓我們瞭解當中的解碼性質和仔細操作。

Acknowledgements

I am indebted to Prof. Tony Lee for the completion of this project in a number of ways. First, he introduced me these topics while I was fretting over for my research topic in the first few months of this program. Second, his intense passion for semi-lattices has encouraged me to hang on with this topic, even after nearly a whole year of initial fruitless attempt. Most importantly, his deep insight and feeling to the problems have brought me important stimuli to the initial conception and the final crystallization leading to this thesis. Besides, I am grateful to the friendly environment of the Broadband Communication Laboratory and the department in general which has made me feel comfortable as a new comer to the engineer's circle. Lastly, to my family and friends who have offered me concern and encouragement to persevere amidst all the vicissitudes in the past two and a half years, are due my deepest gratitude.

Chapter 1
Layout of the thesis

As implied by the title of this thesis, its core studies involve a hierarchical order of attack to problems rather than a single strike. Using brackets to dissect the successive levels of investigation: by "[Cyclic [Probabilistic [[Reasoning] Networks]]]", the hierarchy realized is:

1. Reasoning;
2. Reasoning Networks;
3. Probabilistic Reasoning Networks;
4. Cyclic Probabilistic Reasoning Networks.

While each item above serves an individual problem in its own right, we need to be cautious when walking down the hierarchy. Roughly speaking, the hierarchy starts from something completely philosophical, and then descends into the real world through the introduction of some intelligible operations - mathematical operations in particular. However, it needs to be emphasized that the descent down the hierarchy is something non-trivial, especially with regard to this thesis which will touch upon certain unconventional points of view. Therefore, in order to elicit the landscape of this project, efforts will be spent on outlining the developments leading to our ultimate crystallization of "Cyclic Probabilistic Reasoning Networks". So the layout of this thesis will, in parallel with the above hierarchy, go as follows:

1. Chapter 2 outlines the philosophy and history of the reasoning problem. This culminates in a discussion of our fundamental rationale for reasoning – the *bottom-up approach* – to be employed throughout this thesis; it will be seen important in being the *operationally fundamental picture.*

2. Chapter 3 is the first step to bringing the philosophy of chapter 2 into actualization: it casts "reasoning" into concrete mathematical operations based

on graphical models and probability - collectively they are called *probabilistic reasoning networks*. In particular, chapter 3 aims at a historical sketch of the development of probabilistic reasoning networks: from the *conceptually fundamental* Bayesian networks to its offspring, namely acyclic cluster graphs / junction trees. On the way of development, we shall point out their underlying inadequacies; in particular, it will be pointed out that such developments are not operationally fundamental, contrasted with our intention to operationally fundamental reasoning.

3. Chapter 4 is the second step to bringing the philosophy of chapter 2 into actualization. In particular, we shall discuss the possibility of alternation to the classical models introduced in chapter 3 in order to recover the operationally fundamental picture. By so doing, the resulting computational models will be generally *cyclic* in character, which is precisely the origin of *"cyclic probabilistic reasoning networks"*. The functioning of cyclic networks differs vastly from their classical (acyclic) counterparts; so in order to apprehend these new models, the emerging theme of complexities in physics will be invoked.

4. Chapter 5 and chapter 6 present the ultimate crystallization of our studies of cyclic probabilistic reasoning networks in which we actualize our ideas to error-control applications. Through the successful applications, non-trivial evidences are obtained that testify to our proposal.

While the main engineering concerns and constructions are developed in chapters 3, 5 and 6, the remaining chapters are important in integrating the whole work to deeper concerns in philosophy and the natural science at which the whole foundation of the present work is grounded.

Chapter 2
Introduction

2.1 What is the reasoning problem?

What is reasoning – a human instinct bestowed by God to conform to His rules, or is it a human faculty capable of self-learning and adapting? How is reasoning achieved – can it be transformed into a deterministic mechanism as Newton's laws of motion predicting comet Shoemaker's collision on Jupiter, or must it be probabilistic in character as in coin tossing?

Great question it is! The reasoning problem has proved itself one of the most perplexing issues throughout human history: not only have we not a definite procedure to deal with it, we actually have no fixed idea of what it really is. There have been numerous attempts in establishing valid reasoning frameworks; nevertheless, it is fair to say that despite strenuous efforts of generations of great thinkers, the state of the art is far from mature and we have no unanimous agreement on what the solution should be; instead, we have many solutions, each addressing the problem in a different perspective. The intricacy underlying the affair is that there is always a certain fluidity on what a *proper reasoning* really means, depending on what the problems are related to and by what *rationales* and *contexts* is that *properness* justified. So while some general schemes can always be conjured, they are almost invariably doomed not exhaustive but limited in scope. Therefore, when proposing a scheme for reasoning, we need to pay attention to the different aspects constituting the problem, in particular we shall differentiate a reasoning problem into the following aspects:

1. the problems,

2. the reasoning structures, and

3. the *correspondence* between the two.

Here let us explain what is meant by our differentiation. The first two items are the primary objects of discussion: by *problems* we mean the distinct classes of problems – may it be a betting problem in tossing a coin, or an "intelligence" problem attacked in modeling human behaviour; by *reasoning structures* we refer to the logical forms underlying the reasoning schemes – may it be a deterministic differential equation or some combinatorics. To determine what *reasoning structure* to go with which *class of problems*, we need to check carefully the *correspondence* between the two. Roughly speaking, a *correspondence* corresponds to a *perspective* under which a reasoning structure is justified to be a *valid* answer to the problems concerned. In general, such *correspondence* should **never** be taken for granted and assumed universal, as it is quite clear a lesson from the history of the "science" of reasoning, or at an even deeper root of philosophy proper, that we have hitherto not a single *perspective* which is testified to be so (although there are many strong claims). And we end up with a rich heritage of approaches distinguished by adjectives like ontological, epistemological, hermeneutical, metaphysical, empirical, pragmatic and so on; each addressing a different perspective, a different way of concern and understanding from which a problem is attacked. Our work is no exception in being limited in scope, and to ground our work on a solid footing we shall take pains in elucidating our line of thought, and aspire at erecting a consistent reasoning scheme in the most natural manner.

2.2 Fundamental nature of Knowledge

In the history of western philosophy, there are two major traditions in positioning man relative to his *outer reality*: the natural world, moral values, etc. Platonism is a tradition that emphasizes *transcendental essence*, an unchanging world of *Being* which belongs to a higher state – the *Absolute Goodness*, and is imposed upon our secular world. From Platonism come naturally the physical laws which maintain order in the physical world: sunrise, sunset and stars tracing across the celestial sphere, and moral values which refrain human societies from falling into chaos; however, such external imposition of order entails an expense of the freedom to varieties, and suffocates the possibility to other possible *"Being"*s. In contrast, Aristotelianism is the other tradition purporting to on-earth experiences – a self-regulating world of *Becoming*; so rather than transcendental essence which is external to oneself there is *immanent essence* within everybody that drives towards the ultimate

Being. Under the spirit of Aristotelianism, there emerges a blossoming of ideas to account for our daily observances on one hand, but conflicts and contradictions on the other hand. Although presented in a way apparently opposing each other, we shall take a more humble and objective point of view, and refrain ourselves from a prejudice against either side[1]. Historically speaking at least, what matters is really a propensity to either emphasis arose out of different natures of problems and/or attitudes adopted towards certain interpretations. Perhaps a better way to depict of the situation is to take the following point of view: *Becoming* is to be viewed as certain *operational procedures* (to phase in modern terms) through which to arrive at the goal – the ultimate and absolute *Being*. In other words, *Being* refers to the ultimate objective, while *Becoming* is the crystallization, the sublimation of the earthly to that ultimate. That means: we may treat *Being* as a disposition to the *objective* while *Becoming* as one towards the *operation*[2].

A natural question which follows after the identification of *Being* and *Becoming* is their relative primacy. While certain external orders look as though they are always true and fundamental: for instance physical laws, moral values such as love, freedom, etc., hence purporting to the primacy of an unchanging *Being*; however, as long as we admit that we need some language to express that *Being* and certain operational means to drive to that, it seems inevitable that we have to admit a fundamental role of *Becoming*. So the question is undoubtedly one of the historic ones in human history which is no less controversial than asking whether the chicken comes first or the egg comes first. We have to admit that we do not know; however, if we simply view the two positions in an *instrumental* way to help position ourselves, it seems that either choice may do. *To orient ourselves in the forthcoming discussion, we shall position our studies of reasoning strategies in the scenario of Becoming since our intended reasoning will be one that based on experiences.*

A position in *Becoming* presents us the second task to ponder on: what at root is *Becoming*, apart from the mere idea of certain *operations*? To answer this question, we have to know the *objects* upon which the operations are supposed to

[1]We have to concede that our general statements actually do not describe the full landscapes of both Platonism and Aristotelianism. The whole picture is a lot more complicated and each "ism" does carry certain elements of concern of the other; what we try to depict here are solely the distinct peculiarities underlying each approach.

[2]Rigorously speaking, our identification of *Being* with *objective* is really no good since *objective* is itself an earthly conception; in reality, we have to admit difficulty even in trying a description of *Being* as Kant in [2] has warned us that *Being* is something that is beyond the limit of *Reason*.

execute, and the precise details of the operations executed. We try to answer the first question in this chapter, the second will fill the body of the remaining text. To appreciate what the *objects* are, we look back to what and how our perception of the surroundings is like. To picture a series of phenomena, it is sometimes convenient to think of a *global* order intrinsic to the system as a whole, and according to which *local*, or *individual* phenomena follow in a *natural* and *coherent* manner; that is how we think of the physical world and it serves the fundamental spirit underlying all physical theories to date; we call this line of perception the *top-down* process since we take as the fundamental a *global* order. To the other extreme we have the *bottom-up* process in which we have concrete knowledge on certain *locally* restricted domains only, and an induced *global* order out of these *local* pieces is a result of how the *local* pieces interact with one another, so the resulting *global* knowledge may not be unique. Instances along this line are ubiquitous: in the human psychology of a population, the stock markets and the world politics, etc; well, we may conjecture some sort of *global* order, but that needs to be carefully justified in accordance with the rationales conjoining the *local* pieces. With regard to the line of reasoning we are going to work on, *we shall take a bottom-up approach since that seems to be what we are doing when confronting an ever increasingly sophisticated world with limited individual capability.*

As far as reasoning as a *bottom-up* process is concerned, we discern the *local* and *global* aspects with regard to the nature of information. To illustrate the point, let us consider a problem: "how do we infer the temperature of a distant star?"; this is clearly a *bottom-up* problem, as we can never place a thermometer in the star to take a direct measurement. To answer the question involves several aspects:

1. Electromagnetic signals emitted from the star to the earth which serve as the signature of the temperature of the star's atmosphere.

2. The collection of the star's signals by a telescope and the subsequent analysis by a spectrometer.

3. A tabulation of temperature against spectral distribution by on-earth experiments.

4. By comparing the star's signals with the on-earth experimental tabulation, the temperature of the star is inferred.

We regard each of the first three items as the *local* information which is obtained on the understanding of some very *closely* related physical phenomena: broadly speaking that involves electromagnetism, optical physics, and black-body radiation. It is through a series of such *local* studies that we build up a *global* knowledge base through which we answer the question *on assuming that the global knowledge is adequately represented by joining the local pieces of knowledge consistently at their common portions.* To rephrase it with regard to our question on the temperature of a distant star: we *infer* the temperature of the distant star by *measuring* its correspondence on earth; and by *assuming* that the *same* physics is going on no matter we are on earth or in the star, we draw the conclusion by consistently equating quantities from the star and from the earth. This example illustrates a typical *reasoning* scenario for all *bottom-up* arguments. However, we need to be careful of the pitfalls underlying the above reasoning: first, there is no guarantee that global knowledge equals the sum of its constituting local ones; second, even if it does, the meaning of the above "joining ... consistently" implies an agreement entailed under certain operation which can be something highly complicated. *Therefore, while we are following the bottom-up approach, we are not just working locally, but also globally with certain global operation which allows communication amongst the local pieces; and what we subsequently call global knowledge has to be checked with respect to a background of that global operation.*

2.3 Fundamental methodology of Reasoning

As the last characterization of our reasoning framework, we characterize the content of what we shall take as *information* and the nature of the associated *reasoning strategy*; by concrete realization of these two we would be able to render reasoning to effective manipulation.

The most fundamental framework of reasoning is right embedded in our daily language: for instance we have sayings like

1. If P is true and Q is false, an inference started from P and obtain Q cannot be true; or

2. All S has property A; given that s is a member of S, then s has A.

Other than their apparent familiarity, we notice specifically that both items show that there exist certain *forms* of reasoning that is *independent* of the precise details of what is going on[3]. In fact, this had already been noticed in ancient Greece, and the above *logical forms* of reasoning were examples of the fundamentals of the Stoic's school of logic and Aristotle's syllogism respectively. Sublimation to such *general and universal forms of reasoning* from pure on-earth *instances* is undoubtedly the first most decisive step towards a real take-off to a *science of reasoning.* However, what is the value of such *forms*? At a fundamental level, that gives us some *general* perspectives and directives through which problems can be dealt with. But this is not the full story yet, why? Because later we find that certain logical forms are liable to manipulation, specifically mathematical manipulation which can be implemented mechanically[4]. This endeavour was headed by Leibniz under which he proposed his logical calculus which unfortunately had not been too successful. The first significant breakthrough in this regard came with Boole, with whom came the celebrated Boolean Algebra which forms the fundamental of contemporary computer science. Boole's success had its significance not just in proposing his Boolean Algebra, but more importantly it had encouraged a whole industry grounded on the studies of *forms* or *structures* to blossom; important advances in this line include what are today known as propositional logic, predicate logic, etc which served the fundamentals of the early rule-based approach to artificial intelligence (AI). Important though all such advancements due to the manipulation of *forms*, do they mark the end of the story?

Later it has become evident that straight logic does not seem to correspond to the real world very well: misunderstanding of instructions by AI machines is an infamous problem plaguing the whole AI community. So, what's wrong? Going back to the beginning of our discussion, we saw that the ancient Greeks extracted *forms* out of *instances*, and then we had a series of subsequent sublimations along this line; but, stop and think more clearly: do *forms* really represent all? No! A common occurrence in daily language is that the same word can mean differently in various *contexts*; the most notorious might be poems in which we have all sorts of rhetorical techniques which render understanding impossible if resorting to straight

[3] A casual "proof" of this statement lies in the fact that we have invoked in the two statements only symbols P, Q, etc without picking up particular incidents, yet the whole thing is appealing, though may not be entirely comprehensible, to most readers.

[4] Actually, a deep aspiration underlying this endeavour is to ground reasoning, in the form of relationships expressible by logical forms, on an objective and universal footing so that it can be cast into unambiguous operations in the same way as Newton's laws of motion prescribing a deterministic world picture.

logic. Thus, there must be something intrinsic to *instances* which cannot be adequately captured by *forms*. So, an *instance* is really like a human being: while we as human share similar physical and mental characteristics, it does not deprive us from being individuals; thus while it is true that certain general *forms* exist amongst *instances*, every *instance* is intact in its own *integrity* that it cannot be engulfed by some simple total of *forms*. Thus, what seems to be the true story is that strict logic will never be adequate for *reasoning* purposes. The common way out is to "loosen" conventional logic a bit, from that come situational logic, fuzzy logic, probability based reasoning schemes like Bayesian reasoning, game theory, etc; alongside there are other "physical" approaches such as neural networks which work in a "curve-fitting" manner to future prediction, and genetic algorithms, etc. The present status of the art is still far from mature, with each approach dealing only with a facet of the problem. We do not know whether we can reach a *universal goal of reasoning*; to be honest, we do not even know whether such a goal really exists or is simply a conceptual ideal. Perhaps, as every methodological approach is essentially just a certain set of *operations* which is tantamount to merely some manifestations of *Becoming*, we would never be able to reach a *universal goal of reasoning* that belongs to the realm of *Being*.

2.4 Our intended approach

Let us review for a moment what we have come across so far. In the beginning we distinguished between *Being* and *Becoming*, and as with all natural sciences we positioned ourselves in the tradition of *Becoming*. Then we investigated the constituent of *Becoming* with reference to the nature of information, and we categorized information into *local* type and *global* type. Our task in this project is to discuss reasoning problem in the following line: we start from knowledge of *local* information, and we would like to know what *global* characterization could be reached out of it – by following this *bottom-up* line, our concern centers on the *operations* executed amongst the pieces of *local* information which give rise to certain *global* order. In order to identify the historical endeavours devoted to this task, we skimmed through a history on the methodology of reasoning problems: from fundamental logical forms embedded in our daily language to contemporary mathematically and/or logically oriented endeavours of decision sciences.

We shall position ourselves in one of the main streams of modern decision sciences, namely *probabilistic reasoning networks.* Actually, we shall target at a specific task on discussing the possibility of cyclic probabilistic reasoning structures (whose special roles will be explicated along the development). To make our studies concrete, we actualize our cyclic reasoning structures to error-control application. Our discussion will be based on a background of all aforementioned concerns; by a clear understanding of such background coupled with suitable exploitation of the freedom allowed, we shall establish some probability based error-control structures which are amenable to not only efficient computation, but also to immense conceptual clarity that enhances understanding.

Chapter 3
Probabilistic reasoning networks

3.1 Overview

As introduced in chapter 2, we shall pick probabilistic reasoning networks as our fundamental approach to the reasoning problem. We shall lay down the fundamental mathematical models in the present chapter. In particular, we shall trace the developments leading to the Bayesian network (BN) and the semi-lattice (SL) approaches, both of which are basic to the contemporary state of the art as introduced in Pearl [9] and Jensen [1] [1]. Our emphasis is to see *why* and *how* these approaches have evolved into its present shape; by so doing, we want to point out certain inadequacy and limitations imposed upon the constructions. And then in the forthcoming chapters, we shall try to lift such constraints and launch another program to a different probabilistic reasoning strategy.

3.2 Causality and influence diagrams

As a fundamental human conception towards his surroundings, inspired primarily by a logical ordering of happenings in nature, there is a drive to every mind to organize his daily observances in certain orderly and correlated manner. For instance, a pin pricks and causes pain, or a running ball hit by another swerves away from its original path. This is the background under which *causality* is devised as a natural means to *impose order* on some possibly correlated happenings around; consequently structures like "A *causes* B", "*if* A, *then* B", and so forth have flourished into some the most fundamental elements of our everyday language. For instance, we have "fever *causes* a temperature", "*if* there is an external net force acting on an

[1]The SL approach is termed under the name "Junction Tree" in Jensen [1] which is a speical case of the more general SL introduced later.

object, *then* the object accelerates", etc. Of course, most correlation relationships are not so direct and explicit, but are revealed only through an indirect means; say while a higher rate of divorce has been recorded with couples married at an younger age, it is by no means true that early marriage causes divorce (at least not in a direct sense with which the word "cause" is most often invoked) – such correlation is statistical in nature which should not be cast into a "cause-effect" perspective in the usual sense. Thus, the simple word *"correlation"* can carry very distinct characters which should not be confused; however, as a matter of convenience in the development of this thesis, we shall skirt over this subtle, though important, discrepancy. Whenever we come across "correlation" or "cause-effect" in future, we shall assume both possibilities without explicit stating.

While a *cause* causes *an effect* is an appealing way of conception, in practical daily situations, we rarely face merely two simple events with which a causal role can be conveniently attached; rather the real world is an entanglement of bunches of causes and effects and each effect can serve as a cause to some other effects. For instance, both facts of "I woke up late in the morning" and "There was a car accident on my way to school" can contribute to my being late for school, then my late for school can further contribute to my failure in obtaining a copy of the lecture note, and then.... Therefore, to render such a sophisticated situation intelligible, the so called *influence-diagram* is born; an example is as follows. Suppose we have a set of events $\{A, B, C, D, E, F, G, H\}$ in which a set of causal relationships read

1. $A \to B, C, F$;
2. $B \to D, E$;
3. $C \to D$;
4. $D \to H$;
5. $F \to G$;
6. $G \to H$.

where a "$\to$" means figuratively "as a cause to". A corresponding influence diagram is depicted in Fig.3.1 which is a simple directed graph, on which each node represents an event and each directed arrow carries the figurative meaning of "as a cause to". So, on an influence diagram, a network of causal influences can be *visualized.*

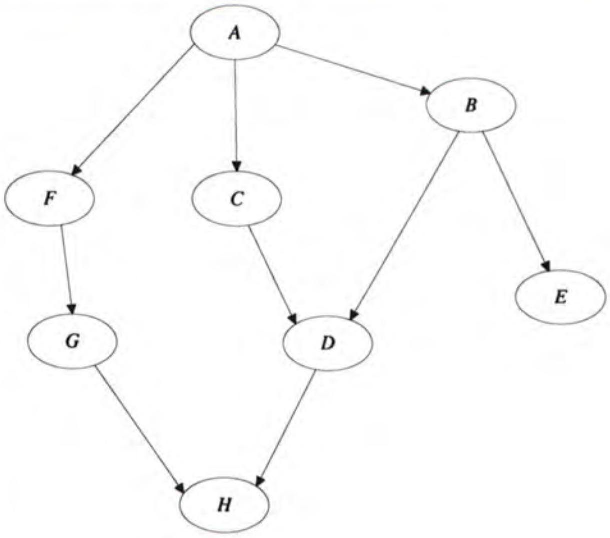


Figure 3.1: An influence diagram depicting the correlation relationships amongst a set of events.

3.3 Bayesian networks – influence diagrams endowed with a probability interpretation

Influence diagrams serve a purpose to *visualize* a network of causal relations, however we do not have yet a language to *talk about* the relationships represented on such diagrams. Since our fundamental concern is on reasoning amongst a set of events, in particular, we shall be concerned with the *likelihood* of occurrence, a useful language to deploy is *probability* – a branch of mathematics that deals specifically with likelihood in a crude sense. The only problem is how this language – probability – should be adapted to our present situation. To prepare the ground for the forthcoming work, let us take a detour on a brief discussion of probability.

3.3.1 A detour to the interpretations of probability

Probability is, in most textbook introduction, a quantitative way to talk of the *likelihood* of occurrences of a random variable. So a natural and intuitive way to get a measure of it is to perform a *large* set of identical experiments/observations on a certain situation upon which everything except the interested random variable is kept constant, then the probability of the random variable taking up a particular value is assigned to be the relative frequency that particular value shows up in the whole set of experiments/observations. This approach gives rise to what is known as the *ensemble-frequency interpretation* of probability – an interpretation based on objective experiments and counting; however, it soon becomes obvious that such ensemble-frequency interpretation is not adequate. One reason is that very often we do **not** have access to a *large* number of experiments which is essential to evaluating a limiting frequency, say we are not going to perform many experiments to see whether a particular government policy affects HK's economy for obvious reasons. Moreover, we have an insurance company talking about the expected probability that a man die at age 30 is something; how can a number that is measured from a total population be applied to a single individual? Such probabilistic assignments, quite inadequately grounded on an objective ground, are usually attributed to the *subjective interpretation* of probability; meaning that *probability* here represents a sense of *belief* rather than an objective measurement. To exacerbate the matter further, quantum theory, which is widely claimed the ultimate theory of nature, incorporates probability as the most fundamental mode of description inexorably. While quantum theory predicts the statistical outcomes of physical experiments with undeniable accuracy, an objective interpretation to the associated probabilistic assignments is thus plausibly adequate; however, if we apply quantum mechanics to the whole universe, can a probabilistic assignment mean an averaged result out of measurements on an ensemble of un*i*verses?

Probability may legitimately be termed one of the most intriguing concepts ever devised in human civilizations. In fact, compared with geometry, analysis so forth which had gained very strong footing by the end of the 19^{th} century, probability as a mathematical discipline has not been unequivocally established even now. What we have are basically certain formalizations due to Von Mise, Kolmogorov, de Finetti, etc.; at the fundamental root of the issue we do not even have a unanimous agreement on the very *concept* of probability. By a *probabilistic reasoning network*, we shall encounter a complicated issue mixing up the above concepts of probability: on one hand we have *objective measurements* as incoming evidence and on the other hand, we have *belief* as the final output. We apologize that by this project we shall not be able to soothe the bitter issue, rather, we shall show one more example of the uncompromising conflict between the different conceptions.

3.3.2 Bayesian networks

Returning to our earlier intention to endow a probability interpretation on an influence diagram, we ask what and how the constituents on the diagram should bear a meaning of probability. As it is the likelihood of events which interest us, it is thus natural to assign a probability at each node characterizing the likelihood of the possible values that event consumes. Next, to correlate two events on an influence diagram in a probabilistic sense, it is natural to conceive each directed arrow as carrying the conditional probability of the event at the arrow-head with respect to the one at the tail – a natural parallel to the notion that the tail's event *influencing* the head's. An influence diagram, endowed with a probability assignment this way, is what known as a *Bayesian network* (BN)[2]. So suppose we have a causal event of C influencing E, with the conditional dependence of E on C being given by $P(e|c)$ [3]; the corresponding BN is depicted in Fig.3.2 below.

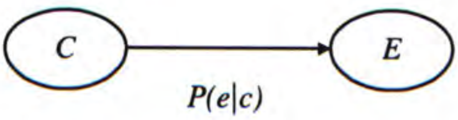


Figure 3.2: A simple BN with C influencing E through the conditional probability $P(e|c)$.

In more complicated circumstances such as that in Fig.3.1, the conditional probability assignment is more intricate: if an event E is caused by two causes, C_1 and C_2 say, then the conditional probability **must** take care of both causes **simultaneously**, so we have $P(e|c_1, c_2)$ – **not** $P(e|c_1)P(e|c_2)$ – even though the two causes may be independent of each other. An example is a betting game: we toss two dices, which serve as two independent causes, and use their sum to determine whether one wins or loses; clearly, while the causes are independent, the outcome none the less depends on both causes jointly. So a corresponding BN for the influence diagram of Fig.3.1 is given as Fig.3.3.

[2]Actually, there is a further constraint on the directed graph being acyclic, with which the significance will be addressed later.

[3]A note on notation: we use *upper-case* letters to stand for events and *lower-case* letters to stand for instances of events.

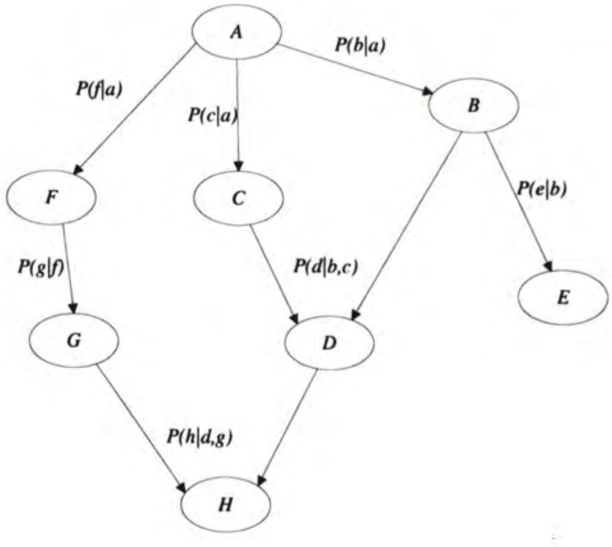


Figure 3.3: A more complicated BN

To have a formal statement to the construction of a BN, let us denote by X_n the nodes on a BN and Φ_n the *parents* of X_n which are the nodes from which arrows terminating on X_n are emanated, then we have

Definition 3.1 *Bayesian Networks (BN):*
Let there be a set of N events, the universe $U = \bigcup_{i \in I} X_i$ *where* X_i *stand for the individual events and I is the index set* $\{1, 2, \cdots, N\}$. *A Bayesian Network (BN) is an acyclic directed graph (DAG) where each node stands for an* X_i *and each arrow running from* X_m *to* X_n *represents an influence* X_m *has on* X_n. *Besides, each node* X_i *is characterized by a local probability of occurrence* $P(x_i)$ *and every group of arrows from* Φ_n *to* X_n *is characterized by a conditional probability* $P(x_n|\Phi_n)$.

In the above construction, a BN is constructed from individual nodes and arrows, so our construction proceeds from knowledge at and amongst individual events, to a universe comprising all the events. Of course, we can go the other way round: with knowledge of a universe, say a global probability distribution of all the events, we can analyze the statistical relationships amongst the constituent events, then a BN can be constructed. In the language of Chapter 2, the former is the *bottom-up* approach while the latter is the *top-down* approach. They carry distinct philosophies that we have noted and here we simply remark that it is the former *bottom-up* approach that we take in this thesis.

3.3.3 Acyclicity and global probability

Recall the two fundamental constituents in constructing the BN: 1.) local probability at each node, and 2.) conditional probability on each group of arrows. A natural question thus arises: what is the global character of the BN in a probability tongue? Well, as both constituents utilized are both *local* in character, in general we cannot say much on a global setting. This situation changes if we impose a *global* constraint, which customarily is taken to be a requirement on the BN being *acyclic* – absence of *directed loops* – as stated in the previous definition. Under this constraint, there is then a *global* probability $P(u)$ *pertaining* to the universe, which with respect to definition 3.1 is given by

$$\begin{aligned} P(u) &\equiv P(x_1, x_2, \cdots, x_N) \\ &= \prod_{i \in I} P(x_i|\phi_i) \end{aligned} \tag{3.1}$$

which can be verified to satisfy the usual properties of a probability distribution like $\sum_{x_1,x_2,\cdots,x_n} P(u) = 1$, etc. Here we heed the special role of *acyclicity*: the absence of directed loops prohibits conditional dependence from a node from feeding back to itself, hence changing itself; thus, it allows a universe to be gradually built from an existing sub-universe – in the sense that a global probability of that portion has existed – through the conditional probabilities connecting in-between. As a summarizing statement, acyclicity induces *a consistent universe, a consistent piece of global information* to be established; this construction is entirely harmonious, one-way and direct; however, does this picture suffice?

3.4 Reasoning on probabilistic reasoning networks I – local updating formulae

Next, we turn to another important topic of BN, namely information *updating* or *inference making*. This task is primarily prompted by a desire to *"infer"* in a causal event the cause from the effects produced, upon a knowledge of the effects through suitable observations. The scenario covers two aspects:

1. A *local* updating formula for each node – an individual event – on receiving new information from neighbouring nodes.

2. A *global* scheduling of updating across the network when certain events within the network are activated by reception of new information.

We shall only discuss the local updating formula in this section, and defer the discussion on global scheduling to a later one. In the present development, we shall emphasize the fundamental rationale underlying our intended *reasoning* which explains *why* and *how* we proceed in the prescribed way.

3.4.1 Rationale of the intended reasoning strategy

An important theme of our chosen reasoning strategy is that we would try to impose a *symmetry* concern on local updating; the reason behind is that on a complicated network most nodes are simultaneously acting as a cause to some nodes and an effect of some other nodes, however, as is usually the case, it **needs not** always be clear and definite which role a node takes in most circumstances; to exacerbate matter further, the causal roles between two events may not be fixed! For example, the two events "early marriage" and "divorce" carry certain statistical correlation that "Couples married at a young age show a higher rate of divorce."; suppose I meet a middle-aged couple who say that they have divorced, then I can think two ways: 1.) They married at a young age, then they have a higher probability to get divorced; or 2.) they have divorced, it is likely that they might have married at a young age. So inferences could be made either ways with respect to the statistical statement aforementioned. The point we try to illustrate is that *under many circumstances there is always a certain degree of arbitrariness in assigning a causal role that depends on how we perceive of the situation: from what starting ground, with what perspective so forth.* Thus when we build up a BN as prescribed before with the arrows *chosen* to respect the global acyclic constraint, we have to be careful that we are actually *limiting* ourselves to *a single*, amongst *many possibly equally probable ways*, to perceive of the situation as a totality – that is precisely what the *dictated global probability* (3.1) means. So, in short, *while a statement on correlation alone is neutral and a natural way to reason, an imposition of causal roles could be nothing more than a mere predilection.*

3.4.2 Construction of the local updating formula

To deal with the first task of establishing a local updating formula, with regard to the *symmetry* concern just mentioned, we deploy the Jeffrey's rule of information updating[4] as our key inference-making device which is stated as follows,

Definition 3.2 *Jeffrey's rule of information updating:*
Given a universe of two events A and B related by a prior conditional probability $P(b|a)$; if, upon suitable measurements/observations it is found that event A shows up with a posterior probability distribution $P^*(a)$, then the posterior probability distribution for B, i.e. $P^*(b)$, is determined by the following relationship,

$$P^*(b) = \sum_a P(b|a)P^*(a) \tag{3.2}$$

The underlying spirit of Jeffrey's rule is that an absorption of evidence *does not* alter the correlation between the two events A and B, that is why we can invoke the *same* $P(b|a)$ upon *inferring* B from A. More precisely, Jeffrey's rule calculates the *posterior* information from the *prior* information on the *assumption* that the posterior information *does not* alter the prior conditional-dependence information; so we assume the conditional probability $P(b|a)$ to stay unchanged upon the attainment of $P^*(a)$ [5]. The question then follows is the way to adapt the Jeffrey's rule to our present concern, this can be illustrated by a simple example. Consider again the situation in Fig.3.2 where we have constructed a universe, composed of two events C and E, through the conditional probability $P(e|c)$. Now, a forward reasoning of inferring the probability of E, i.e. $P^*(e)$, from a knowledge of C — $P^*(c)$ — is trivially given by

$$P^*(e) = \sum_c P(e|c)P^*(c) \tag{3.3}$$

Next, we revert the question and ask: What is the posterior probability of C, $P^*(C)$, provided that an evidence on E — $P^*(e)$ — is received? By invoking

[4] Adapted from Pearl [9] P.62-70.

[5] It ought to be stressed that while the assumption seems not bad in our daily world, such expectation does not hold in general; for instance, in a universe of quantum particles, a measurement will in general introduces disturbance to a system that renders prior conditioning information of that system untrue *after* the measurement.

Jeffrey's rule of updating, we have immediately

$$P^*(c) = \sum_e P(c|e)P^*(e) \tag{3.4}$$

However, (3.4) does not make up the whole story: how can we determine $P(c|e)$? The trick here is the Bayes' theorem on probabilistic relationships, which states

$$\begin{aligned} P(c|e) &= \frac{P(c,e)}{P(e)} && (3.5) \\ &= \frac{P(e|c)P(c)}{P(e)} \\ &= \frac{P(e|c)}{P(e)}P(c) && (3.6) \end{aligned}$$

from which $P(c|e)$ can be derived from a knowledge of $P(e|c)$, $P(e)$ and $P(c)$, which are *completely* specified in the set of prior information.

Two important points worth emphasis in the present development:

1. The subjective character of information updating. Our present development involves an active manipulation of probabilities, unlike the objective interpretation of probability which is a passive counting until a limiting frequency is obtained. Probabilities upon active manipulations have to be interpreted in the subjectivist tradition; which is obviously an adequate tradition to situate our reasoning strategy.

2. We have hitherto deployed in our development the conditional probability as a starting point to build up the global universe of information. While this approach is intuitive, and probably is the easiest way to proceed in daily thinking, it is nonetheless doubtful whether the conditional probability really enjoys a fundamental role: the reason is that an assignment of a conditional probability is determined in turn by a joint probability from which we have

$$\begin{aligned} P(b|a) &\equiv \frac{P(b,a)}{P(a)} \\ &= \frac{P(b,a)}{\sum_b P(b,a)} && (3.7) \end{aligned}$$

   Thus what follows from (3.7) is a fundamental role played not by the conditional probability, but by the joint probability; in an operational perspective the joint probability bears the crude measurement results whereas the conditional probability is a *derived* concept which *imposes* a *causal order* amongst the events observed [6]. Same consideration is also evident directly from (3.6) where the conditional probabilities are tantamount to two different counting on the *same* joint probability – they are nothing more than the two sides of the same coin!

As a consequence of the symmetry embedded in (3.6), we conclude that (3.3) and (3.4) are actually one; from now on, we update not $P(c)$ or $P(e)$ individually, rather we update the joint probability $P(c,e)$ directly as follows,

$$P^*(c,e) = P(c,e)\frac{P^*(u)}{P(u)} \tag{3.8}$$

where $''u''$ stands for either $''c''$ or $''e''$ depending on which one is the receptor of new evidence; from (3.8) $P^*(c)$ or $P^*(e)$ are immediately obtained by usual marginalization. We call (3.8) the *absorption map* which can be easily shown to exhibit the following properties:

1. Updating of the joint probability preserves the new evidence: if the new evidence is $P^*(e)$, $u = e$ and it follows by simple arithmetic that $\sum_c P^*(c,e) = P^*(e)$; hence the evidence itself does not change on updating the original joint probability – this justifies "absorption" in the absorption map.

2. Let us take $u = e$, then it follows from (3.8) that

$$\begin{aligned} P^*(c|e) &= \frac{P^*(c,e)}{P^*(e)} \\ &= \frac{P(c,e)}{P(e)} \\ &= P(c|e) \end{aligned} \tag{3.9}$$

   while in general $P^*(e|c) \neq P(e|c)$. Thus conditional dependence is preserved one-way under the absorption map.

The crucial point to notice here is that the local updating formula (3.8) is founded on *joint probabilities* which carry no preferred causal-role assignments,

[6]This distinction becomes very acute in quantum mechanics; for example, in Khrennikov [3], it is argued that it is due to an improper incorporation of conditional probabilities which has induced the violation to Bell's inequality by quantum mechanics. Such violation carries the important consequence that the quantum world is fundamentally *non-local*, hence our beloved notion of *(Einstein) locality* is wrong, meaning that there can be some form of physical "influence" which propagates at a speed faster than that of light!

reflecting our earlier intention to drop the more or less conceptual artifact. We shall see later that this change in perspective gives rise to an alternative way to deal with the reasoning problem which will form the hard core of this thesis. In the mean time, we just stress that our focus from now on will shift from "two nodes connected by a conditional probability" on a BN to a joint probability encompassing all three.

As a further remark which serves another hard core of our thesis, the absorption map can easily be generalized to a universe with more than two variables by simply regarding "c" and "e" as sets of variables like the individual components of a vector; however, a more intricate question comes when there are several sources of updated information feeding in. Suppose with respect to (3.8) we have multiple independent sources contributing to $P^*(e)$, namely

$$P^*(e) = \prod_{i=1}^{N} P_i^*(e)$$

Then (3.8) is adjusted to

$$\begin{aligned} P^*(c,e) &= P(c,e)\frac{P^*(e)}{P(e)^N} \\ &= P(c,e)\prod_{i=1}^{N}\frac{P_i^*(e)}{P(e)} \end{aligned} \tag{3.10}$$

The rationale behind (3.10) is that if it happens that the pieces of new evidence $P_i^*(e) = P(e)$ for $i = 1, 2\cdots N$, which is tantamount to a situation that the new evidence serves a confirmation to, rather than a revision of the old knowledge, then $P^*(c,e)$ should simply be reduced to $P(c,e)$. Similarly, in occasions that there are two pieces of evidence on different variables feeding in, say if we have both $P^*(c)$ and $P^*(e)$, then we may treat them independently and write the updating formula as

$$P^*(c,e) = P(c,e)\left[\frac{P^*(c)}{P(c)}\right]\left[\frac{P^*(e)}{P(e)}\right] \tag{3.11}$$

Augmentation of (3.11) due to multiple independent sources of new evidence can be implemented in parallel with (3.10); so the essence underlying the absorption map is the updating factor $P^*(u)/P(u)$ in (3.8) for *every* independent source of evidence. We conclude this section with a formal definition of the absorption map:

Definition 3.3 *Absorption map:*

Let there be a set of N events, the universe $U = \bigcup_{i\in I} X_i$ where X_i stand for the *individual events and I is the index set $\{1, 2, \cdots, N\}$, and upon U is defined a probability distribution $P(x_1, x_2, \cdots x_N)$. Suppose to U is fed a set of evidence $\{P^*(e_j)|E_j \subset U\}$ where E_j are not necessarily disjoint, then the updated joint probability distribution reads*

$$P^*(x_1, x_2, \cdots x_n) = P(x_1, x_2, \cdots x_n) \prod_j \frac{P^*(e_j)}{P(e_j)} \tag{3.12}$$

3.5 Cluster graphs – another perspective to reasoning problems

In the last section, we discussed the local updating formula on a BN, which culminated in the updating formulae (3.8) to (3.12). The fundamental spirit underlying is that rather than reasoning on a "cause-effect" predilection dictated by the directed arrows on a conventional BN, reasoning can also proceed in a manner that treats the "cause" and the "effect" on the same footing, thereby depriving the probably unnatural causal assignment of a fundamental status in our discussion. However, the foregoing discussion focuses on a universe with two variables only; thus in order to boost our picture to a global setting of many varibles, we need a way to distinguish pieces of "joint probabilities" out of a given BN; each such joint probability corresponds to a sub-piece of the global BN which is an independent sub-universe within the global universe. The following scenario is proposed[7]:

Identification of the independent sub-universes from a Bayesian Network (BN):

1. *Given a BN with nodes $X_i, i = 1, 2, \cdots, N$ and parent set Φ_i mediated by a set of conditional probabilities $P(x_i|\phi_i)$ in between, we proceed as follows:*

2. *For each node X_i, add edges between every pair of parents if they are not connected.*

3. *Ignore all arrow heads on the original BN, so what left is an undirected graph with the same topology as the original BN, together with a few added edges. The graph resulted is called the moral graph of the BN concerned.*

[7]The earliest approach along this line can be found in Lauritzen and Spiegelhalter [5] and similarly in Jensen [1]; however, notice that we are doing *less* than theirs in the present treatment, namely we do not strive for a globally *acyclic* structure.

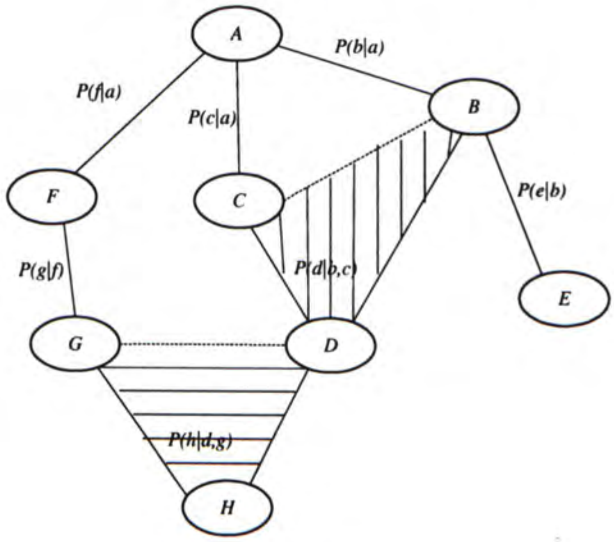


Figure 3.4: The moral graph corresponding to Fig.3.3.

4. *In the moral graph so formed, There are probably several fully connected sub-graphs associated with each node, each of which we shall call a clique. Each clique* $U_i \equiv \{X_i \cup \Phi_i \subseteq U\}$ *stands for an independent sub-universe – an independent set of measurements on the collection of variables* U_i *which incorporates the information* $P(x_i, \phi_i)$.

5. *Pick any ordering on the nodes, and read off every distinct clique corresponding to each conditional probability.*

As an example, the moral graph of Fig.3.3 is shown in Fig.3.4, with which we note that there are two added edges BC and DG since D has both B and C while H has both D and G as their parents respectively.

From Fig.3.4, the cliques identified are AF, AC, AB, BE, BCD, FG and DGH. These cliques can be arranged in accordance with the BN's topology from which they are derived; we shall call the graph resulted a *cluster graph*[8]. The cluster graph corresponding to Fig.3.4 is shown in Fig.3.5. A cluster graph is a hypergraph which depicts how a collection of subsets of a graph join together. Two neighbouring cliques are connected by their intersecting elements which serve as the "common knowledge" between the cliques. As an additional structure, each clique on a cluster graph inherits from the original BN a joint probability of the

[8] A special type of cluster graphs, namely the *acyclic* cluster graph, is customarily called the *junction tree* in the Bayesian-network community.

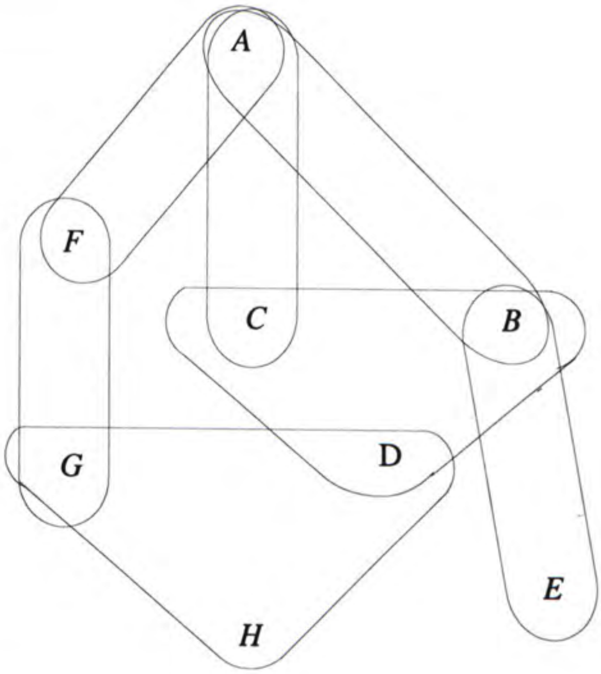


Figure 3.5: The cluster graph corresponding to Fig.3.4.

variables contained therein.

Cluster graphs convey a clear picture to conceive of the construction of the underlying BNs in an *operationally fundamental* point of view: each local sub-universe stands for a test, an experiment or a set of observations on the variables of that sub-universe; then in order to comprehend a global universe from its constituting pieces of local information, we join the local sub-universes together at their intersecting portions. Compared to a BN, we have instead of the following two constituting structures:

1. local probabilities at individual nodes, and

2. conditional probabilities of each node with respect to its parents;

the following parallel structures on the corresponding cluster graph:

1. local joint probabilities on each clique, and

2. a global topology according to which the cliques join together.

The distinction between these two perspectives is non-trivial: the cluster graph perspective carries with it no directionality which as previously argued is a mere conceptual construct, hence it is a more fundamental picture. As a matter of fact, a cluster graph is free from the BN's global acyclic constraint, as is evident on comparing Fig.3.3 and Fig.3.5. Recall that the global acyclic constraint underlying a BN contributes certain advantages, say the existence of a global probability, hence bypassing this constraint implies that certain advantages are to be lost. This is going to be a game of "gain" against "loss"; and the fundamental purpose of this thesis is to see how far we can go when adopting the cluster graph as our fundamental picture in the reasoning problem.

3.6 Semi-lattices – another representation of Cluster graphs

Cluster graphs are a natural representation to depict how different sets of experiments/observations join together to constitute our conceptual world. However, it soon becomes evident that for a highly connected network there exists too many clusters and the graph will just be a mess; thus, to render analysis efficient we cast the whole thing into a more manageable and rigorous mathematical framework – semi-lattices (SL). As we shall take the SL solely as a way of representation and shall not invoke any algebraic structures entailed under them, we shall not delve into a formal definition but shall just give a schematic illustration on the construction of SL's pertinent to our use.

3.6.1 Construction of semi-lattices

First, we introduce a concept of ordering on set. Let there be a set of sets $V = \{Y_i | i \in I\}$ where I is some index set and Y_i are assumed distinct; upon V we introduce an ordering relationship "$>$" such that for any $i, j \in I$ we have $Y_i > Y_j$ if and only if $Y_i \supset Y_j$; analogously, we have "$<$" so that $Y_i < Y_j$ if and only if $Y_i \subset Y_j$. With respect to both "$>$" and "$<$", we call elements on the "smaller" side *successors* and on the "bigger" size *ancestors*. This concept of ordering introduces us a easy way to depict of the intersecting relationship within a general set of sets.

Suppose we have a set of sets $U = \{X_i | i \in I\}$, a *Hasse diagram* is a pictorial illustration of the intersecting relationships on U. The construction of a Hasse diagram goes as follows,

1. Form the *intersection closure* $\bar{U}$ of U – the intersection closure is a set which takes as elements *all* the intersecting products generated from the elements of U. For instance, suppose we have the set $U = \{ABD, BCF, ABCE, FG\}$, then the intersection closure is $\bar{U} = \{ABD, BCF, ABCE, FG, AB, BC, B, F\}$[9].

2. The Hasse diagram corresponding to U incorporates all elements in $\bar{U}$, arranging them on different levels as illustrated in Fig.3.6: on the top level we put in all elements of U, on the lower levels we put in elements of $S \equiv \bar{U} \setminus U$ such that when we cross the levels along any line from top to bottom the elements encountered form an ordered chain; for instance, from Fig.3.6 we have the ordered chains $ABD > AB > B$, $BCF > F$, etc. It suffices for our purpose to call a set-theoretic structure which can be represented by a Hasse diagram a *semi-lattice* (SL for short). Clear enough, if we take U to comprise all the sub-universes identified in a cluster graph, the resulting SL will be a systematic representation of the cluster graph, with each elements in U represented by an oval and each element in S represented by a square. It is important to notice that SLs actually carry rich information in their own right; however, as the present state of this project is far from exploiting the whole thing, we shall be content with taking them simply as a way of representation.

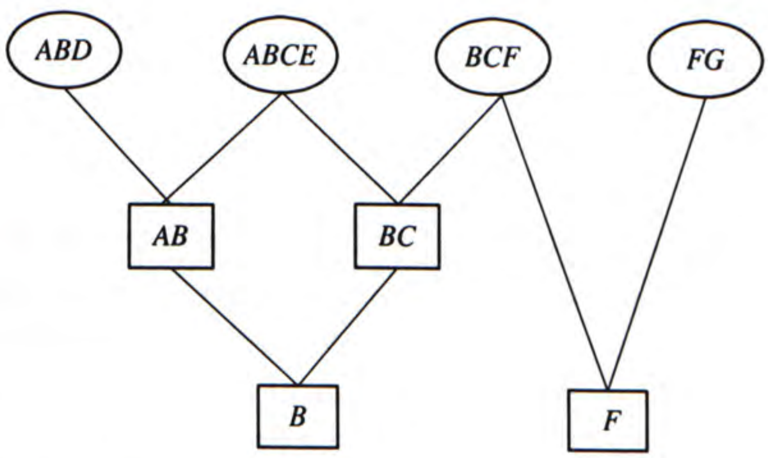


Figure 3.6: The Hasse diagram corresponding to the set $U = \{ABD, BCF, ABCE, FG\}$.

[9] The name "closure" suggests that the set is *closed* under the operation of intersection. Besides, as a note on terminology, elements in U are also called schemata while elements in $\bar{U} \setminus U$ are called separators.

3.7 Bayesian networks and semi-lattices

Recall that acyclicity underlying a BN permits an associated global probability, the associated SL resulted by the above transformation in general do not enjoy this privilege. In general, a SL permits a global probability only when the constituting sub-universes are joined in an acyclic manner – the acyclic joint dependency (AJD) condition. The AJD condition originates from relational database theory; for our project, it suffices to treat it purely as a condition on sets[10]. Graphically, AJD means that the local sub-universes are joined in an acyclic manner on the cluster graph, forming no loop; thus, the cluster graph on Fig.3.5 does not satisfy the AJD condition since there is a loop: $AC - CDB - AB$; however, if say AC and AB are removed, then the AJD condition is satisfied. In fact, the AJD condition plays an important role in the studies of relational database theory, to which SL's serve the fundamental mathematical objects for description. It has been rigorously proved in Lee [6] and [7] that for a SL satisfying the AJD condition there exists a global probability that can be expressed in terms of local ones and vice-versa; we refer to [7] for a proof on it, here we just write down the identification formula as discussed below. On a SL satisfying the AJD condition, for each element of $S = \{\xi_j | j \in J\}$, we identify a characteristic number c_j defined as follows: c_j is the number of connected components on the SL which carries ξ_j upon the deletion of ξ_j and all its successors. So with respect to Fig.3.6, $c_{AB} = 2$ and $c_B = 1$, etc. By this identification, the global probability can be written as a product of the local ones as

$$P(u) = \frac{\prod_{i \in I} P(x_i)}{\prod_{j \in J} P(\xi_j)^{c_j - 1}} \tag{3.13}$$

where u stands for an instance of $\cup_{i \in I} X_i$ [11];thus for example, we have for the

[10]The set operation which checks whether a given set of sets satisfy the AJD condition is the Graham-reduction procedure: Given a set of sets $U = \{X_i | i \in I\}$,

1. for any $j \in I$, delete X_j if $X_j \subseteq X_i$ for some $i \in I$; and
2. for all $i, j \in I$, delete $R_{ji} \equiv X_j \setminus X_i$ if $R_{ji} \subseteq X_k$ for no $k \in I$.

It can be proved that U is Graham-reducible to the null set if and only if U satisfies the AJD condition.

[11]A point we have skirted over in (3.13) is that on constructing the global probability from local ones we must have the *consistency* condition saying that neighbouring sub-universes marginalize to the same probabilities for their intersecting elements.

situation in Fig.3.6,

$$P(abcdef) = \frac{P(abce)P(abd)P(bcf)P(fg)}{P(ab)P(bc)P(f)} \tag{3.14}$$

An interesting application of the global probability (3.13) is to derive the local updating formula studied in section 3.4.2; an example in this line is done in Appendix A with the situation depicted in Fig.3.6 and the global probability (3.14).

3.7.1 Bayesian networks to acyclic semi-lattices

Recall that our former transformation from a BN to a SL does not guarantee an acyclic SL to be resulted. Without acyclicity, (3.13) does not apply to the SL resulted and that will not be a pleasant situation to deal with. Therefore, to circumvent this difficulty, additional procedures called triangulation is commonly introduced to make the resulting SL acyclic such as done in [1]. We shall not show the construction and just refer the interested readers to [1]. What is important here is that it is right at this point that our proposal branches out: *we shall not introduce triangulation to ensure acyclicity on the resulting SLs, we start from where we end up with after transformation!* Actually, as our title suggests, we are interested in making cyclic SLs our probabilistic reasoning framework which will be the task in the remaining chapters of this thesis.

3.8 Reasoning on (acyclic) probabilistic reasoning networks II – global updating schedules

The global updating schedule is the second step towards an overall information updating of the information network. However, unlike the local updating formula which is essentially local, the global updating schedule is global in character and hence needs to take into account the global topology of the information structure. We shall devote discussion to global scheduling on an *acyclic* topology in this section, the *cyclic* counterparts will be the target in later chapters. Here, since what matters is the topology, we do not have to differentiate between BNs or acyclic SLs, but simply bear in mind that it is with respect to each node for a BN, and each sub-universe for an acyclic SL that acts as the fundamental units for sending and receiving information.

Acyclic topologies give rise to especially easy ways to possible global scheduling. The reason is that *acyclicity* implies definite beginning and end: the simplest case might be a string for which the two ends play the obvious roles. In general, global scheduling can be scheduled according to the following rationales (cited from Jensen [1]),

Definition 3.4 *Message Passing Scheme (Jensen [1] P.75):*
A node[12] *V can send exactly one message to a neighbour W, and it may only be sent when V has received a message from each of its other neighbours.*

This Message Passing Scheme carries the essential properties that mark the definite beginning and end of the overall information updating as proved in Jensen [1]. Here we shall not delve into a discussion of it but simply notice the character of global scheduling here: basically it is an *economy* concern which requires each node to absorb *all* evidence prior to sending information to other uninformed nodes; of course, this statement is justified only on the ground of *acyclicity* since otherwise the global knowledge base will not allow a clear cut into definite *updated*, *partially-updated* and *non-updated* pieces. In an alternative perspective, *acyclicity* means that there is no feedback of information, so a piece of new evidence runs into the information structure and remains unchanged if there is no other evidence. Clearly, we shall not be able to incorporate such scheme into *cyclic* networks.

3.9 Conclusion

We have traced the two most important developments in the upcoming of probabilistic reasoning networks: BNs and SLs; and have talked about their intimate relationships. While BNs serve the conceptually fundamental picture to a reasoning strategy, we argued that it is the approach based on SLs which is operationally fundamental and natural. So we focused ourselves on SLs. However, upon a comparison between the two we knew that their mutual relationships are really not that direct, with differences come up in the global topological characteristics of the respective structures upon transformation. *While conventionally we can introduce some operations to render them on the same footing, we just emphasize that we shall not invoke them in our proposal but be content with what the natural con-*

[12] A node is respectively taken as a node for a BN or a sub-universe for an acyclic SL in our terminology.

struction gives rise to; and it is the endeavour in the forthcoming chapters to build up a reasoning framework starting right at this point.

Chapter 4
Cyclic reasoning networks – a possibility?

4.1 Overview

The past chapter has introduced the general landscape of contemporary probabilistic reasoning networks. The main line of thought, in both the BN and the SL approaches, is that in certain occasions a *global* piece of knowledge (in the form of a global probability) of a universe can be *decomposed* into that of smaller pieces of constituting *local* sub-universes, and in the other way round from pieces of *local* knowledge we can *build up* a global piece of knowledge of the universe. The constraint which defines such occasions is that the local pieces have to be joined in an *acyclic* manner – directed *acyclic* graphs for BNs and *acyclic* joint dependency for SLs. This in essence means that there is no information feedback, so that a piece of new evidence is fully trusted and overrides all past experiences; however, is it really the *usual* way that reasoning proceed? Is it really meaningful? In this chapter, we shall try to lift the constraint of *acyclicity*, and try to foresee whether it is a possibility to implement something like reasoning on a cyclic structure.

4.2 A meaningful cyclic structure – derivation of the ideal gas law

Let us broaden our view a bit first to appreciate that cyclicity does indeed pose something useful, and most importantly something *true*. As a first example, let us look at a physical case. In the 17^{th} to 18^{th} century, people were interested in how gas behaved. There were three parameters describing the state of gas well known by that time: pressure (P), temperature (T) and volume (V). So a natural question arose: how were P, V and T related? Therefore people conducted experiments to investigate their relationships. For the ease of experiments and analysis, relationships between two parameters were measured at a time on a fixed mass of gas and the following conclusions were drawn:

1. $PV = K_T$ where K_T means a constant of T; this is known as Boyle's Law.

2. $P = K_V T$ where K_V means a constant of V.

3. $V = K_P T$ where K_P means a constant of P; this is known as Charles' Law.

So we see the three parameters: P, V and T, are indeed intimately related. In order to understand the relationships amongst these three experiments, we cast them on a SL setting as depicted in Fig.4.1; further, it is easy to verify that the SL resulted is a cyclic structure as the set $\{PV, PT, VT\}$ is not Graham-reducible to the null set. In Fig.4.1, each of the above experiments is represented by a schema (an oval), and the common information amongst is represented by the separators (the squares). However, while the SL depicts the set-relationships from which we construct our universe, we do not have probability in the schemata and the links now; instead, what we have is a set of numerical experimental values. Besides, we understand consistency between schemata as their common elements taking up identical numerical values. Next, based on the information embedded in the SL, we derive the ideal gas law in the following discussion.

Since the above three experiments are only specific measurements on the same gas, all three relationships concluded above must hold simultaneously for the given gas in equilibrium. From the experiments, we have

$$\begin{aligned} PV &= K_T &(4.1)\\ P &= K_V T &(4.2)\\ V &= K_P T &(4.3)\end{aligned}$$

for a set of values P, V and T. By (4.2) and (4.3), we have

$$PV = K_V K_P T^2 \tag{4.4}$$

Comparing (4.1) with (4.4), we have

$$K_V K_P T^2 = K_T \tag{4.5}$$

Our derivation so far has based entirely on the relationships of each schema, together with the consistency of values imposed on understanding the SL of the three experiments. The next step is entirely a logical consequence of the mathematical equality as expressed in (4.5); we have, as the R.H.S. of (4.5) is entirely a function of T,

$$K_V K_P = K f(T) \tag{4.6}$$

where K is a constant independent of P, V and T and $f(T)$ is a function of T. Put (4.6) into (4.4), we have

$$\begin{aligned} PV &= K f(T) T^2 && (4.7) \\ P &= \frac{K}{V} f(T) T^2 && (4.8) \end{aligned}$$

On comparing (4.2) and (4.8), we have

$$\begin{aligned} f(T)T^2 &= T \\ f(T) &= \frac{1}{T} && (4.9) \end{aligned}$$

Lastly, put (4.9) into (4.7), we have the ideal gas equation:

$$PV = KT \tag{4.10}$$

which we are all familiar with. Notice that the above derivations proceed by identifying values P, V and T in the three schemata, i.e. the three experiments. In summary, what we have done are:

1. Extracting information from each schema, i.e. from each experiment.
2. Comparing information amongst the schemata by some consistency condition.

While the first step is within each schema, hence a *local* concern, the second step is the key to a *global* concern incorporating the several pieces of local information. Notice that we must have all three pieces of information in order to derive the ideal gas law, we cannot afford missing one!

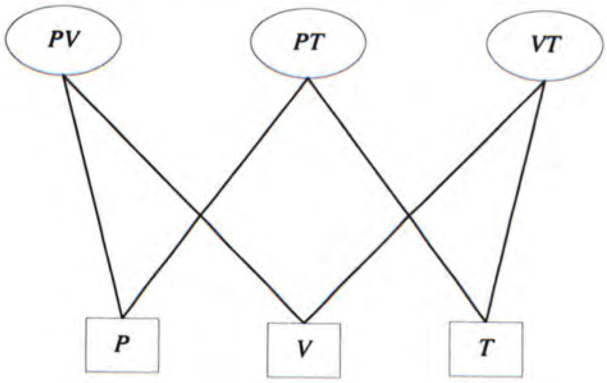


Figure 4.1: A semi-lattice on the sets PV, PT and VT.

Cyclic processes, more precisely processes involving feedback, are prevalent in the physical world. For instance, a water molecule with one oxygen (O) atom covalent-bonded to two hydrogen (H) atom exhibits a characteristic oscillation frequency because the dynamical feedback of forces on the H-O-H structure picks out that characteristic frequency *pertaining* to the structure itself[1]. Of course, the sense of "feedback" in physical processes is different from that of a reasoning process; however, the point raised here is an emphasis on the background structure, the topology according to which the sub-universes join together. Recall how we have constructed an information network: we perform a set of experiments, and then "glue" up the sub-universes – the experiments – to construct the global universe. So, the background structure, the topology of how the sub-universes join together, enjoys a fundamental role. Probabilistic attributes are some further characterizations employed to make sense of the universe. In this perspective, it is not unreasonable to doubt why we need to confine ourselves to the *acyclic*-world when talking about reasoning process. Of course, that is not without a reason, the following is a partial account for this.

4.3 What's "wrong" to be in a cyclic world

The first thing one might immediately think of is a matter on consistency: suppose we start from some point in a cyclic world and go to another point in all possible paths available, a question immediately comes up is whether the information from

[1]In this case, it can be shown that the frequency is a function of the masses of the atoms and the parameters which characterize the bonding force between the atoms.

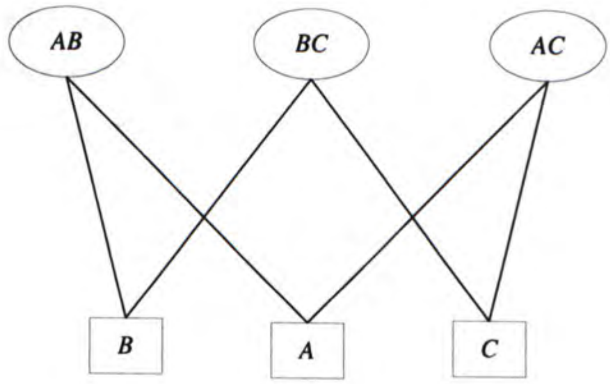


Figure 4.2: A semi-lattice on the sets AB, BC and AC.

different paths going to be consistent. The answer to this question is in general "no!". Suppose we propagate information on the SL depicted in Fig.4.2. Each variable A, B and C takes on a value of 0 or 1, with the information constraint within each schema be that the two variables be of different values, i.e. if one is 0 the other must be a 1 and vice-versa. We start from $A = 0$, then this piece of information is fed into schemata AB and AC, and according to the assignment constraint we have $B = 1$ and $C = 1$; next, the different pieces of information from AB and AC are fed into schema BC, however what we shall have is $BC = \{11\}$, a violation to the assignment constraint in BC! Therefore, cyclic structure in this example is undesirable, and that is why it has never assumed popularity amongst the database theorists despite they can be some really natural structures that we encounter in the real world. So is a cyclic structure inadequate in terms of information representation?

Perhaps a more perplexing issue stemmed from cyclic structures is as follows. Take a look at the SL in Fig.4.2, it can be viewed to consist of three acyclic sub-universes: $AB - BC$, $AC - AB$ and $AC - BC$. As with the reasoning networks we have investigated in the previous chapter, we assign a joint probability to each schema and we understand consistency on the links by the same marginalized probabilities from the two schemata connected. Suppose the SL in Fig.4.2 is a *complete* information representation for a universe containing variables A, B and C, then the pair of sub-universes $AB - AC$, which contains all the variables in the universe, should represent the universe too; a question thus arises: as $AB - AC$ represents the universe, we can deduce a joint distribution on BC since a "global" probability can be derived on $AB - AC$, then is this deduced probability the same as that read from the sub-universe BC in the SL upon marginalization? Unfortunately, the answer is "No!". An example is provided by the following three joint probability tables, assuming that A, B and C are binary variables of value either 0 or 1:

	11	10	01	00
$P_{AB}(ab)$	0.191	0.385	0.097	0.327
$P_{BC}(bc)$	0.243	0.045	0.160	0.551
$P_{AC}(ac)$	0.318	0.258	0.085	0.339

From the above table the consistency conditions between neighbouring schemata can be verified at once; for instance, between AC and AB, we have

$$\begin{aligned} P_{AC}(a=1) &= P_{AB}(a=1) &= 0.576 \equiv P_{AB-AC}(a=1) \\ P_{AC}(a=0) &= P_{AB}(a=0) &= 0.424 \equiv P_{AB-AC}(a=0) \end{aligned}$$

where $P_{AC}(a = 1) \equiv \sum_c P_{AC}(a = 1, c)$, etc. Now suppose we form a "global" probability on $AB - AC$, and we have by (3.13) for an acyclic SL

$$P_{AB-AC}(abc) = \frac{P_{AB}(ab)P_{AC}(ac)}{P_{AB-AC}(a)} \tag{4.11}$$

It can be verified at once from (4.11) that

$$\begin{aligned} P_{AB-AC}(b=1,c=1) &= \sum_a P_{AB-AC}(a,b=1,c=1) \\ &= 0.124 \\ &\neq 0.243 \\ &= P_{BC}(b=1,c=1) \end{aligned}$$

Therefore, when invoking a cyclic configuration we have to either

1. accept *non-unique* global probabilities which dictate a number of universes encompassing all the variables; or else

2. relinquish the concept of a global probability. If so, we shall be content with a *globally consistent* universe which comprises only local probabilities connected in compliance with certain topology, with consistency between neighbouring sub-universes understood to be equal marginalized probabilities on their intersecting elements.

We opt for the second, due to its relative simplicity. In fact, our example has testified to a result proved in [11] stating that for a cyclic SL, it is in general not possible to have a global probability distribution which marginalizes into the respective local joint-probabilities. A lack of global probability distribution do seem something uncomfortable, especially with respect to most physical measurements where a global probability is always available by a frequency evaluation on a set of measurement data; however, when involved in situations dealing with probability as a measure of the degrees of belief do we have to worry the absence of a global probability? It is one of the fundamental points wanted to be addressed in this thesis that it is possible to give up the idea of a global probability; of course, that implies we shall have to accept a non-frequency interpretation of probability. The rationale behind such a choice is that what is really intrinsic to our building up of a information universe are the *individual* sets of measurements: we start from several local sub-universes to build up a global universe, not vice-versa. It is only when a global universe is taken as the starting point that a global probability is something natural. An analogy with differential geometry is immediate: if two initially parallel vectors are to be transported on a curved Riemannian surface to the same destination along different paths, it is in general not guaranteed that they will end up parallel again at the destination; however, if we start from a higher dimensional embedding space, then we can always suitably adjust the vectors along the paths so that they will end up parallel. From an operational point of view, probability is no more than a convenient short-hand for making sense of the data in each sub-universe, therefore it does not carry a fundamental significance on a global setting. This discussion justifies our abolishment of a global probability. Of course, the drawback is a really difficult interpretation to what meaning should we attach to the computed quantities, other than a corny saying of "belief" which is to the technical minded an unfathomable concept. Perhaps, it is instructive to adopt a similar attitude to the concept of mass when Einstein formulated the general theory of relativity: there is no clear reason why the inertial mass and the gravitational mass have to be identical; however, upon the adoption of this empirical fact, up to experimental errors, the result is beautiful. That is why we do not differentiate the two different concepts and be content with (actually rejoice over) a complete geometrization of the theory of gravitation today.

Natural though cyclic structures, there are indeed difficulties and constraints associated with a proper treatment of cyclic SLs. Can anything come to the rescue?

Our next consideration suggests a possible way out.

4.4 Communication – Dynamics – Complexity

Recall the essential ingredients underlying the construction of an acyclic reasoning network are:

1. Individual sub-universes of information. By a sub-universe we mean a set of experiments or observations which is represented by a joint probability on the local variables.

2. A global topological structure constituted by the sub-universes by conjoining them consistently at their intersecting elements.

3. A global information updating schedule across the network.

Our discussion with cyclic SLs so far has been focused entirely on the first two procedures, not the last one; however, while this last procedure is well defined with an acyclic topology, it is not that obvious on a cyclic structure: for instance, how should information updating be scheduled on a cyclic setting which has no definite *beginning* and *end*? Difficult it may be, we believe information updating, upon *suitable* adaptations, will serve the right key to the recruitment of cyclic reasoning networks. The conjecture is: let us regard information, expressed at present in a probability distribution, as if it is a physical attribute like the intermolecular force within a water molecule, can we pass information amongst the sub-universes until an ultimate globally consistent universe be reached? Thus, we shift the problem from whether cyclic structures are adequate (surely they are!), to how suitable *communication* to be enacted amongst different parties to promote agreement. This reduces the establishment of reasoning on a cyclic SL to the following two broad concerns:

1. Local updating formulae: they dictate how a sub-universe updates its own information upon reception of new evidence.

2. Global updating schedules: they dictate how information flow be scheduled across the network.

In fact, we shall see from examples in chapter 5 that different couplings between *local updating formulae* and *global updating schedules* will give rise to all sorts of possible fantastic behaviour. Meanwhile, we shall take a detour to look at two simple examples, one from mathematics and the other from physics, which share the same theme of *local-global* couplings, from which we draw inspiration towards what *communication* on a cyclic SL is going to give.

1. In numerical mathematics, an algorithm stands for an operation on a single or a system of equations, which is essentially *local* in the sense that subsequent iterations depend on the knowledge of the *immediate* state of the system. The algorithm itself does not take into account any *global constraint* imposed on the system; however, it is the global constraint which determines what the iterative results to be. As an example, suppose we are given functions $f(x)$ and $g(x)$ as illustrated in Fig.4.3, and we try to solve for the zeros by the Newton's method. The algorithm of the Newton's method reads:

$$x_n = x_{n-1} - \frac{f'(x_n)}{f(x)} \qquad n = 1, 2, \cdots \tag{4.12}$$

   and similarly for $g(x)$. Graphically, (4.12) tells that each x_n is determined by the intersection of the x-axis and the tangent dropped at $f(x_{n-1})$; this procedure is illustrated in Fig.4.3. While it is evident that x_n approaches the desired solution x_s for $f(x)$ due to its concaveness, convexness of $g(x)$ deprives the iteration from converging to anything. Thus, the topology of functions is the global constraint which determines whether iterations by Newton's method are going to converge to the desired solutions. Here it is evident that couplings between *local iteration algorithms* and *global constraints* determine what the outcomes under iterations will be.

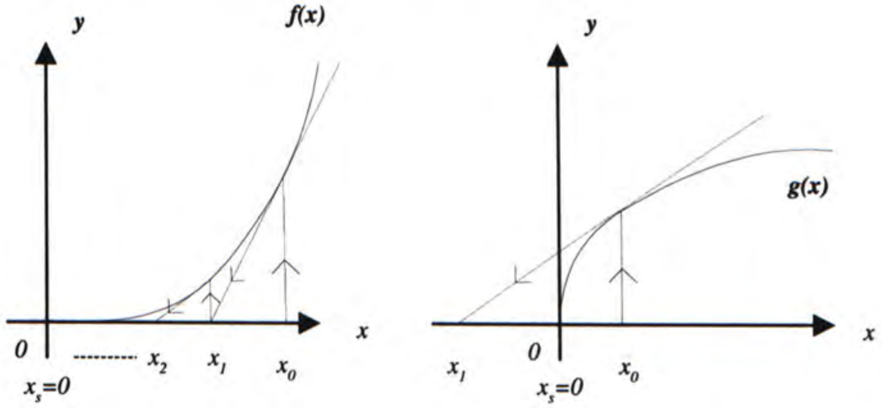
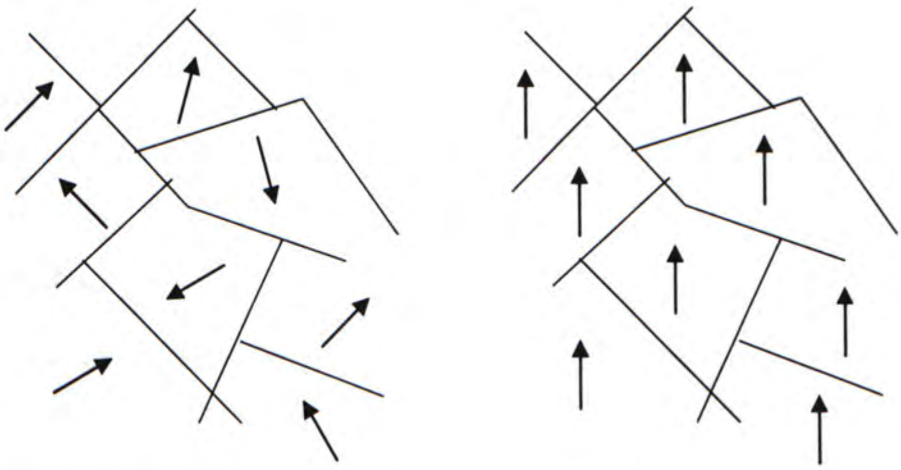


Figure 4.3: Topological dependence of Newton's method.

2. The theme of *local-global* coupling is not limited to numerical computation, but is an important theme of modern science as well. An important example is magnetization of ferro-magnetic materials. Inside those materials, each atom carries a certain magnetic moment so that each one is essentially a small magnet. When the material is subject to a temperature higher than the Curie's temperature T_c, the *local microscopic magnetic moments* are randomly oriented as suggested in Fig.4.4 so that there is no *global magnetization*; however, if the temperature is lowered to below T_c, then the *local microscopic magnetic moments* get aligned to give a *global magnetization*. The transition is very sharp and sudden like water boils at $100^0 C$. This physical example shares an interesting feature with iterative mathematics: both of them involve the coupling between *local* and *global* quantities. In this case, individual atomic magnetic moments are the local constituents and the temperature T is the global constraint governing the emergence of global magnetization. Due to such strong resemblance, a branch of inter-disciplinary science is developed in the past two decades trying to investigate the general relationship underlying such local-global correlations. This field is, appropriately termed due to the very complicated behaviour it studies, Complexity.



$T > T_c$ $T < T_c$

Figure 4.4: Temperature dependence of ferro-magnetic materials.

Other examples which fall into the category of Complexity include chaos, fractal geometry, etc. The associated convergence behaviour of chaos is complicated: it can be a single fixed point, a set of periodic fixed points or even an apparently random behaviour.

4.4.1 Communication as dynamics; dynamics to complexity

Communication in daily usage means the exchange of information amongst two or more parties. An important feature underlying is that communication is mutual in character and often lacks a distinct end, thereby a *dynamics* of information flow is entailed by communication processes. Recall that with the communication schemes with acyclic SLs an important ingredient is the global updating schedule across the network. However, unlike *acyclic* topology which entails a definite beginning and end, *cyclic* topology by definition carries no boundary; thus global scheduling on *cyclic* topology entails a form of *iterative calculations*: with the *local* and *global* roles played by the *local updating formulae* and the *global information-flow schedulings* respectively. From the mathematical and physical analogues, it is just natural to anticipate certain complex behaviour to show up in the cyclic SL upon introducing communication. In fact, it is indeed the general case exemplified by most conceivable computational experiments on the cyclic SL. *However, with carefully designed global networks coupled with well calibrated updating schemes, it is possible to render very nice iterative results* which shall form our focus in the following chapters. Further, we shall see that the cyclic reasoning networks that we attack carry a natural interpretation related to error-control applications; that will be our first testimony to a rightful adoption of the cyclic reasoning network.

4.5 Conclusion

While the theories and interpretations of *acyclic* reasoning networks are full-fledged, we come to a conjecture on the possibility to *cyclic* reasoning networks. We have come to see that cyclic information structures exist, say in the derivation of the ideal gas law. However, by incorporating a cyclic structure to an information network, we shall have to refrain ourselves from certain beloved and deep-rooted concepts. That means if we are to step into the cyclic world, we have to be ready for a quite different world-view. Towards an actual functioning of cyclic reasoning networks, we heed a theme on the coupling between *local* and *global* characteristics which will form the backbone of our forthcoming studies.

Chapter 5
Cyclic reasoning networks – error-control application

5.1 Overview

In this chapter, we shall have a first taste of the fruitful results embodied in the class of cyclic reasoning networks. We shall study a series of probabilistic reasoning networks, which are subsequently shown to carry very nice algebraic properties that enable a natural error-control interpretation. The inspiration is that in recent years people have noticed that BNs served an efficient tool in turbo-code decoding problems; however, unlike classical constructions which centered on tree structures carrying no loop, the BNs conduced from turbo-codes carry loops. The implication from the presence of loops is immense: there is information feedback; so information updating for turbo-code structures fall exactly into the realm of our considerations. Our schedule in this chapter is: upon introducing the SLs for some code structures (not refined to turbo-code type), we develop several dynamical updating schemes, we shall see the couplings between SL structures and updating schemes manifest in a non-trivial manner.

5.2 Communication schemes on cyclic reasoning networks directed to error-control applications

Similar to the acyclic counterparts, the communication schemes on cyclic SLs comprise two portions:

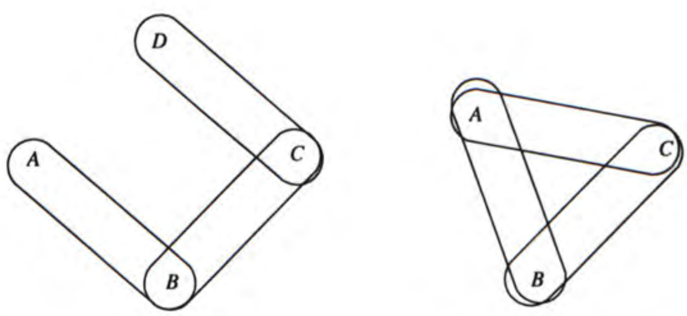
1. Local updating formulae at each sub-universe upon reception of new evidence.

2. Global updating schedule across the network.

We shall develop both steps in parallel with that of the acyclic counterparts as follows.

5.2.1 Part I – Local updating formulae

In Fig.5.1 is shown the cluster graphs of two different global configurations, one *acyclic* and the other *cyclic*. Now, suppose we sit inside the local sub-universe BC, then what shall we see in the two different configurations? Recall that a sub-universe has information transfer with other sub-universes only via the common elements. In both situations shown, BC is connected to two neighbouring sub-universes through elements B and C; so other than knowledge on B and C, one sitting in BC is absolutely ignorant of his outside world. Resultedly, he would not be able to tell whether other sub-universes are connected or not – such qualities can only be told from a global perspective! Thus the local updating formulae, which are essentially *local* operations, should not be affected by the global topology. It follows that it would be a good idea to deploy what we have in the *acyclic* counterparts, that means we should employ *similar* Bayesian updating formulae – the Absorption map in (3.12)[1].



(a) An acyclic universe (b) A cyclic universe

Figure 5.1: Local perspectives of sub-universes in different global settings.

Although we take the same form of local updating formulae as the *acyclic* counterparts, there is an additional complication now due intrinsically to the *cyclic*

[1]This idea is exactly the same in spirit with Einstein's equivalence principle in formulating the general theory of relativity: we cannot discern a gravitational effect from an accelerating effect inside a local frame – so suppose we sit inside a lift and see an apple falling, we cannot tell the difference between whether there is a gravitational attraction acting on the apple, or that the lift is being accelerated upwards.

nature: simultaneous updating. While simultaneous updating does not carry much significance in the acyclic case (in fact, they occur scarcely); in contrast, on a cyclic network, simultaneous updating under iterative calculations is ubiquitous. This leads us to the following important concern regarding our intended application to error-control problems: in iterative decoding, we want to maintain the structure of the code; however, simultaneous updating erodes structure (as to be shown below), further exacerbation comes along with the iterative character of calculations. So we need to find a way to circumvent this difficulty.

Upon simultaneous updating, *independent* pieces of new evidence are absorbed into a local sub-universe. For instance, let us consider an influence diagram as in Fig.5.2 (a) and its corresponding cluster graph in Fig.5.2 (b). If E_A and E_B receives new evidence, clusters E_AA and E_BB will update cluster ABC simultaneously. Now suppose our problem is such that the conditional probability $P(c|a,b)$ is a defining feature of the system, a naive application of the updating formula

$$P^*(abc) = P(abc)\frac{P^*(a)}{P(a)}\frac{P^*(b)}{P(b)} \tag{5.1}$$

may not be desirable since by (5.1) we shall have

$$\begin{aligned} P^*(c|ab) &\equiv \frac{P^*(abc)}{P^*(ab)} \\ &= \frac{P^*(abc)}{\sum_c P^*(a,b,c)} \\ &\neq \frac{P(abc)}{P(ab)} \\ &= P(c|ab) \end{aligned} \tag{5.2}$$

This problem is exacerbated by the iterative nature of the communication scheme pertaining to a *cyclic* network, which gradually erodes away the dependence of C on the combinations of AB. To surmount this difficulty, we group A, B together into AB, and modify the updating formula (5.1) to

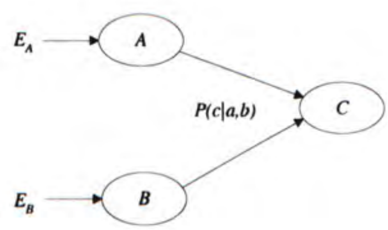
$$P^*(abc) = P(abc)\frac{P^*(a)P^*(b)}{P(ab)} \tag{5.3}$$

with which we have

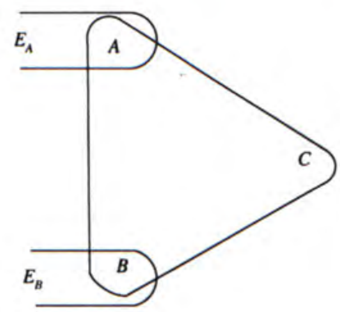
$$\begin{aligned} P^*(c|ab) &\equiv \frac{P^*(abc)}{P^*(ab)} \\ &= \frac{P^*(abc)}{P^*(a)P^*(b)} \end{aligned}$$

$$= \frac{P(abc)}{P(ab)}$$
$$= P(c|ab) \quad (5.4)$$

Thus by (5.4) the defining nature of the system is preserved. Notice that this problem is pertinent to the particular problem at hand which dictates what of the original measurement $P(abc)$ would have to be maintained: may it be a conditional dependence like $P(c|ab)$ or anything else. The updating procedure remains the same in spirit with the simple absorption map; what makes a difference is an emphasis on the roles played by some possible combinations of variables.



(a) Influence diagram



(b) Cluster graph

Figure 5.2: Simultaneous updating on an information network.

5.2.2 Part II – Global updating schedules across the network

Another problem is the global scheduling of updating across the network. Unlike scheduling on an *acyclic* network which is well-defined, the presence of loops on a *cyclic* network renders new evidence be assessed everywhere, which as a whole invalidates a locally based (i.e. based on individual sub-universes) scheduling scheme on a *cyclic* setting. So, instead of a locally based scheme, we advocate a semi-global scheme by incorporating several sub-universes carrying similar characteristics into one cluster of sub-universes; then updating should be scheduled amongst such clusters. The rationale underlying is again drawn by analogy with many physical problems. As an example, when two magnets interact, we will not consider their interactions in terms of the individual atomic magnetic moments; rather, we know the global magnetizations of the individual magnets, and consider the interaction between in terms of the global magnetizations. In our following studies on error-control problems, we shall be led by the above philosophy to construct different global updating schemes, and see how interesting outcomes are conduced.

5.3 Probabilistic reasoning based error-control schemes

In this section we talk about the rationale underlying probabilistic reasoning based error-control approach. Our examples to be constructed aim at showing the logic underlying the problem, and the feasibility of attacking the problem in the way prescribed. Our focus is on individual decoding, so we shall not pay attention to any efficiency concern and performance issues, etc., although they are no doubt important factors in justifying the efficacy of an error-control structure.

5.3.1 Local sub-universes and global universe underlying the error-control structure

An error-control structure is composed of two parts:

1. Some independent information bits, $\{X_i, i = 1, 2, \cdots, N\}$.

2. Some error-control bits, $\{C_j, j = 1, 2, \cdots, K\}$.

Each of the error-control bits is constructed from a subset of the information bits. Thus a code, which consists of both information and error-control bits, forms a structure in its own right; in contrast with a plain collection of information bits which is nothing but individual entities. The idea of error-control is that if some bits, information or error-control ones, happen to be transmitted with errors, then by resorting to the underlying correlation within the code structure, we would be able to infer which bits are tainted and how to correct them. The classical scheme towards such an approach is algebraic in character, in which we work out a set of algebraic rules for error correction; however, the difficulty underlying such an approach is that the algebraic rules can be very intricate that escape the untrained minds. To circumvent this difficulty we notice the following: no matter how the algebraic rules of encoding work, what they do in essence is to identify a particular C_j with some X_i, by which certain topological structures of the code are entailed on top of the precise algebraic assignments; a natural question thus follows is whether decoding may proceed entirely on the basis of such topological structures, which triggers the framework of probabilistic reasoning based error-control approach to be investigated below. A dynamical analog is of order here: an error is very much like a perturbation to the structure; if the perturbation is not too large as to destroy the structure, the disturbance caused by the perturbation abates on subsequent iterations and the original structure can be restored to its untainted state. This line of thought can be found in studies like [4] and [8], etc.

5.4 Error-control structure I

Our first code is constructed from three binary information bits: X_1, X_2 and X_3, and there are three error-control bits constructed according to the following rules:

$$\begin{aligned} C_1 &= (X_1 + X_2) \bmod 2 \\ C_2 &= (X_2 + X_3) \bmod 2 \\ C_3 &= (X_1 + X_3) \bmod 2 \end{aligned} \tag{5.5}$$

The code bits $\{X_i, C_i | i = 1, 2, 3\}$ are subsequently transmitted across a noisy channel, and the received bits are correspondingly $\{Y_i, D_i | i = 1, 2, 3\}$. This code structure can be cast in the form of a reasoning network, with which the corresponding BN is shown in Fig.5.3. The BN can be roughly divided into two big categories: the inner structure layer involves only X_i and C_i, and the outer evidence layers involve observations Y_i and D_i. The probability assignment is according to the following conventions:

1. Since the construction of the error-control bits are given by (5.5) which is a deterministic assignment, we may put $P(c_1 = 0|x_1 = 0, x_2 = 0) = 1$ and $P(c_1 = 1|x_1 = 0, x_2 = 0) = 0$, etc. However, in practical computations such deterministic assignment may not be convenient; in such cases we assign instead a probability of assignment $0 \leq ass \leq 1$ such that $P(c_1 = 0|x_1 = 0, x_2 = 0) = ass$ and $P(c_1 = 1|x_1 = 0, x_2 = 0) = 1 - ass$. [2]

2. Conditional probabilities between Y_i and X_i, D_i and C_i represent the chance that a code bit gets tainted on transmission across the noisy channel. We shall

[2]We may imagine a machine that not simply assigns either a value 0 or 1 to a signal on comparison with a certain threshold, but assigns a *degree of belief* to the value proportional to the actual value of the signal received. For example, if the full-reading of a signal is $1V$, and we take an observed signal larger than $0.5V$ be a value 1, and a value 0 otherwise, then we may assign for a received signal of $0.7V$ a value 1 with probability 0.7, whereas a received signal of $0.4V$ is assigned a value 1 with probability 0.4 (or a value 0 with probability 0.6).

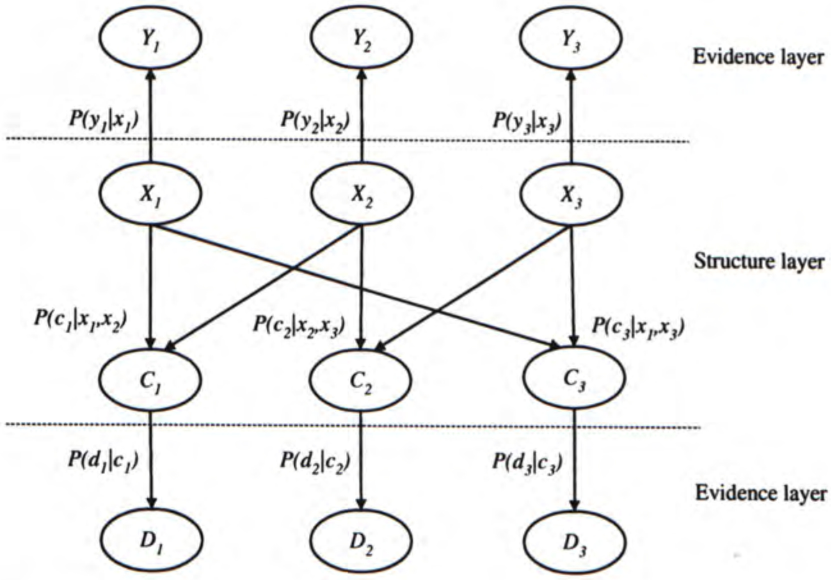


Figure 5.3: Error-control structure I– Bayesian network perspective.

assume the probability for a bit to suffer a transmission error be $0 \leq err \leq 1$. So, with a pretty good quality channel, we may assume $err \sim 0.1$, hence $P(y_1 = 1|x_1 = 0) = 0.9$ and $P(y_1 = 1|x_1 = 0) = 0.1$, etc.

In the language of classical maximum-likelihood decoding, the code constructed by (5.5) is a one-error correcting code as the minimum Hamming distance of the code is three as exemplified by the two code-words 000000 and 111000. Our decoding is going to be a maximum a posteriori one, so how is ours going to be different?

Next, we are going to transplant this code into a probabilistic reasoning based decoding framework. The (conceptually simpler) BN corresponding to the code is shown in Fig.5.3; however, recall that it is the (operationally fundamental) SL that we favour, so we cast Fig.5.3 into its corresponding SL. From Fig.5.3, we read off the schemata – the independent sub-universes – as $\{X_iY_i, i = 1, 2, 3\}$, $X_1X_2C_1$, $X_2X_3C_2$, and $X_1X_3C_3$; these schemata form a cyclic set as they are not Graham-reducible to the null set. A word on the schemata is in order, as they actually give rise to the conditional probabilities of the BN, there is a clear meaning on the term "independent sub-universes" ascribed to them:

1. Each X_iY_i or C_iD_i stands for an experiment to measure the reliability of the transmission channel. Suppose we make up a set of X_i, we can then observe the transmitted signals Y_i, these observations allow a frequency interpretation of the conditional probabilities $P(y_i|x_i)$.

2. The remaining schemata represent the defining structure of the code, namely the assignment rules in (5.5). Though the assignment rules are deterministic in character, they can nonetheless be rendered a probability interpretation as described before. Anyway, no matter what the rationales behind those probabilities are, it is just important that they are defined within *one set of rules*, or *one set of experiments* which is tantamount to *an independent sub-universe.* Through an articulation of the local sub-universes with respect to the topology of the BN in Fig.5.3, the global universe is obtained.

The SL corresponding to Fig.5.3 is shown in Fig.5.4 with the sub-universes shown as ovals and separators as squares. Notice that this SL is drawn in a slightly different way to emphasize the *functional* roles played inside the SL structure: first, we have two *evidence layers* as indicated, one upper and one lower, they are schemata where observed evidence enters the structure through Y_i and D_i; second, we have the inner *structure layer* played by the schemata involving only X_i and C_i related in a way dictated by the defining formulae of the code structure in (5.5). The spirit underlying probabilistic reasoning based decoding is that by absorbing evidence (making observations) at Y_i and D_i, a further check against the structure of the code will recover the untainted code-words.

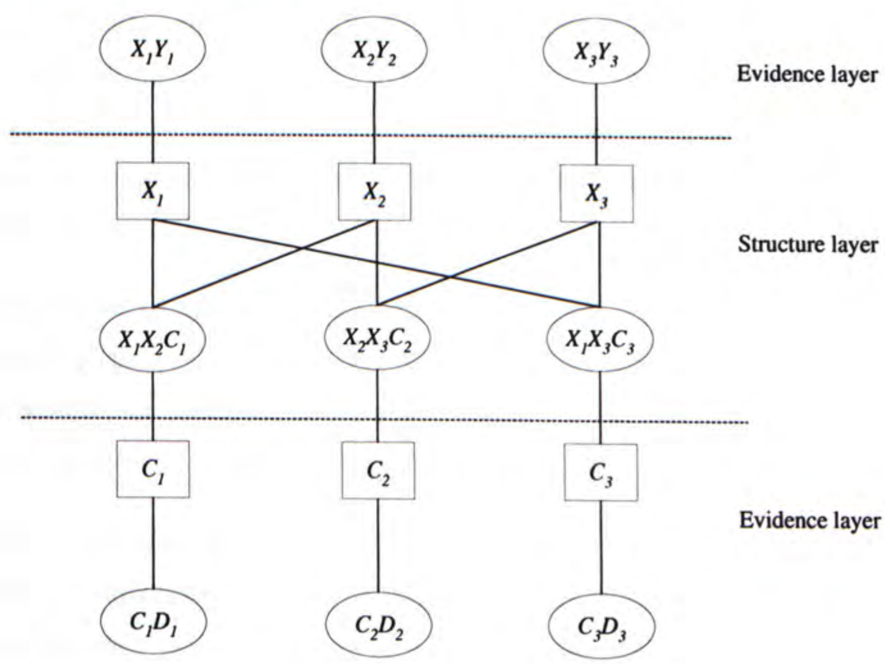


Figure 5.4: Error-control structure I– Semi-lattice perspective.

5.4.1 Decoding algorithm – Communication between local sub-universes in compliance with the global topology

Now, we come to the heart of the whole program: decoding. Decoding is, as introduced before, a series of communication amongst the local sub-universes in compliance with the global topology. We shall develop the whole program under the spirit of section 5.2. Decoding in this case broadly comprises two steps:

1. Absorption of evidence.

2. Checking the evidence against the pre-defined code structure.

As an iterative decoder, the above two steps repeat until a final converged solution is obtained; with respect to our problems, it is tantamount to the ultimate attainment of a globally consistent SL – that is when neighbouring schemata marginalize to the same probabilities on their intersecting elements. Suppose we have a set of observed evidence $\{y_i = \bar{y}_i, c_i = \bar{c}_i | i = 1, 2, 3\}$, we propose our algorithm as follows:

Decoding algorithm for error-control structure I:

1. *Initialization of the code structure:*
   We let $P(x_i) = 0.5$ for both possible outcomes of the root events $x_i, i = 1, 2, 3$; that means we have no prior evidence of what the code is likely to be. Then we can initialize the joint probabilities of individual schemata in the cyclic SL successively according to the original BN; this results in a globally consistent SL to begin with.

2. *Assignment of updated evidential probabilities:*
   Let $P^*_{X_iY_i}(y_i = \bar{y}_i) = P^*_{C_iD_i}(d_i = \bar{d}_i) = k$ where $0.5 \leq k \leq 1$ stands for a probability of belief of an observed event [3]. For the other possible value of y_i, we assign $P^*_{X_iY_i}(y_i \neq \bar{y}_i) = P^*_{C_iD_i}(d_i \neq \bar{d}_i) = 1 - k$.

3. *Absorption of evidence I:*
   The sub-universes $\{X_iY_i, i = 1, 2, 3\}$ in the evidence layers are updated as

[3]Notice that this k needs not always be unity; there are conditions which enfeeble an observation. For instance, although it is true that a machine returns either a 0 or 1 for an observed signal, it actually does so by *coarse-graining* the observed signal: there is a threshold value in hardware implementation above which a signal is assigned a value 1 or a value 0 otherwise. Such approximations represent an additional uncertainty.

follows:

$$P^*_{X_iY_i}(x_iy_i) \leftarrow P_{X_iY_i}(x_iy_i)\frac{P^*_{X_iY_i}(y_i)}{P_{X_iY_i}(y_i)} \tag{5.6}$$

and similarly for $\{C_iD_i, i = 1, 2, 3\}$

Set do-loop counter: *count* $= 0$

Do while [the SL is not globally consistent]

4. *Absorption of evidence II:*

   Next, evidence is absorbed from the evidence layers into the structure layer. Each schema within the structure layer absorbs evidence from schemata in the evidence layers through the intersecting elements. So, for instance, the schema $X_1X_2C_1$ absorbs evidence from the set of schemata $\{X_1Y_1, X_2Y_2, C_1D_1\}$; besides, noting that within the schema $X_1X_2C_1$, the group X_1X_2 determines C_1, thus taking into account of (5.3), the updating formula to $X_1X_2C_1$ reads[4]:

   if count $= 0$:

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2}P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right)\left(\frac{P^*_{C_1D_1}(c_1)}{P_{X_1X_2C_1}(c_1)}\right) \tag{5.7}$$

   elseif count > 0 :

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow P_{X_1X_2C_1}(x_1x_2c_1)\frac{\prod_{i=1,2}P^*_{X_iY_i}(x_i)}{P_{X_1X_2X_4}(x_1x_2)} \tag{5.8}$$

   endif

   Similar calculations hold for $P^*_{X_2X_3C_2}(x_2x_3c_2)$ and $P^*_{X_1X_3C_3}(x_1x_3c_3)$.

5. *Checking evidence against the code structure:*

   Next, a schema within the structure layer is updated by other schemata in the structure layer through their intersecting elements. So for instance, the schema $X_1X_2C_1$ is connected to $X_1X_3C_3$ through X_1 and $X_2X_3C_2$ through

[4] A note on notations: Along the course of information updating our formulae involve not only information absorptions, but also a renormalization within individual sub-universes; however, as a renormalization (within each sub-universe) amounts only to a multiplication factor, we can thus take the convenience to denote all renormalization constants by an "α" and just remind ourselves that each α has to be evaluated locally.

X_2, the corresponding updating thus proceeds through X_1 and X_2; besides, noting that within $X_1X_2C_1$, the group X_1X_2 determines C_1; taking into account of (5.3), updating to $X_1X_2C_1$ reads

$$P^{**}_{X_1X_2C_1}(x_1x_2c_1) \leftarrow P^{*}_{X_1X_2C_1}(x_1x_2c_1)\frac{P^{*}_{X_1X_3C_3}(x_1)P^{*}_{X_2X_3C_2}(x_2)}{P^{*}_{X_1X_2C_1}(x_1x_2)} \tag{5.9}$$

Similar calculations hold for $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ and $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$.

6. *Output of evidence:*

   After a check against the code structure, the updated values of X_i are fed back to the evidence layers through again the intersecting elements. Thus, for example, we have

$$P^{**}_{X_1Y_1}(x_1y_1) \leftarrow \alpha P^{*}_{X_1Y_1}(x_1y_1)\frac{P^{**}_{X_1X_2C_1}(x_1)P^{**}_{X_1X_3C_3}(x_1)}{p^{*}_{X_1Y_1}(x_1)^2} \tag{5.10}$$

   and similary for $P^{**}_{X_iY_i}(x_iy_i), \quad i = 2,3$; besides,

$$P^{**}_{C_1D_1}(c_1d_1) \leftarrow P^{*}_{C_1D_1}(c_1d_1)\frac{P^{**}_{X_1X_2C_1}(c_1)}{P^{*}_{C_1D_1}(c_1)} \tag{5.11}$$

   and similarly for $P^{**}_{C_iD_i}(c_id_i), \ i = 2,3$.

7. *Renaming items:*

   We rename items for further iterations:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^{*}_{X_iY_i}(x_iy_i)$ for $i = 1,2,3$.

   (b) $P^{**}_{C_iD_i}(c_id_i) \to P^{*}_{C_iD_i}(c_id_i)$ for $i = 1,2,3$.

   (c) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_2C_1}(x_1x_2c_1)$, and similary for $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ and $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$.

   $count = count + 1$

   enddo

8. Upon the attainment of global consistency, beief for individual variables can be extracted from the individual sub-universes as follows:

$$P^t(x_i) = \sum_{y_i} P^{**}_{X_iY_i}(x_iy_i)$$
$$P^t(c_i) = \sum_{d_i} P^{**}_{C_iD_i}(c_id_i) \tag{5.12}$$

9. To read off the decoded code-word, we pick

$$\begin{aligned} X_i &= \arg\left\{\max_{\mathrm{x_i}}[P^t(x_i)]\right\} \\ &\equiv \begin{cases} x_{i_0} & \text{if } P^t(x_{i_0}) > P^t(x_{i_1}) \\ x_{i_1} & \text{if } P^t(x_{i_1}) > P^t(x_{i_0}) \end{cases} \end{aligned} \tag{5.13}$$

for $i = 1, 2, 3$ and we let x_{i_0} and x_{i_1} to stand for the instances of x_i. Similar assignments hold for C_i. If it happens that $P^t(x_{i_0}) = P^t(x_{i_1})$, we pick one value arbitrarily.

Remarks:

1. Steps 1 to 3 describe the absorption of evidence from external observations; notice that by step 1 the structure of the code dictated by the set of conditional probabilities is engraved into the corresponding SL. Steps 4 to 7 are the iterative checks of the evidence against the code structure; it is by these iterative checks that the untainted code-word is to be restored.

2. In step 4, the updating formulae differ for *count* $= 0$ and *count* > 0 due to the different roles the schemata X_1Y_1 and X_2Y_2 play from that of C_1D_1. Of course, they share the similarity that they are all where initial observations are made, that is the reason why the updating formula incorporate all three schemata at *count* $= 0$. However, while X_1Y_1 and X_2Y_2 are both connected to two schemata in the structure layer, C_1D_1 is connected to one only; this structural difference suggests that X_1Y_1 and X_2Y_2 share an additional purpose of *mediating* evidences between the schemata in the structure layer to which they are connected. Therefore for *count* > 0 where the iteration aims at striving for a consistency of information, the updating formula involves only X_1Y_1 and X_2Y_2, but not C_1D_1.

3. In practical computations it can be the case that the beliefs of all the possible decoded values are very close, then the decoded bit is admittedly unreliable. If many bits of a decoded code-word behave that way, it would be better to abandon the code-word completely in practical applications.

5.4.2 Decoding rationales

A problem that is evident from the last remark above is that we need to have an objective rationale for assessing the confidence level of the decoded results. We devise a measure as follows: suppose a random variable X has possible values x_0 and x_1, we take the ratio

$$\kappa_X \equiv \begin{cases} P(x_0)/P(x_1) & \text{if } P(x_0)/P(x_1) \geq 1 \\ P(x_1)/P(x_0) & \text{if } P(x_1)/P(x_0) > 1 \end{cases} \tag{5.14}$$

as a quantitative measure of the contrast between possible outcomes. So $\kappa \geq 1$; the bigger κ is, the more confident is the outcome. Of course, this confidence is solely a confidence measure on the possible outcomes regarding a probability assignment, it has nothing to do with the confidence related to the decoded result being the true and untainted bit.

5.4.3 Computational results

In this subsection, we are going to plug in some numbers to see if our decoding structure works. We shall test the code on assuming the following parameters:

1. probability of a transmission error, $err = 0.1$;
2. the assignment probability, $ass = 0.999$;
3. initial confidence of observation, $\kappa = 9$.

Besides, we shall agree on the following convention on division by 0: in a product like (5.6), if corresponding to a particular y_0 of the random variable Y the denominator $P(y_0) = 0$, the corresponding numerator $P^*(y_0)$ must also be zero and the quotient $0/0 \equiv 1$; otherwise, the outcome is undefined. A set of experimental decoding results is given in Table 5.1, which testifies that a single-error tainted code can be correctly decoded, with typical confidence $\kappa \approx 2 \sim 9$ (i.e. a degree of belief in the range 0.65 to 0.9). So, how does decoding with two-error codes proceed? This is what we study next.

Decoding on two-error codes is a bit more intricate. Suppose κ varies, a question then arises: if there are actually two errors incurred along the transmission, but it is known that they are both close to the threshold (so the confidence for such observations $\kappa \approx 1$), is it likely that we may still be able to decode it? The answer

untainted codes	1-error tainted codes	recovered codes
001011	$\underline{1}01011$	$\underline{0}01011$
100101	$\underline{0}00101$	$\underline{1}00101$
000000	$00\underline{1}000$	$00\underline{0}000$
⋮	⋮	⋮

Table 5.1: Single-error decoding of error-control structure I.

untainted bits	transmitted bits	degree of belief	decoded bits	degree of belief
0	1	0.60	0	0.58
0	0	0.90	0	0.90
1	1	0.90	1	0.92
0	0	0.90	0	0.57
1	1	0.90	1	0.83
1	0	0.60	1	0.57

Table 5.2: Correct decoding of a two-error tainted code.

is: probably! Tables 5.2 and 5.3 show respectively the decoded results of a two-error tainted code with different degrees of belief on the observations: it suggests that if all the remaining bits other than the tainted bits are received confidently, then we may still be able to retrieve the pristine state of the code. So our probability based decoder takes into account the **weight** of evidence.

Further properties:

To prepare the background for the following chapter, let us experiment one step further: we shall iterate on an initially *inconsistent* structure (w.r.t. the allowed code structure), and see what the iterated outcome will be. We stripe the structure of an error-control interpretation, and regard it simply as a dynamical structure. The details are reported in Table 5.4 and Table 5.5; it is found in both cases that a *consistent* structure is recovered after iterations, it is especially astounding in Table 5.5 that all the initial bits are transformed to attain a consistent structure, albeit

untainted bits	transmitted bits	degree of belief	decoded bits	degree of belief
0	1	0.60	1	0.53
0	0	0.80	0	0.77
1	1	0.70	1	0.77
0	0	0.75	1	0.52
1	1	0.80	1	0.65
1	0	0.70	0	0.52

Table 5.3: Incorrect decoding of a two-error tainted code.

original bits	degree of belief	iterated bits	degree of belief
0	0.74	0	0.64
1	0.72	1	0.63
0	0.57	1	0.60
0	0.62	1	0.53
0	0.70	0	0.53
1	0.83	1	0.53

Table 5.4: Iteration on an inconsistent structure (1).

most of the bits are accompanied by a quite low confidence $\kappa \approx 1$. In fact, this structure preserving property will be found generally true with all our proposed structures, which we shall explain in the coming chapter.

5.5 Error-control structure II

5.5.1 Structure of the code and the corresponding decoding algorithm

The second error-control structure we are going to study is an elaboration of structure I: we add one more set of error-control bits to the code devised previously. So, in error-control structure II, we have

original bits	degree of belief	iterated bits	degree of belief
0	0.90	1	0.56
1	0.90	0	0.56
0	0.90	1	0.56
0	0.90	1	0.51
0	0.90	1	0.51
1	0.90	0	0.51

Table 5.5: Iteration on an inconsistent structure (2).

1. Some independent information bits: $S_0 = \{X_i, i = 1, 2, 3\}$.
2. A first set of error-control bits: $S_1 = \{C_j, j = 1, 2, 3\}$.
3. Another set of error-control bits: $S_2 = \{C'_j, j = 1, 2, 3\}$.

where S_1 is defined in (5.5), and S_2 is defined as follows,

$$\begin{aligned} C'_1 &= (X_1 + X_2 + 1) \bmod 2 \\ C'_2 &= (X_2 + X_3 + 1) \bmod 2 \\ C'_3 &= (X_1 + X_3 + 1) \bmod 2 \end{aligned} \tag{5.15}$$

Thus S_1 and S_2 carry exactly opposite bits: if $C_1 = 1$, we would have $C'_1 = 0$, etc. This is again a one-error correcting code in maximum-likelihood decoding since the Hamming distance is 3 as exhibited by the code-words 000000111 and 111000111.

To recapitulate our line of concern regarding error-control applications, recall the ingredients underlying our error-control structures:

1. local joint probabilities at individual sub-universes;
2. a global topology according to which the sub-universes join together to form a universe;
3. dynamical updating at individual sub-universes and a global scheduling on the flow of information updating across the universe.

A comparison between the present problem and the last one shows a difference in the global topology – a difference in the configuration of the global universe, which

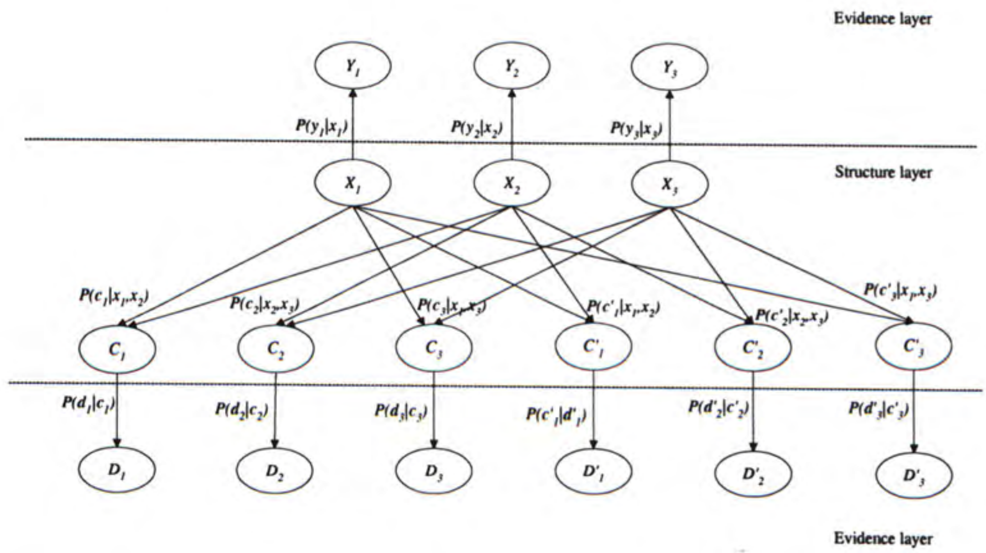


Figure 5.5: Error-control structure II – Bayesian network perspective.

demands a different scheduling on the global information flow. We shall deploy a naive direct adaptation of the updating scheme of error-control structure I for the present problem, and observe the resulting error correcting capability. However, recall that there is no obligation between global topology and global information flow scheduling, we are completely *free* to pick any flow schedule; a question thus follows to all active minds: what would show up if different global flow schedules are deployed? This is a question which triggers our subsequent dig into error-control structures III and IV. To pinpoint the problem: we want to study the correlations, the differences arise out of different couplings of global topology and global updating dynamics; bear this in mind, we set forth in the present section a study on error-control structure II.

As with error-control structure I, we cast this code in the perspectives of a BN and a SL depicted in Fig.5.5 and Fig.5.6 respectively. Again, we assume the set of observations be $\{y_i = \bar{y}_i, d_i = \bar{d}_i, d'_i = \bar{d}'_i | i = 1, 2, 3\}$, and in parallel with the decoding algorithm with error-control structure I, we propose the following decoding scheme for error-control structure II:

Decoding algorithm for error-control structure II:

1. *Initialization of the code structure:*
   We let $P(x_i) = 0.5$ for both possible outcomes of the root events in S_0; that means we have no prior evidence of what the code is likely to be. Then

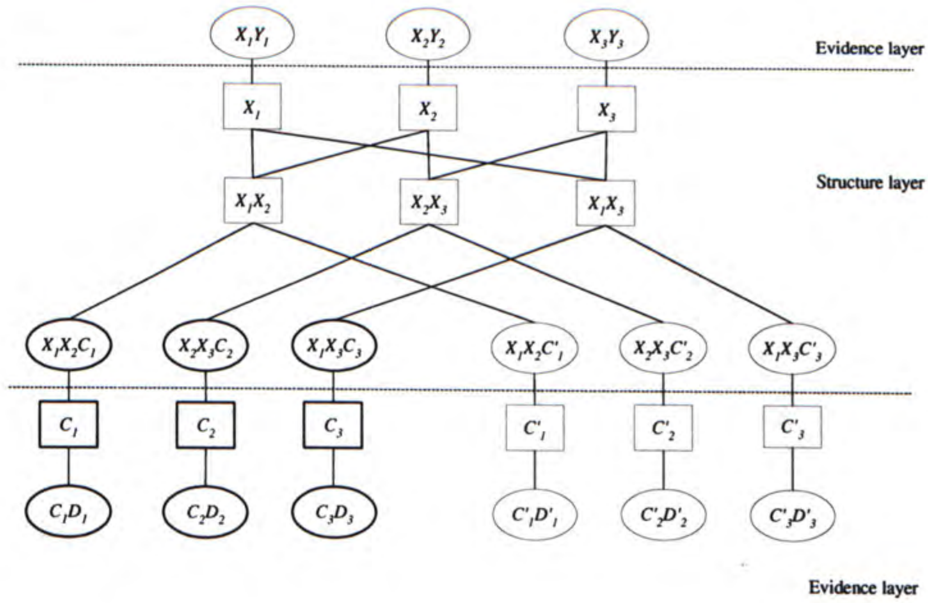


Figure 5.6: Error-control structure II – Semi-lattice perspective.

we initialize the joint probabilities of individual schemata in the cyclic SL successively according to the original BN; this results in a globally consistent SL to begin with.

2. *Assignment of updated evidential probabilities:*

   As with error-control structure I, we put $P^*_{X_iY_i}(y_i = \bar{y}_i) = k$ where $0.5 \leq k \leq 1$ stands for a probability of belief of an observed event. For the other possible value of y_i, we assign $P^*_{X_iY_i}(y_i \neq \bar{y}_i) = 1 - k$. Similar assignments hold for schemata C_iD_i and $C_i^{'}D_i^{'}$.

3. *Absorption of evidence I:*

   The individual sub-universes in the evidence layers $\{X_iY_i, i = 1, 2, 3\}$ are updated as follows:

$$P^*_{X_iY_i}(x_iy_i) \leftarrow P_{X_iY_i}(x_iy_i)\frac{P^*_{X_iY_i}(y_i)}{P_{X_iY_i}(y_i)} \tag{5.16}$$

   and similarly for C_iD_i and $C_i^{'}D_i^{'}$.

   Set do-loop counter: $count = 0$

   Do while [the SL is not globally consistent]

4. *Absorption of evidence II:*

   Next, evidence is absorbed from the evidence layers into the structure layer.

Each schema within the structure layer absorbs evidence from schemata in the evidence layers through the intersecting elements. So, for instance, the schema $X_1X_2C_1$ absorbs evidence from the set of schemata $\{X_1Y_1, X_2Y_2, C_1D_1\}$; besides, noting that within the schema $X_1X_2C_1$, the group X_1X_2 determines C_1, thus taking into account of (5.3), the updating formulae to $X_1X_2C_1$ read:

if count $= 0$:

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right)\left(\frac{P^*_{C_1D_1}(c_1)}{P_{X_1X_2C_1}(c_1)}\right) \tag{5.17}$$

elseif count > 0 :

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)} \tag{5.18}$$

endif

and similarly for other schemata within the structure layer.

5. *Checking evidence against the code structure:*
Next, each schema in the structure layer is updated by other schemata belonging to the structure layer through their intersecting elements. However, unlike error-control structure I, we have to take care of the fact that there are two sets of intersection within the structure layer:

   (a) that within S_1 and S_2 individually, and

   (b) that between S_1 and S_2.

   We take the updating for schema $X_1X_2C_1$ as an example. Within S_1, the schema $X_1X_2C_1$ is connected to $X_1X_3C_3$ through X_1 and $X_2X_3C_2$ through X_2; besides, $X_1X_2C_1$ is connected to S_2: through X_1X_2 to $X_1X_2C_1'$, through X_1 to $X_1X_3C_3'$ and through X_2 to $X_2X_3C_2'$. Analogous to error-control structure I, updatings to $X_1X_2C_1$ are accomplished through the intersecting elements. Noting further that within the schema $X_1X_2C_1$, the group X_1X_2 determines C_1, taking into account of (5.3), updating to $X_1X_2C_1$ reads

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)T_{S_1}T_{S_2} \tag{5.19}$$

where T_{S_1} and T_{S_2} stand for updatings due to S_1 and S_2 respectively; they are given by

$$\begin{aligned} T_{S_1} &= \frac{P^*_{X_1X_3C_3}(x_1)P^*_{X_2X_3C_2}(x_2)}{P_{X_1X_2C_1}(x_1x_2)} \\ T_{S_2} &= \frac{P^*_{X_1X_2C_1'}(x_1x_2)p^*_{X_1X_3C_3'}(x_1)P^*_{X_2X_3C_2'}(x_2)}{p_{X_1X_2C_1}(x_1x_2)^2} \end{aligned}$$

Similar constructions are immediate to other schemata in the structure layer.

6. *Output of evidence:*

   After a check against the code structure, the updated values of X_i, C_i and C_i' are fed back to the evidence layers through the intersecting elements. Thus, for example, we have

$$P^{**}_{X_1Y_1}(x_1y_1) \leftarrow \alpha P^*_{X_1Y_1}(x_1y_1)\frac{P^*_{X_1X_2C_1}(x_1)P^*_{X_1X_3C_3}(x_1)P^*_{X_1X_2C_1'}(x_1)P^*_{X_1X_3C_3'}(x_1)}{p^*_{X_1Y_1}(x_1)^4} \tag{5.20}$$

   and similarly for $P^{**}_{X_iY_i}(x_iy_i)$ for $i = 2, 3$; besides,

$$P^{**}_{C_1D_1}(c_1d_1) \leftarrow P^*_{C_1D_1}(c_1d_1)\frac{P^{**}_{X_1X_2C_1}(c_1)}{P^*_{C_1D_1}(c_1)} \tag{5.21}$$

   and similarly for other schemata in S_1 and S_2.

7. *Renaming items:*

   We rename items for further iterations:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^*_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

   (b) $P^{**}_{C_iD_i}(c_id_i) \to P^*_{C_iD_i}(c_id_i)$ for $i = 1, 2, 3$.

   (c) $P^{**}_{C_i'D_i'}(c_i'd_i') \to P^*_{C_i'D_i'}(c_i'd_i')$ for $i = 1, 2, 3$.

   (d) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_2C_1}(x_1x_2c_1)$, and similary for other schemata inside the structure layer.

   count = *count* + 1

   enddo

8. On attaining global consistency, belief of individual variables can be extracted from the individual sub-universes as follows:

$$P(x_i) = \sum_{y_i} P^*_{X_iY_i}(x_i, y_i)$$

$$P(c_i) = \sum_{d_i} P^*_{C_iD_i}(c_i, d_i)$$
$$P(c'_i) = \sum_{d'_i} P^*_{C'_iD'_i}(c'_i, d'_i) \tag{5.22}$$

9. To read off the decoded code word, we pick

$$\begin{aligned} X_i &= \arg\left\{\max_{x_i}[P^t(x_i)]\right\} \\ &\equiv \begin{cases} x_{i_0} & \text{if } P^t(x_{i_0}) > P^t(x_{i_1}) \\ x_{i_1} & \text{if } P^t(x_{i_1}) > P^t(x_{i_0}) \end{cases} \end{aligned} \tag{5.23}$$

for $i = 1, 2, 3$; similar assignments hold for C_i and C'_i. If it happens that $P^t(x_{i_0}) = P^t(x_{i_1})$, we pick one value arbitrarily.

Remarks:

1. Since error-control structure II is essentially an elaboration of structure I, it is thus expected that both structures share the characteristics discussed with structure I.

2. Error-control structure II and the forthcoming structures III and IV mimic a *turbo-* type code which deploys two or more encoders; however, we shall take into account no *interleaver* since our interest is directed to individual decoding rather than the overall performances.

5.5.2 Computational results

We shall carry out a set of experiments in parallel with that of error-control structure I. Again, we start from a simple survey on the error-correcting capability of a single-error tainted code and take parameters as follows,

1. probability for a transmission error, $err = 0.1$;

2. the assignment probability, $ass = 0.999$;

3. initial confidence of observation , $\kappa = 9$.

Some results are reported in Table 5.6 which testify to our claim, with the confidence of decoding $\kappa \approx 4$ up to 99. Further, as with error-control structure I, our

untainted codes	1-error tainted codes	recovered codes
001011100	101011100	001011100
100101010	000101010	100101010
000000111	001000111	000000111
000000111	000100111	000000111
⋮	⋮	⋮

Table 5.6: Single-error decoding of error-control structure II.

untainted bits	transmitted bits	degree of belief	decoded bits	degree of belief
0	1	0.60	0	0.55
0	0	0.90	0	0.99
1	1	0.90	1	0.96
0	1	0.60	0	0.55
1	1	0.90	1	0.95
1	1	0.90	1	0.55
1	0	0.60	1	0.55
0	0	0.90	0	0.95
0	0	0.90	0	0.55

Table 5.7: Correct decoding with a three-error tainted code.

present structure decodes with regard to the **weight** – the degree of belief – of the code bits, so occasionally we are capable of decoding a code with more errors than the restriction set by the minimum Hammming distance; an example is shown in Table 5.7 in which a three-error tainted code may still be correctly decoded if the tainted bits do not err too much; of course, as the transmitted code errs by more than a Hamming distance, it is anticipated that decoding will not be stable, actually if there is one more error incurred along the transmission, as is suggested in Table 5.8, the decoded code turns out to be totally another one dictated by the code structure.

untainted bits	transmitted bits	degree of belief	decoded bits	degree of belief
0	1	0.60	1	0.96
0	0	0.90	0	0.83
1	0	0.60	1	0.81
0	1	0.60	1	0.81
1	1	0.90	1	0.70
1	1	0.90	0	0.78
1	0	0.60	0	0.81
0	0	0.90	0	0.70
0	0	0.90	1	0.78

Table 5.8: Incorrect decoding with a four-error tainted code.

To prepare further ground for comparison dedicated specifically to error-control structures II to IV, we study the decoding of two-error tainted codes in greater details. In particular, we are interested in how *error-patterns* affect the decoding capability. The error-patterns of the two errors can be categorized into the following:

1. both errors occur in S_0;
2. both errors occur in S_1;
3. both errors occur in S_2;
4. one error in S_0 and the other in S_1;
5. one error in S_0 and the other in S_2;
6. one error in S_1 and the other in S_2;

Table 5.9 enlists the computational results regarding the two-error tainted codes of 000000111, and we assume the following parameters:

1. probability for a transmission error, $err = 0.01$;
2. the assignment probability, $ass = 0.99999$;
3. initial confidence of observation , $\kappa = 99$.

2-error tainted codes	recovered codes	confidence level (κ)
$\underline{11}0000111$	$\underline{111}000111$	1000
$\underline{1}0\underline{1}000111$	$\underline{111}000111$	1000
$000\underline{11}0111$	000000111	100
$000\underline{1}0\underline{1}111$	000000111	100
$000000\underline{00}1$	000000111	100
$000000\underline{0}1\underline{0}$	000000111	100
$\underline{1}00\underline{1}00111$	000000111	3
$\underline{1}000\underline{1}0111$	$0\underline{111}0\underline{10}1\underline{0}$	100
$0\underline{1}0\underline{1}00111$	000000111	3
$0\underline{1}00\underline{1}0111$	000000111	3
$\underline{1}00000\underline{0}11$	000000111	3
$\underline{1}000001\underline{0}1$	$0\underline{111}0\underline{10}1\ \underline{0}$	100
$000\underline{1}00\underline{0}11$	000000111	100
$000\underline{1}001\underline{0}1$	000000111	100
⋮	⋮	⋮

Table 5.9: Two-error decoding of error-control structure II.

It is evident from the results that decoding capability varies with the exhibited error-patterns; we shall see the same manifestation in the forthcoming error-control structures. Here we pay special attention to the confidence of decoded results: they actually vary; even amongst the correctly decoded results some are as good as 1000 but some are just 3. Actually, if we decrease say κ, we shall come up with incorrect results. This is something fragile in the coupling of the parameters err, ass and κ that underlies all our probabilistic decoding schemes. Can we track such upcoming? That will be one of the important tasks next chapter.

5.6 Error-control structure III

Error-control structure III shares the same code structure, hence the local sub-universes and global topology with error-control structure II; what differs is that a different global information flow scheduling is deployed. Recall that with error-control structure II, iterations between the two decoders – $\{S_0, S_1\}$ and $\{S_0, S_2\}$ – are implemented *simultaneously* and on the basis of *individual sub-universes*; with error-control structure III, iterations are to be implemented between the decoders *consecutively*. That means, we shall run the decoders *in turn*, and iterate information in between. Our proposal to error-control structure III reads:

Decoding algorithm for error-control structure III:

1. *Initialization of the code structure:*
   We let $P(x_i) = 0.5$ for both possible outcomes of the root events in S_0; that means we have no prior evidence of what the code is likely to be. Then initialize the joint probabilities of individual schemata in the cyclic SL successively according to the original BN; this results in a globally consistent SL to begin with.

2. *Assignment of updated evidential probabilities:*
   We put $P^*_{X_iY_i}(y_i = \bar{y}_i) = k$ where $0.5 \le k \le 1$ stands for a probability of belief of an observed event. For the other possible value of y_i, we assign $P^*_{X_iY_i}(y_i \ne \bar{y}_i) = 1 - k$. Similar assignments hold for schemata C_iD_i and $C_i'D_i'$.

3. *Absorption of evidence I:*
   The individual sub-universes in the evidence layers $\{X_iY_i, i = 1, 2, 3\}$ are updated as follows:

$$P^*_{X_iY_i}(x_iy_i) \leftarrow P_{X_iY_i}(x_iy_i)\frac{P^*_{X_iY_i}(y_i)}{P_{X_iY_i}(y_i)} \tag{5.24}$$

   and similarly for C_iD_i and $C_i'D_i'$.

   Set counter *count* $= 0$

   Do while [the SL is not globally consistent]

4. *Decoder 1: Input* $\{P^*_{X_iY_i}(x_iy_i), P^*_{C_iD_i}(c_id_i) | i = 1, 2, 3\}$

   (a) *Absorption of evidence II:*
   Next, evidence is absorbed from the evidence layers into the structure layer. Each schema within the structure layer absorbs evidence from schemata in the evidence layers through the intersecting elements.

So, for instance, the schema $X_1X_2C_1$ absorbs evidence from the set of schemata $\{X_1Y_1, X_2Y_2, C_1D_1\}$; updating formulae to $X_1X_2C_1$ read:
if *count* $= 0$, we have

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right)\left(\frac{P^*_{C_1D_1}(c_1)}{P_{X_1X_2C_1}(c_1)}\right) \tag{5.25}$$

or if *count* > 0, we have instead

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right) \tag{5.26}$$

Similar calculations hold for $P^*_{X_2X_3C_2}(x_2x_3c_2)$ and $P^*_{X_1X_3C_3}(x_1x_3c_3)$.

(b) *Checking evidence against the code structure (Decoder 1):*
Next, a schema within the structure layer is updated by other schemata in the structure layer through their intersecting elements; thus updating to $X_1X_2C_1$ reads

$$P^{**}_{X_1X_2C_1}(x_1x_2c_1) \leftarrow P^*_{X_1X_2C_1}(x_1x_2c_1)\frac{P^*_{X_1X_3C_3}(x_1)P^*_{X_2X_3C_2}(x_2)}{P^*_{X_1X_2C_1}(x_1x_2)} \tag{5.27}$$

Similar calculations hold for $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ and $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$.

(c) *Output of evidence:*
After a check against the code structure, the updated values of X_i are fed back to the evidence layers through again the intersecting elements. Thus, for example, we have

$$P^{**}_{X_1Y_1}(x_1y_1) \leftarrow \alpha P^*_{X_1Y_1}(x_1y_1)\frac{P^{**}_{X_1X_2C_1}(x_1)P^{**}_{X_1X_3C_3}(x_1)}{p^*_{X_1Y_1}(x_1)^2} \tag{5.28}$$

and similary for $P^{**}_{X_iY_i}(x_iy_i), \quad i = 2, 3$; besides,

$$P^{**}_{C_1D_1}(c_1d_1) \leftarrow P^*_{C_1D_1}(c_1d_1)\frac{P^{**}_{X_1X_2C_1}(c_1)}{P^*_{C_1D_1}(c_1)} \tag{5.29}$$

and similarly for $P^{**}_{C_iD_i}(c_id_i), \quad i = 2, 3$.

(d) *Rename items:*
We rename $P^{**}_{X_iY_i}(x_iy_i) \rightarrow P^*_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

5. *Decoder 2: Input* $\{P^*_{X_iY_i}(x_iy_i), P^*_{C_i'D_i'}(c_i'd_i')|i = 1, 2, 3\}$

Information processing with decoder 2:

Since both decoders are of the same structure, decoding processed with decoder 2 is the same as that with decoder 1. Therefore, we simply repeat step 4a to step 6 with the the present input.

6. *Renaming items:*

   Rename items for further iterations:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^{*}_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

   (b) $P^{**}_{C_iD_i}(c_id_i) \to P^{*}_{C_iD_i}(c_id_i)$ for $i = 1, 2, 3$.

   (c) $P^{**}_{C_i'D_i'}(c_i'd_i') \to P^{*}_{C_i'D_i'}(c_i'd_i')$ for $i = 1, 2, 3$.

   (d) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_2C_1}(x_1x_2c_1)$, and similary for other schemata inside the structure layer.

   count = *count* + 1

   enddo

7. Upon the attainment of global consistency, belief of individual variables can be extracted from the individual sub-universes as follows:

$$\begin{aligned} P(x_i) &= \sum_{y_i} P^*_{X_iY_i}(x_i, y_i) \\ P(c_i) &= \sum_{d_i} P^*_{C_iD_i}(c_i, d_i) \\ P(c_i') &= \sum_{d_i'} P^*_{C_i'D_i'}(c_i', d_i') \end{aligned} \tag{5.30}$$

8. To read off the decoded code-word, we pick

$$\begin{aligned} X_i &= \arg\left\{\max_{x_i}[P^t(x_i)]\right\} \\ &\equiv \begin{cases} x_{i_0} & \text{if } P^t(x_{i_0}) > P^t(x_{i_1}) \\ x_{i_1} & \text{if } P^t(x_{i_1}) > P^t(x_{i_0}) \end{cases} \end{aligned} \tag{5.31}$$

   for $i = 1, 2, 3$; similar assignments hold for C_i and C_i'. If it happens that $P^t(x_{i_0}) = P^t(x_{i_1})$, we pick one value arbitrarily.

5.6.1 Computational results

Without enlisting detailed results again, we simply record here that all single-error tainted codes are correctly decodable with high confidence of belief. Here, let us enter discussion into decoding of two-error codes related to the error-patterns directly. As is assumed previously, we put $err = 0.01$, $ass = 0.99999$ and $\kappa = 99$; in table 5.10 we enlist results implemented with the two-error tainted codes of the original code-word 000000111.

2-error tainted codes	recovered codes	confidence level (κ)
$\underline{11}0000111$	$\underline{111}000111$	1000
$\underline{1}0\underline{1}000111$	$\underline{111}000111$	1000
$000\underline{11}0111$	000000111	100
$000\underline{1}0\underline{1}111$	000000111	100
$000000\underline{00}1$	000000111	100
$000000\underline{0}1\underline{0}$	000000111	100
$\underline{1}00\underline{1}00111$	000000111	100
$\underline{1}000\underline{1}0111$	$\underline{111}000111$	80
$0\underline{1}0\underline{1}00111$	000000111	100
$0\underline{1}00\underline{1}0111$	000000111	100
$\underline{1}00000\underline{0}11$	000000111	100
$\underline{1}000001\underline{0}1$	000000111	100
$000\underline{1}00\underline{0}11$	000000111	100
$000\underline{1}001\underline{0}1$	000000111	100
⋮	⋮	⋮

Table 5.10: Two-error decoding of error-control structure III.

From Table 5.10 we observed that in comparison with error-control structure II, the present decoding is more confident when dealing with the two-error combinations of one error in the information set and one in the error-control set. This is an imporant difference due to the decoding dynamics which we shall discuss later in details. Lastly, as an observation resonating with the previous structures, we notice that decoding with structure III is again *structure preserving* in general.

5.7 Error-control structure IV

Our last error-control structure is error-control structure III with a parallel implementation of decoding with both decoders; however, unlike error-control structure II, decoding is strictly confined within each decoder. We propose the decoding algorithm as follows.

Decoding algorithm for error-control structure IV:

1. *Initialization of the code structure:*
   We let $P(x_i) = 0.5$ for both possible outcomes of the root events in S_0; that means we have no prior evidence of what the code is likely to be. Then initialize the joint probabilities of individual schemata in the cyclic SL successively according to the original BN; this results in a globally consistent SL to begin with.

2. *Assignment of updated evidential probabilities:*
   We put $P^*_{X_iY_i}(y_i = \bar{y}_i) = k$ where $0.5 \leq k \leq 1$ stands for a probability of belief of an observed event. For the other possible value of y_i, we assign $P^*_{X_iY_i}(y_i \neq \bar{y}_i) = 1 - k$. Similar assignments hold for schemata C_iD_i and $C'_iD'_i$.

3. *Absorption of evidence I:*
   The individual sub-universes in the evidence layers $\{X_iY_i, i = 1, 2, 3\}$ are updated as follows:

$$P^*_{X_iY_i}(x_iy_i) \leftarrow P_{X_iY_i}(x_iy_i)\frac{P^*_{X_iY_i}(y_i)}{P_{X_iY_i}(y_i)} \tag{5.32}$$

   and similarly for C_iD_i and $C'_iD'_i$.

   Set counter *count* $= 0$

   Do while [the SL is not globally consistent]

4. *Absorption of evidence II: (within individual decoders)*
   Next, evidence is absorbed from the evidence layers into the structure layer. Each schema within the structure layer absorbs evidence from schemata in the evidence layers through the intersecting elements. So, for instance, the schema $X_1X_2C_1$ absorbs evidence from the set of schemata $\{X_1Y_1, X_2Y_2, C_1D_1\}$; updating formula to $X_1X_2C_1$ reads:

if *count* $= 0$, we have

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right)\left(\frac{P^*_{C_1D_1}(c_1)}{P_{X_1X_2C_1}(c_1)}\right) \tag{5.33}$$

or if *count* > 0, we have instead

$$P^*_{X_1X_2C_1}(x_1x_2c_1) \leftarrow \alpha P_{X_1X_2C_1}(x_1x_2c_1)\left(\frac{\prod_{i=1,2} P^*_{X_iY_i}(x_i)}{P_{X_1X_2C_1}(x_1x_2)}\right) \tag{5.34}$$

Similar calculations hold for $P^*_{X_2X_3C_2}(x_2x_3c_2)$ and $P^*_{X_1X_3C_3}(x_1x_3c_3)$, together with the counterparts of decoder 2.

5. *Checking evidence against the code structure: (within individual decoders)*
   Next, a schema within the structure layer is updated by other schemata within the structure layer and belonging to the same decoder; thus updating to $X_1X_2C_1$ reads

$$P^{**}_{X_1X_2C_1}(x_1x_2c_1) \leftarrow P^*_{X_1X_2C_1}(x_1x_2c_1)\frac{P^*_{X_1X_3C_3}(x_1)P^*_{X_2X_3C_2}(x_2)}{P^*_{X_1X_2C_1}(x_1x_2)} \tag{5.35}$$

   Similar calculations hold for $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ and $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$ of S_1. For decoder 2 we have in parallel that

$$P^{**}_{X_1X_2C_1'}(x_1x_2c_1') \leftarrow P^*_{X_1X_2C_1'}(x_1x_2c_1')\frac{P^*_{X_1X_3C_3'}(x_1)P^*_{X_2X_3C_2'}(x_2)}{P^*_{X_1X_2C_1'}(x_1x_2)} \tag{5.36}$$

   and similarly for other schemata in S_2.

6. *Output of evidence:*
   After a check against the code structure, the updated values of X_i are fed back to the evidence layers through the intersecting elements. Thus, for example, we have

$$P^{**}_{X_1Y_1}(x_1y_1) = \alpha P^*_{X_1Y_1}(x_1y_1)\frac{P^{**}_{X_1X_2C_1}(x_1)P^{**}_{X_1X_3C_3}(x_1)P^{**}_{X_1X_2C_1'}(x_1)P^{**}_{X_1X_3C_3'}(x_1)}{p^*_{X_1Y_1}(x_1)^4} \tag{5.37}$$

   and similary for $P^{**}_{X_iY_i}(x_iy_i)$, $i = 2, 3$; besides,

$$P^{**}_{C_1D_1}(c_1d_1) = P^*_{C_1D_1}(c_1d_1)\frac{P^{**}_{X_1X_2C_1}(c_1)}{P^*_{C_1D_1}(c_1)} \tag{5.38}$$

   and similarly for $P^{**}_{C_iD_i}(c_id_i)$, $i = 2, 3$. Adaptation to the schemata of S_2 is immediate.

7. *Renaming items:*

   Rename items for further iterations:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^{*}_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

   (b) $P^{**}_{C_iD_i}(c_id_i) \to P^{*}_{C_iD_i}(c_id_i)$ for $i = 1, 2, 3$.

   (c) $P^{**}_{C'_iD'_i}(c'_id'_i) \to P^{*}_{C'_iD'_i}(c'_id'_i)$ for $i = 1, 2, 3$.

   (d) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_2C_1}(x_1x_2c_1)$, and similary for other schemata inside the structure layer.

   count = *count* + 1

   enddo

8. On attaining global consistency, belief of individual variables can be extracted from the individual sub-universes as follows:

$$
\begin{aligned}
P(x_i) &= \sum_{y_i} P^{*}_{X_iY_i}(x_i, y_i) \\
P(c_i) &= \sum_{d_i} P^{*}_{C_iD_i}(c_i, d_i) \\
P(c'_i) &= \sum_{d'_i} P^{*}_{C'_iD'_i}(c'_i, d'_i)
\end{aligned}
\tag{5.39}
$$

9. To read off the decoded code-word, we pick

$$
\begin{aligned}
X_i &= \arg\left\{\max_{\mathrm{x_i}}[P^t(x_i)]\right\} \\
&\equiv \left\{ \begin{array}{ll} x_{i_0} & \text{if } P^t(x_{i_0}) > P^t(x_{i_1}) \\ x_{i_1} & \text{if } P^t(x_{i_1}) > P^t(x_{i_0}) \end{array} \right.
\end{aligned}
\tag{5.40}
$$

   for $i = 1, 2, 3$; similar assignments hold for C_i and C'_i. If it happens that $P^t(x_{i_0}) = P^t(x_{i_1})$, we pick one value arbitrarily.

5.7.1 Computational results

We carried out similar investigations with error-control structure III. Without listing out the results, we just remark that the decoding capability of the present structure is very similar to that of structure III. However, we notice the following important difference. In the decoding of two-error code-words, we noticed in table 5.10 that with regard to the following tainted codes: $\underline{1}000\underline{1}0111$ and $\underline{1}00000\underline{1}01$, only the latter one is correctly decodable; but notice, decoding by structure IV renders both *no* remedy. Clearly, while the two tainted code-words play the same role since they share fundamentally the same structure (it is only a precise assignment of a 1 or a 0 that differs), the difference in decoding *must* be due to the decoding dynamics pertinent to individual structures; this is an important mystery that we shall crack in the coming chapter. Lastly, to resonate with previous observations, we notice here that decoding with structure IV is again *structure preserving*.

5.8 Conclusion

We have implemented in this chapter several versions of cyclic probabilistic reasoning networks based on SL, aspiring at specifically error-control applications. With regard to the computational procedures, what we have implemented amount to a coupling of the following:

1. a local joint probability structure within each sub-universe;
2. a global topology according to which the sub-universes join together – with which we confronted cyclic settings specifically;
3. an updating scheme comprising first local information updating formulae, and second a global scheduling on the overall information flow throughout the network.

There are two important aspects noticed this chapter. First, regarding the structure as a mathematical entity we came up with the important observation that the dynamics *seem* to favour a result that complies with the prescribed structures. We shall show in the next chapter that this is an important property which enables the functioning towards error-control applications. Second, regarding the error correcting capability of the structures proposed, all four of which are found *unconditionally* single-error correcting; but for two-error codes decoding is *conditional* on the error-patterns. All these seem mystical on a first look, however, we shall show in the next chapter that they are in fact liable to effective explanation.

Chapter 6
Dynamics on cyclic probabilistic reasoning networks

6.1 Overview

In the previous chapter, we have successfully implemented some error-control structures on the basis of cyclic probabilistic reasoning networks, and thus testified to the possible usefulness of such structures in our daily world. However, as is true with nearly all studies of complex systems, what has been most puzzling is *why* and *how* the observed phenomena come about. *Why* and *how* does a bar magnet gain a global magnetization on lowering the surrounding temperature? Similarly, *why* and *how* does the desired solution, the untainted code, come up on iterating initial observations on the cyclic networks we have constructed? Interesting though, this is unfortunately a question which in most circumstances escapes explanation since the computations involved inside a complex system are almost invariably untraceable! Even if analysis is possible, it is usually extremely complicated[1]. Nevertheless, as a probably blissful structure our decoding networks may be, we are going to show that analytic solutions for all our previous works are in fact feasible! Furthermore, by the solutions constructed, we shall gain a fresh perspective on *how* different elements of the decoding scheme join hands to explain *why* the desired results come about.

[1]For example, Richardson [10] deals with a very good analysis on the geometry of turbo-decoding dynamics; which however is technically so involved that it may be elusive to the untrained minds.

6.2 Decoding rationales

As *decoding* involves a suitable manipulation of observed evidence to produce the desired solution in compliance with certain pre-defined code structure, the reasoning underlying must involve the following aspects:

1. the pre-defined code structure;
2. the incorporation of observed evidence; and
3. the prescriptions which *reshape* the received observation back into the untainted code.

While a *decoding scheme* is supposed to be a suitable incorporation of the above three, an actual *implementation* of the decoding procedures involves merely some simple arithmetical operations. A question follows immediately: how are the two going to merge? What are the intrinsic meanings underlying a so-called *decoded result* and what are the rationales that qualify such a *decoded result* to be a valid answer to the *decoding problem*? In particular, with respect to the iterative decoding schemes we have proposed, several questions are in order:

1. Does the decoding scheme give an *interpretable* answer? Directed specifically to our problem, we ask whether an iteration converge at all to a solution, or does it gives rise to chaotic behaviour?
2. Is the decoding scheme structure preserving? That is, whether the decoded results conform to the pre-defined code structure.
3. If the decoded results are legitimate code-words, can we determine which particular code-word has the highest propensity to showing up corresponding to a given set of evidence? Besides, what are the criteria that govern the limit of decoding? For instance, we might conjecture the decoding capability be limited by say the minimum Hamming distance of the code-words as in usual maximum-likelihood decodings.

The first point addresses solely the dynamics of the iterative schemes, while the latter two address the characters and interpretations of the converged results. Our computational experiments in the last chapter have affirmed us positive answers to all these questions; the error-control structures do work. In this chapter, we shall carry out some analytical treatments on the underlying reasoning dynamics; by so doing, we shall be able to answer the three questions in a fresh perspective.

6.3 Error-control structure I – exact solutions

In this section we shall attack error-control structure I; however, instead of naive and straight forward computations we shall delve into the heart of the problem and derive its analytical solutions.

6.3.1 Dynamical invariant – a key to tackle many dynamical problems

Dynamics is historically born as a branch of mechanics to deal with changes of body motions in a physical background of interacting forces. There are in fact two distinct broad perspectives in dealing with these situations:

1. Local perspective: it is precisely the equations of motion, for instance we have Newton's second law in classical mechanics, Einstein's field equation in general relativity and Schrödinger's equation in quantum mechanics, etc. Mathematically, they are *differential laws* defined *locally* in space and time (or spacetime).

2. Global perspective: it includes the conservation laws, for instance we have conservation of energy, conservation of linear and angular momenta, etc. They are global constraints which limit possible evolutions of physical processes.

In order to tackle dynamical problems effectively, a suitable combination of both local and global perspectives is usually the knack. In retrospect, the decoding algorithms proposed last chapter represent the local dynamics by which we drive observations to the desired solutions, but we have hitherto touched upon no global characterization. Can we ferret some global invariants? Well, could be; but not quite: there is no guarantee that some global invariant exists; even if it does, it may not be tractable. Therefore, we should be grateful if there really exists some invariants and yet we be able to find it; otherwise, we would have to resort to other methods. With respect to the error-control structures proposed, there is indeed a simple class of invariants for structures I, II and IV, but we fail in looking up the counterpart for structure II.

6.3.2 Dynamical invariant for error-control structure I

A clue to our intended invariant is suggested by the very purpose of our work, namely a decoding scheme which incorporates observations to retrieve the untainted codes; so it is immediately apparent that the structure of the code has to be *nearly* such an invariant. Actually, it is not hard to show that an invariant pertaining to error-control structure I can be realized as follows.

We focus our consideration on the structure layer, as all information regarding the code structure is summarized there. Besides, as the three sub-universes $X_1X_2C_1$, $X_2X_3C_2$ and $X_1X_3C_3$ are symmetric with respect to their roles inside the code, our focus can be further narrowed down to either one of them, $X_1X_2C_1$ say. We let

$$\begin{aligned}\mathcal{K}_{x_1x_2c_1} &\equiv \frac{P_{X_1X_2C_1}(x_1x_2c_1)}{P_{X_1X_2C_1}(x_1x_2)}\frac{P^*_{C_1D_1}(c_1)}{P^*_{C_1D_1}(c_1)} \\ &= P_{X_1X_2C_1}(c_1|x_1x_2)\frac{P^*_{C_1D_1}(c_1)}{P_{C_1D_1}(c_1)}\end{aligned} \tag{6.1}$$

We claim that

$$L_{x_1x_2c_1} \equiv \frac{\mathcal{K}_{x_1x_2c_1}}{\Omega_{x_1x_2}} \tag{6.2}$$

where $\Omega_{x_1x_2} \equiv \sum_{c_1} \mathcal{K}_{x_1x_2c_1}$ fits our purpose. To verify this, we notice that within the decoding algorithm of error-control structure I, $X_1X_2C_1$ is updated only in steps 4 and 5 within the do-loop. Further, we notice that except step 4 when *count* $= 0$, every updating to $X_1X_2C_1$ involves a structure that preserves the conditional dependence $P_{X_1X_2C_1}(c_1|x_1x_2)$. To proceed along iterations, we have for step 4 when *count* > 0:

$$\begin{aligned}P^*_{X_1X_2C_1}(c_1|x_1x_2) &= \frac{P^*_{X_1X_2C_1}(x_1x_2c_1)}{P^*_{X_1X_2C_1}(x_1x_2)} \\ &= \frac{P^*_{X_1X_2C_1}(x_1x_2c_1)}{\prod_{i=1,2} P^*_{X_iY_i}(x_i)} \\ &= \frac{P_{X_1X_2C_1}(x_1x_2c_1)}{P_{X_1X_2C_1}(x_1x_2)} \\ &= P_{X_1X_2C_1}(c_1|x_1x_2)\end{aligned} \tag{6.3}$$

and for step 5:

$$P^{**}_{X_1X_2C_1}(c_1|x_1x_2) = \frac{P^{**}_{X_1X_2C_1}(x_1x_2c_1)}{P^{**}_{X_1X_2C_1}(x_1x_2)}$$

$$
\begin{aligned}
&= \frac{P^{**}_{X_1X_2C_1}(x_1x_2c_1)}{P^{**}_{X_1X_3C_3}(x_1)P^{**}_{X_2X_3C_2}(x_2)} \\
&= \frac{P^{*}_{X_1X_2C_1}(x_1x_2c_1)}{P^{*}_{X_1X_2C_1}(x_1x_2)} \\
&= P^{*}_{X_1X_2C_1}(c_1|x_1x_2) \\
&= P_{X_1X_2C_1}(c_1|x_1x_2) \qquad (6.4)
\end{aligned}
$$

with the last equality follows from (6.3). Therefore, the conditional probability $P(c_1|x_1x_2)$ is a constant of motion. To determine what this constant is, we return to the very beginning of the iteration, namely step 4 when *count* = 0. With (6.1), we can reformulate step 4 when *count* = 0 as

$$P^{*}_{X_1X_2C_1}(x_1x_2c_1) = \alpha\,\mathcal{K}_{x_1x_2c_1} \prod_{i=1,2} P^{*}_{X_iY_i}(x_i) \qquad (6.5)$$

from which we have

$$P^{*}_{X_1X_2C_1}(x_1x_2) = \alpha\,\Omega_{x_1x_2} \prod_{i=1,2} P^{*}_{X_iY_i}(x_i) \qquad (6.6)$$

and determine that

$$
\begin{aligned}
P^{*}_{X_1X_2C_1}(c_1|x_1x_2) &= \frac{P^{*}_{X_1X_2C_1}(x_1x_2c_1)}{P^{*}_{X_1X_2C_1}(x_1x_2)} \\
&= \frac{\alpha\,\mathcal{K}_{x_1x_2c_1}\prod_{i=1,2} P^{*}_{X_iY_i}(x_i)}{\alpha\,\Omega_{x_1x_2}\prod_{i=1,2} P^{*}_{X_iY_i}(x_i)} \\
&= L_{x_1x_2c_1} \qquad (6.7)
\end{aligned}
$$

Thus $L_{x_1x_2c_1}$ is a constant of motion as claimed; we can similarly define $L_{x_1x_3c_3}$ and $L_{x_2x_3c_2}$ as the other constants of motion.

6.3.3 Iteration dynamics

With a constant of motion identified, we are fully equipped for a detailed study on the iteration dynamics of the decoding algorithm. We shall derive in this section the exact solutions of error-control structure I, thereby showing that the decoding scheme converges in two iterations, i.e. it terminates at *count* = 1!

In the following discussion we shall take as a representative the schema $X_1X_2C_1$ when dealing with updating of the structure layer's elements, and the schema X_1Y_1 when dealing with the evidence layer's counterparts. The knack underlying the upcoming discussions is to reformulate the decoding algorithm by suitably incorporating constants like $L_{x_1x_2c_1}$ and so forth. By tracing through the computations involved, we shall prove the claimed results.

Assume new evidence has been absorbed to the schemata X_iY_i and C_iD_i, so we have the set of incoming evidence: $\{P^e_{X_iY_i}(x_iy_i), P^e_{C_iD_i}(c_id_i)|i = 1,2,3\}$[2], which serves as the perturbation to trigger the decoding dynamics.

we proceed from step 4 onwards which initiates the following computations:

The first iteration, *count* $= 0$

1. *Absorption of evidence II:*

   We reformulate (5.7) by incorporating $\mathcal{K}_{x_1x_2c_1}$, by which we obtain[3]

$$P^*_{X_1X_2C_1}(x_1x_2c_1) = \alpha\, \mathcal{K}_{x_1x_2c_1} \prod_{i=1,2} P^e_{X_iY_i}(x_i) \tag{6.8}$$

2. *Checking evidence against the code structure:*

   By incorporating (6.8), (5.9) can be reformulated as follows,

$$\begin{aligned} P^{**}_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\, \frac{\mathcal{K}_{x_1x_2c_1}\prod_{i=1,2} P^e_{X_iY_i}(x_i)}{(\sum_{c_1}\mathcal{K}_{x_1x_2c_1})\prod_{i=1,2} P^e_{X_iY_i}(x_i)}\, \lambda_{x_1}\lambda_{x_2} P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \\ &= \alpha\, \frac{\mathcal{K}_{x_1x_2c_1}\prod_{i=1,2} P^e_{X_iY_i}(x_i)}{\Omega_{x_1x_2}\prod_{i=1,2} P^e_{X_iY_i}(x_i)}\, \lambda_{x_1}\lambda_{x_2} P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \\ &= \alpha\, \lambda_{x_1}\, \lambda_{x_2}\, L_{x_1x_2c_1} P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \end{aligned} \tag{6.9}$$

   where

$$\begin{aligned} \lambda_{x_1} &\equiv \sum_{x_3c_3} \mathcal{K}_{x_1x_3c_3} p^e_{X_3Y_3}(x_3) \\ \lambda_{x_2} &\equiv \sum_{x_3c_2} \mathcal{K}_{x_2x_3c_2} p^e_{X_3Y_3}(x_3) \end{aligned} \tag{6.10}$$

   To facilitate further discussions, we take the trouble writing out $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$ and $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ explicitly:

   For $X_2X_3C_2$, we have

$$P^{**}_{X_2X_3C_2}(x_2x_3c_2) = \alpha\, \eta_{x_2}\eta_{x_3} L_{x_2x_3c_2} P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3) \tag{6.11}$$

[2]In the analytical approach this chapter, we denote evidence by an explicit superscript ^{e} instead of a * as used in presenting the decoding algorithms last chapter.

[3]In the analytic treatment we shall replace the assignment symbol "←" by the equality sign "=" following the common practice of dynamical treatment.

with

$$\eta_{x_2} = \sum_{x_1 c_1} \mathcal{K}_{x_1 x_2 c_1} p^e_{X_1 Y_1}(x_1)$$
$$\eta_{x_3} = \sum_{x_1 c_3} \mathcal{K}_{x_1 x_3 c_3} P^e_{X_1 Y_1}(x_1) \quad (6.12)$$

whereas for $X_1 X_3 C_3$, we have

$$P^{**}_{X_1 X_3 C_3}(x_1 x_3 c_3) = \alpha\, \zeta_{x_1} \zeta_{x_3} L_{x_1 x_3 c_3} P^e_{X_1 Y_1}(x_1) P^e_{X_3 Y_3}(x_3) \quad (6.13)$$

with

$$\zeta_{x_1} = \sum_{x_2 c_1} \mathcal{K}_{x_1 x_2 c_1} p^e_{X_2 Y_2}(x_2)$$
$$\zeta_{x_3} = \sum_{x_2 c_2} \mathcal{K}_{x_2 x_3 c_2} P^e_{X_2 Y_2}(x_2) \quad (6.14)$$

We remark on the special roles played by λ, η and ζ: they will be seen to represent the *correcting force* which *alters* and *reshapes* a tainted code back its pristine state – that is the significance which deserves us the trouble writing them out explicitly.

3. *Output of evidence:*
After iterating the new evidence against the pre-defined code structure, information is fed back to the evidence layers. However, the roles played by $X_i Y_i$ and $C_i D_i$ are different: each $X_i Y_i$ is connected to *two* schemata in the structure layer, whereas each $C_i D_i$ is connected to only *one*, therefore, each $X_i Y_i$ is a *mediator* between the two schemata it is connected to upon reception of information from them; in contrast each $C_i D_i$ is a passive information receptor of the fed back information. Due to this difference, in the investigation of the iteration dynamics, it suffices to bother ourselves with only the fed back information to $X_i Y_i$. Therefore, we have

$$P^{**}_{X_1 Y_1}(x_1 y_1) = \alpha\, \frac{P^{*}_{X_1 Y_1}(x_1 y_1)}{P^{*}_{X_1 Y_1}(x_1)^2} P^{*}_{X_1 X_2 C_1}(x_1) P^{*}_{X_1 X_3 C_3}(x_1) \quad (6.15)$$

Heed what $P^{*}_{X_1 X_2 C_1}(x_2)$ and $P^{*}_{X_1 X_3 C_3}(x_1)$ give; for say $P^{*}_{X_1 X_2 C_1}(x_1)$ reads[4]

$$P^{*}_{X_1 X_2 C_1}(x_1) = \sum_{x_2 c_1} \lambda_{x_1} \lambda_{x_2} L_{x_1 x_2 c_1} P^e_{X_1 Y_1}(x_1) P^e_{X_2 Y_2}(x_2)$$

[4]Recall that we have absorbed all computations involving renormalization constants by the "α", so we can forget such "α" henceforth in the following computations.

$$= \lambda_{x_1} P^e_{X_1Y_1}(x_1) \left(\sum_{x_2c_1} L_{x_1x_2c_1} \lambda_{x_2} P^e_{X_2Y_2}(x_2) \right) \tag{6.16}$$

with which we notice that

$$\begin{aligned} \sum_{x_2c_1} L_{x_1x_2c_1} \lambda_{x_2} P^e_{X_2Y_2}(x_2) &= \sum_{x_2} \lambda_{x_2} P^e_{X_2Y_2}(x_2) \sum_{c_1} L_{x_1x_2c_1} \\ &= \sum_{x_2} \lambda_{x_2} P^e_{X_2Y_2}(x_2) \frac{\sum_{c_1} \mathcal{K}_{e_1x_2c_1}}{\Omega_{x_1x_2}} \\ &= \sum_{x_2} \lambda_{x_2} P^e_{X_2Y_2}(x_2) \end{aligned} \tag{6.17}$$

which does not depend on x_1 and hence amounts to a constant for (6.15) and (6.16); therefore, by (6.17) we can rewrite (6.16) as

$$P^*_{X_1X_2C_1}(x_1) = \lambda_{x_1} P^e_{X_1Y_1}(x_1) \tag{6.18}$$

Similarly, we have

$$P^*_{X_1X_3C_3}(x_1) = \zeta_{x_1} P^e_{X_1Y_1}(x_1) \tag{6.19}$$

By (6.18) and (6.19), we have

$$\begin{aligned} P^{**}_{X_1Y_1}(x_1y_1) &= \alpha\, \lambda_{x_1} \zeta_{x_1} P^e_{X_1Y_1}(x_1y_1) \\ &= \alpha\, \gamma_{x_1} P^e_{X_1Y_1}(x_1y_1) \end{aligned} \tag{6.20}$$

where $\gamma_{x_1} \equiv \lambda_{x_1}\zeta_{x_1}$. Similar results are obtained in the same way and we have

$$\begin{aligned} P^{**}_{X_2Y_2}(x_2y_2) &= \alpha\, \gamma_{x_2} P^e_{X_2Y_2}(x_2y_2) \\ P^{**}_{X_3Y_3}(x_3y_3) &= \alpha\, \gamma_{x_3} P^e_{X_3Y_3}(x_3y_3) \end{aligned} \tag{6.21}$$

where $\gamma_{x_2} \equiv \lambda_{x_2}\eta_{x_2}$ and $\gamma_{x_3} \equiv \eta_{x_3}\zeta_{x_3}$ respectively.

4. *Renaming items:* To proceed with further iterations, we rename items:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^*_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

   (b) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_{2}c_1}(x_1x_2c_1)$, and similary for $P^{**}_{X_1X_3C_3}(x_1x_3c_3)$ and $P^{**}_{X_2X_3C_2}(x_2x_3c_2)$.

So this concludes the first iteration, next we proceed to the second.

The second iteration, *count* $= 1$

5. *Absorption of evidence II:*

The mediated information (6.21) is fed back to the structure layer. We have

$$\begin{aligned} P^*_{X_1X_2C_1}(x_1x_2c_1) &= \alpha \frac{P_{X_1X_2C_1}(x_1x_2c_1)}{P_{X_1X_2}(x_1x_2)} P^*_{X_1Y_1}(x_1)P^*_{X_2Y_2}(x_2) \\ &= \alpha\, L_{x_1x_2c_1}\left(\sum_{y_1} P^e_{X_1Y_1}(x_1y_1)\gamma_{x_1}\right)\left(\sum_{y_2} P^e_{X_2Y_2}(x_2y_2)\gamma_{x_2}\right) \\ &= \alpha\, \gamma_{x_1}\gamma_{x_2} L_{x_1x_2c_1} P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \end{aligned} \tag{6.22}$$

Similar calculations hold for other schemata for which we have

$$\begin{aligned} P^*_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\, \gamma_{x_2}\gamma_{x_3} L_{x_2x_3c2} P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3) \\ P^*_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\, \gamma_{x_1}\gamma_{x_3} L_{x_1x_3c_3} P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3) \end{aligned} \tag{6.23}$$

Prior to any further elucidation of (6.22) and (6.23) we turn to the next step immediately to see a "miracle".

6. *Checking evidence against the code structure:*

Next, we check the updated information in the structure layer again, we see

$$\begin{aligned} P^{**}_{X_1X_2C_1}(x_1x_2c_1) &= \alpha \frac{P^*_{X_1X_2C_1}(x_1x_2c_1)}{P^*_{X_1X_2C_1}(x_1x_2)} P^*_{X_2X_3C_2}(x_2)P^*_{X_1X_3C_3}(x_1) \\ &= \alpha\, L_{x_1x_2c_1} P^e_{X_2X_3C_2}(x_2)P^e_{X_1X_3C_3}(x_1) \end{aligned} \tag{6.24}$$

Heed what $P^*_{X_2X_3C_2}(x_2)$ and $P^*_{X_1X_3C_3}(x_1)$ will give us this time. We consider $P^*_{X_2X_3C_2}(x_2)$ say, we have

$$\begin{aligned} P^*_{X_2X_3C_2}(x_2) &= \sum_{x_3c_2} \gamma_{x_2}\gamma_{x_3} L_{x_2x_3c_2} P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3) \\ &= \gamma_{x_2} P^e_{X_2Y_2}(x_2)\left(\sum_{x_3c_2} \gamma_{x_3} L_{x_2x_3c_2} P^e_{X_3Y_3}(x_3)\right) \end{aligned} \tag{6.25}$$

Notice again

$$\begin{aligned} \sum_{x_3c_2} \gamma_{x_3} L_{x_2x_3c_2} P^e_{X_3Y_3}(x_3) &= \sum_{x_3} \gamma_{x_3} P^e_{X_3Y_3}(x_3) \sum_{c_2} L_{x_2x_3c_2} \\ &= \sum_{x_3} \gamma_{x_3} P^e_{X_3Y_3}(x_3) \frac{\sum_{c_2} \mathcal{K}_{x_2x_3c_2}}{\Omega_{x_2x_3}} \\ &= \sum_{x_3} \gamma_{x_3} P^e_{X_3Y_3}(x_3) \end{aligned} \tag{6.26}$$

which amounts to a constant for (6.25). Thus we have

$$P^*_{X_2X_3C_2}(x_2) = \gamma_{x_2} P^e_{X_2Y_2}(x_2) \tag{6.27}$$

By similar calculations we have

$$P^{*}_{X_1X_3C_3}(x_1) = \gamma_{x_1} P^{e}_{X_1Y_1}(x_1) \tag{6.28}$$

By (6.27) and (6.28), (6.24) reads

$$P^{**}_{X_1X_2C_1}(x_1x_2c_1) = \alpha\, \gamma_{x_1}\gamma_{x_2} L_{x_1x_2c_1} P^{e}_{X_1Y_1}(x_1) P^{e}_{X_2Y_2}(x_2) \tag{6.29}$$

which is **identical** to (6.22)!! This means the iterations have converged; not just converged, but we have obtained actually the exact analytic solutions! Likewise, we can derive the corresponding analytic solutions for the other schemata which are briefly summarized in (6.35). To confirm that the iterations have terminated, we check the next updating to the evidence layer.

7. *Output of evidence:*

   Let us consider X_1Y_1, the updating reads

$$P^{**}_{X_1Y_1}(x_1y_1) = \alpha\, P^{*}_{X_1Y_1}(x_1y_1) \frac{P^{**}_{X_1X_2C_1}(x_1) P^{**}_{X_1X_3C_3}(x_1)}{P^{*}_{X_1Y_1}(x_1)^2} \tag{6.30}$$

   With reference to (6.27) and (6.28), the fed back information $P^{**}_{X_1X_2C_1}(x_1)$ and $P^{**}_{X_1X_3C_3}(x_1)$ can be immediately inferred/read; so we have

$$P^{**}_{X_1X_2C_1}(x_1) = P^{*}_{X_1X_3C_3}(x_1) = \gamma_{x_1} P^{e}_{X_1Y_1}(x_1) \tag{6.31}$$

   However, by (6.20) we also have

$$P^{*}_{X_1Y_1}(x_1) = \gamma_{x_1} P^{e}_{X_1Y_1}(x_1)$$

   Thus (6.30) gives

$$P^{**}_{X_1Y_1}(x_1y_1) = P^{*}_{X_1Y_1}(x_1y_1) \tag{6.32}$$

   testifying to the termination of the iterative calculations[5].

8. *Output results:*

   So we have got the analytic solutions for error-control structure I; here, let us briefly summarize the results as follows,

[5]The renormalization constant must be simply unity since $P^{*}_{X_1Y_1}(x_1y_1)$ is assumed normalized.

(a) First, we have the fundamental correctives given by

$$\begin{aligned}
\lambda_{x_1} &\equiv \sum_{x_3c_3} \mathcal{K}_{x_1x_3c_3} p^e_{X_3Y_3}(x_3) \\
\lambda_{x_2} &\equiv \sum_{x_3c_2} \mathcal{K}_{x_2x_3c_2} p^e_{X_3Y_3}(x_3) \\
\eta_{x_2} &= \sum_{x_1c_1} \mathcal{K}_{x_1x_2c_1} p^e_{X_1Y_1}(x_1) \\
\eta_{x_3} &= \sum_{x_1c_3} \mathcal{K}_{x_1x_3c_3} P^e_{X_1Y_1}(x_1) \\
\zeta_{x_1} &= \sum_{x_2c_1} \mathcal{K}_{x_1x_2c_1} p^e_{X_2Y_2}(x_2) \\
\zeta_{x_3} &= \sum_{x_2c_2} \mathcal{K}_{x_2x_3c_2} P^e_{X_2Y_2}(x_2)
\end{aligned} \tag{6.33}$$

from which we derive the correctives for individual information bits as follows:

$$\begin{aligned}
\gamma_{x_1} &= \lambda_{x_1}\zeta_{x_1} \\
\gamma_{x_2} &= \lambda_{x_2}\eta_{x_2} \\
\gamma_{x_3} &= \eta_{x_3}\zeta_{x_3}
\end{aligned} \tag{6.34}$$

(b) Next, by invoking the correctives in (6.34), we have the converged solutions for each schemata in the structure layer, namely[6]

$$\begin{aligned}
P^t_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\,\gamma_{x_1}\gamma_{x_2} L_{x_1x_2c_1} P^e_{X_1Y_1}(x_1) P^e_{X_2Y_2}(x_2) \\
P^t_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\gamma_{x_2}\gamma_{x_3} L_{x_2x_3c_2} P^e_{X_2Y_2}(x_2) P^e_{X_3Y_3}(x_3) \\
P^t_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\gamma_{x_1}\gamma_{x_3} L_{x_1x_3c_3} P^e_{X_1Y_1}(x_1) P^e_{X_3Y_3}(x_3)
\end{aligned} \tag{6.35}$$

(c) From (6.35), we can derive the desired solutions, namely the degree of belief for the information bits X_i, we have

$$P^t(x_i) = \alpha\gamma_{x_i} P^e_{X_iY_i}(x_i) \quad \text{for } i = 1,2,3 \tag{6.36}$$

where we have not included any subscript since the same answer is to be obtained no matter from which schema marginalization is carried out.

(d) From (6.35), we can also derive the terminated solutions for $P^t(c_i)$; so for $i = 1$ say, we have

$$\begin{aligned}
P^t(c_1) &= \sum_{x_1x_2} P^t_{X_1X_2C_1}(x_1x_2c_1) \\
&= \sum_{x_1x_2} \alpha\, L_{x_1x_2c_1} P^t(x_1) P^t(x_2)
\end{aligned} \tag{6.37}$$

[6] We use superscript ^{t} to denote the terminated solutions.

By now, we can answer the first question raised for our decoding scheme: the scheme does converge to a solution; in fact, we have constructed the analytic solutions for the decoding dynamics. Next, we shall see how these analytic solutions help us answer the next couple of questions, namely whether the solutions belong to the pre-defined class, and what characteristics the decoded results exhibit in relation to the tainted codes received.

6.3.4 Structure preserving property and the maximum a posteriori solutions

The solutions obtained in (6.35) are very general: the dynamical constant defined in (6.2) represents far more than what we need for a practical coding structure. In a real code, the error-control bits are related to the information bits in a *deterministic* manner; therefore, the conditional probabilities of the error-control bits with respect to the information bits should (ideally) be given by a straight 0 or 1. For example, with $X_1X_2C_1$ we have

$$P_{X_1X_2C_1}(c_1|x_1x_2) = \begin{cases} 1 & \text{if } c_1 = (x_1 + x_2) \bmod 2 \\ 0 & \text{otherwise} \end{cases} \tag{6.38}$$

and similarly for the other schemata. To make notations compact, we introduce a generalized Kronecker's delta-function defined as follows,

$$\delta_{x_1x_2c_1} \equiv \begin{cases} 1 & \text{if } c_1 = (x_1 + x_2) \bmod 2 \\ 0 & \text{otherwise} \end{cases} \tag{6.39}$$

by which we can simply write

$$P_{X_1X_2C_1}(c_1|x_1x_2) = \delta_{x_1x_2c_1} \tag{6.40}$$

Now, by invoking (6.40) on (6.1), we have

$$\mathcal{K}_{x_1x_2c_1} = \delta_{x_1x_2c_1} \frac{P^e_{C_1D_1}(c_1)}{P_{C_1D_1}(c_1)} \tag{6.41}$$

through which the dynamical constant (6.2) is tantamount to[7]

$$\begin{aligned} L_{x_1x_2c_1} &= \frac{\delta_{x_1x_2c_1} P^e_{C_1D_1}(c_1)/P_{C_1D_1}(c_1)}{\sum_{c_1} \delta_{x_1x_2c_1} P^e_{C_1D_1}(c_1)/P_{C_1D_1}(c_1)} \\ &= \delta_{x_1x_2c_1} \end{aligned} \tag{6.42}$$

[7]With (6.42) we notice on one hand that $P_{C_1D_1}(c_1) = 1/2$ upon initialization so that individual fractions are always defined; on the other hand, we remind that in this thesis P^e is always assumed non-vanishing (though it can be very small) as justified before.

Thus for a real code, L is a highly constrained function on its arguments; consequently, the decoded results (6.35), being proportional to L, are represented more precisely by

$$\begin{aligned} P^t_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\,\gamma_{x_1}\gamma_{x_2}\delta_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \\ P^t_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\gamma_{x_2}\gamma_{x_3}\delta_{x_2x_3c_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3) \\ P^t_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\gamma_{x_1}\gamma_{x_3}\delta_{x_1x_3c_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3) \end{aligned} \tag{6.43}$$

Our discussion of error-control structure I culminates in the set of equations (6.43) which serves a compact answer to all the questions we have raised. However, prior to digging into the questions, let us notice an important observation first: from (6.43) we see that every line is proportional to a respective delta-function; meaning that local structures – the allowed code-word fragments like $x_1x_2c_1$ – are preserved. *Compared with BNs which realize relations through directed arrows, the delta-functions here play the same role; however, unlike a BN where local directed arrows dictate a global acyclic constraint, we are now free of this unnatural global constraint! Therefore, for SL based reasoning networks, a global topology comes only as a natural descendent from the local sub-universes chosen to build up the global knowledge base, not from any unnatural local predilection!*

To prepare the ground for further discussions, we need to place some numerical constraint on (6.43). We shall always assume in the following that the confidence of belief of an initial observation κ be large, which is a reasonable assumption regarding practical coding problems. More precisely, we shall call the limit of an extremely large κ, namely $\kappa \to \infty$ the *algebraic limit* [8] because it is where our *probabilistic* reasoning scheme coincide with conventional *algebraic* decoding scheme. To answer our previous queries, we pay attention to the structure of the solutions; notice that it comprises three portions:

1. δ – generalized Kronecker's delta-functions defining the code structure.

2. P^e – belief corresponding to the received evidence.

3. γ – correctives which drive the tainted codes back to the untainted ones.

[8] In order that the algebraic limit serves a good approximation, we need on one hand a large κ for the initial observations, and on the other hand the channel's transmission error probability *err* being small. Here we shall take it simply as a theoretical limit that approximates actual functioning.

Furthermore, we have *global consistency* between every two equations in (6.43) entailed from the decoding structure being in dynamical equilibrium. What are the consequences from all these beautiful things? First, it follows by global consistency that the same value for the information bits are to be deduced by marginalization over whatever schemata in the structure layer; in fact, by (6.36), the decoded solutions read

$$P^t(x_i) \propto \gamma_{x_i} P^e_{X_iY_i}(x_i) \tag{6.44}$$

Recall the definitions for γ, we have for γ_{x_1} say,

$$\gamma_{x_1} = \lambda_{x_1}\zeta_{x_1} \tag{6.45}$$

where in the *deterministic* limit accounted by (6.42), we have

$$\begin{aligned} \lambda_{x_1} &= \sum_{x_3c_3} \delta_{x_1x_3c_3} p^e_{X_3Y_3}(x_3) p^e_{C_3D_3}(c_3) \\ \zeta_{x_1} &= \sum_{x_2c_1} \delta_{x_1x_2c_1} p^e_{X_2Y_2}(x_2) p^e_{C_1D_1}(c_1) \end{aligned} \tag{6.46}$$

Now, ponder on what such λ_{x_1} and ζ_{x_1} mean: the delta-functions in (6.46) confine contributions to those *allowed* code fragments associated with x_1; for instance let us consider $\lambda_{x_1=0}$, by $\delta_{x_1=0,x_3c_3}$ we know that the only contributing terms are the pairs $(x_3 = 0, c_3 = 0)$ and $(x_3 = 1, c_3 = 1)$, thus

$$\lambda_{x_1=0} = p^e_{X_3Y_3}(x_3 = 0)p^e_{C_3D_3}(c_3 = 0) + p^e_{X_3Y_3}(x_3 = 1)p^e_{C_3D_3}(c_3 = 1)$$

Similarly, we have

$$\lambda_{x_1=1} = p^e_{X_3Y_3}(x_3 = 0)p^e_{C_3D_3}(c_3 = 1) + p^e_{X_3Y_3}(x_3 = 1)p^e_{C_3D_3}(c_3 = 0)$$

For decoding purpose it is the ratio $\gamma_{x_1=0}/\gamma_{x_1=1}$ that matters; we thus define a *propensity*

$$\chi^{X_1X_3C_3}_{x_1} \equiv \frac{\lambda_{x_1=0}}{\lambda_{x_1=1}} \tag{6.47}$$

with which we have: if $\chi^{X_1X_3C_3}_{x_1} > 1$, the corrective drives to $x_1 = 0$; if $\chi^{X_1X_3C_3}_{x_1} < 1$, the drive shifts to $x_1 = 1$. Hence if the received code fragment is $(x_3 = 0, c_3 = 0)$, we have $\chi^{X_1X_3C_3}_{x_1} > 1$ and the evidence from $X_1X_3C_3$ suggests $x_1 = 0$. Similarly we can define the *propensity* due to $X_1X_2C_1$,

$$\chi^{X_1X_2C_1}_{x_1} \equiv \frac{\zeta_{x_1=0}}{\zeta_{x_1=1}} \tag{6.48}$$

which works in the same way. Of course, apart from these two indirect pieces of evidence we must not leave the direct evidence $P^e_{X_1Y_1}(x_1)$ in oblivion. Taken as a whole, the determination of X_1 reduces to a study of the following ratio:

$$\begin{aligned} R_{x_1} &\equiv \frac{P^t(x_1=0)}{P^t(x_1=1)} \\ &= \chi_{x_1}^{X_1X_2C_1}\chi_{x_1}^{X_1X_3C_3}r_{x_1} \end{aligned} \tag{6.49}$$

where $r_{x_1} = P^t(x_1=0)/P^t(x_1=1)$. By (6.49), we have: if $R_{x_1} > 1$, $x_1 = 0$; or $x_1 = 1$ if $R_{x_1} < 1$ instead. The determination of R_{x_1} is a tug of war amongst the three voters:

1. r_{x_1} which represent a vote due to an initial belief on direct observation.
2. $\chi_{x_1}^{X_1X_2C_1}$ which represent another vote due to the indirect evidence X_2 and C_1 regulated by $\delta_{x_1x_2c_1}$.
3. $\chi_{x_1}^{X_1X_3C_3}$ which represent the remaining vote due to the indirect evidence X_3 and C_3 regulated by $\delta_{x_1x_3c_3}$.

Notice that the voters are *all* and the *only* schemata within the SL which carry X_1 as an element, so (6.49) incorporates *all possible immediate* information available from the *global* knowledge base; in other words, this represents a *first-order correction* which appeals to *immediate information* for correction, in reminiscences of calculus which treats only first-order incremental changes. Of course, the particular combination that makes up such correctives is intrinsic to the particular information updating implementation and so is in no way a necessity – it is one amongst probably many different manifestations of *Becoming*.

While the information bits X_i are determined by (6.36), the error-control bits are determined by equation (6.37) that can be shown to read essentially as

$$P^t(c_1) \propto \delta_{x_1x_2c_1}P^t(x_1)P^t(x_2)$$

in the *algebraic limit*, which is dominated by a single term related to the particular values of x_1 and x_2 decoded by (6.35) in a manner that *complies with* the code structure dictated by the delta-functions. The attainment of the other error-control bits are similar; so we arrive at the following conclusion:

Lemma 6.1 *In the algebraic limit, the decoded code-words belong to the class of the pre-defined codes; in other words, the decoding is structure-preserving.*

Of course, if we are not working in the *algebraic* limit, or in cases that the global structure is far from being *deterministic*, then the determination is a lot more subtle and we need to work out the detailed computations in (6.37).

Now, we have shown that the decoded results must be valid code-words, however, we still do not know whether such decoded results are the ones we want. For an extreme example, if we have an original code-word 001011 corrupted to 110011, we do not expect our decoder to retrieve the original code-word correctly since the corrupted one is again a valid code-word. This implies the existence of certain limit to the error patterns under which a tainted code-word is correctly decodable. To elucidate this matter, our first task is to go deeper into the nature of the iterative solutions. First, as is important in every iterative calculations, we need to know what the fixed points correspond to. In fact, we have

Lemma 6.2 *In the algebraic limit, all valid code-words are fixed points of the iteration dynamics.*

Proof:
The proof is trivial since by examining the decoding of each information bit we see that the ratios χ and r must favour the received information bit since the received code-word is a valid one; then by lemma 6.1 the same error-control bits are to be retrieved. Thus, as all bits are unchanged, such valid code-words are fixed points of the iteration dynamics.

Combining lemmas 6.1 and 6.2, we have the following theorem:

Theorem 6.1 *In the algebraic limit, the terminated solutions of error-control structure I are and only are valid code-words defined by the code.*

Theorem 6.1 is an important characterization of structure I as a mathematical object; by which further dig into the decoding behaviour proceed as follows.

A little bit more involved is the determination of the decoded code-words in relation to the received ones; however, it can be simplified a bit since by lemma 6.1 we know that the entire decoding can be reduced essentially to a discussion on the decoding of the information bits. According to the decoding formulae (6.49) for the information bits, it is appreciated that in the algebraic limit the decoding is practically *topological* rather than *algebraic*, since the algebraic endowment of the code simply cast itself into the number of correcting factors χ, and the decoding capability is really controlled by a competition of those χ present. To arrive at more illustrative results, we deliberately generalize (6.36) a bit and regard N_γ as the number of constituting χ factors embodied in the corrective γ (obviously, $N_\gamma = 2$ for error-control structure I); from (6.49), through direct counting we deduce at once the following lemma:

Lemma 6.3 *For error-control structure I in the algebraic limit, decoding of the information bits are governed by the following:*

1. *If the received information bit is in error, we can allow at most* $\lceil N_\gamma/2 \rceil - 1$ *incorrect factors of* χ *for a correct decoding of the concerned information bit; and*

2. *if the received information bit is not in error, we can allow at most* $\lceil N_\gamma/2 \rceil$ *incorrect factors of* χ *for a correct decoding of the concerned information bit.*

From lemma 6.3 we can deduce the maximum number of errors of a tainted code for it to be correctible by error-control structure I. Let us consider the worst condition: with reference to (6.49), suppose we have the maximum number of errors M_I occurring in an equation, we have two cases to consider. First, suppose the concerned information bit is erred, so r is mistaken and by the first statement of lemma 6.3 we can allow *at most* $\lceil N_\gamma/2 \rceil - 1$ incorrect factors of χ for a correct decoding of the concerned information bit; in the worst case, we have all remaining errors showing up in *different* instances of χ, so to ensure correct decoding we have

$$\begin{aligned} M - 1 &\leq \lceil N_\gamma/2 \rceil - 1 \\ M &\leq \lceil N_\gamma/2 \rceil \end{aligned} \tag{6.50}$$

Second, suppose the concerned information bit is received intact; by the second statement of lemma 6.3, we can allow *at most* $\lceil N_\gamma/2 \rceil$ incorrect factors of χ to ensure correct decoding of the concerned information bit; in the worst case, we have all errors showing up in *different* instances of χ, by which we have

$$M \leq \lceil N_\gamma/2 \rceil \tag{6.51}$$

Therefore, by both (6.50) and (6.51) we arrive at the following theorem:

Theorem 6.2 *(Sufficient condition for correct decoding:)*
For error-control structure I in the algebraic limit, the maximum number of errors M_I allowed for a tainted code-word for it remains to be correctible is

$$M_I \leq \lceil N_\gamma/2 \rceil \tag{6.52}$$

For example, in the example studied, we have $N_\gamma = 2$, so the maximum number of errors allowed is $M_I = 1$, explaining the observations in the computational experiments.

The decoding behaviour of structure I will serve as the prototype for other structures. In the forthcoming analysis, we shall see decoding formulae of similar types show up again by which we can do very effective comparison.

6.4 Error-control structures III & IV – exact solutions

As might be implied by the computational experiments for both structures III and IV, their behaviour are really similar. In fact, we shall derive in this section the analytic solutions of both, thereby the close ties between the two will become obvious.

6.4.1 Error-control structure III

Error-control structure III is nothing more than a series of iterations between two error-control structure I's, it is thus expected that solutions of both must be intimately related. In parallel with the discussions with error-control structure I, we shall first identify the underlying dynamical invariants and then derive the exact solutions.

6.4.1.1 Dynamical invariants for error-control structure III

As is evident from the decoding algorithm of error-control structure III, updating of schemata inside the structure layer, when viewed decoder by decoder, is essentially the same as the counterparts with structure I; the differences being that inputs to X_iY_i are the modified ones due to the other decoder, and that we are implementing only half of the whole decoding algorithm. Following the same discussion with structure I, it is not difficult to show that the same L – $L_{x_1x_1c_1}$, $L_{x_1x_2c_1'}$, etc. – are the constants of motion desired. By the way, it is easy to determine that their values are given by expressions like (6.1) and (6.2), with obvious modifications from c_i to c_i' when dealing with different decoders. As a partial summary, we have the constants of motion

$$\begin{aligned} L_{x_1x_2c_1} &= \frac{\mathcal{K}_{x_1x_2c_1}}{\Omega_{x_1x_2}} \\ L_{x_1x_2c_1'} &= \frac{\mathcal{K}_{x_1x_2c_1'}}{\Omega_{x_1x_2}} \end{aligned} \tag{6.53}$$

and so on.

6.4.1.2 Iteration dynamics

With knowledge of the iterative results of error-control structure I, that of error-control structure III are almost immediate. In the forthcoming discussion, we shall thus proceed not step by step, but *decoder* by *decoder*.

1. *count=0, decoder 1:*

   At this first stage of information updating, the input pieces of evidence are: $P^e_{X_iY_i}(x_iy_i)$ and $P^e_{C_iD_i}(c_id_i)$ with $i = 1, 2, 3$. Following the same iteration dynamics of error-control structure I up to step 3, we arrive at the local joint probabilities for the schemata within the structure layer given by (6.9), (6.11) and (6.13). To summarize, we have

   $$\begin{aligned} P^{**}_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\,\lambda_{x_1}\,\lambda_{x_2}\,L_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \\ P^{**}_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\eta_{x_2}\,\eta_{x_3}\,L_{x_2x_3c_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3) \\ P^{**}_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\zeta_{x_1}\,\zeta_{x_3}\,L_{x_1x_3c_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3) \end{aligned} \tag{6.54}$$

   where λ, η and ζ are given by (6.33). Further, the joint probabilities $P^{**}_{X_iY_i}(x_iy_i)$ are determined to be

   $$\begin{aligned} P^{**}_{X_1X_2}(x_1y_1) &= \alpha\,\gamma_{x_1}P^e_{X_1Y_1}(x_1y_1) \\ P^{**}_{X_2Y_2}(x_2y_2) &= \alpha\,\gamma_{x_2}P^e_{X_2Y_2}(x_2y_2) \\ P^{**}_{X_3Y_3}(x_3y_3) &= \alpha\,\gamma_{x_3}P^e_{X_3Y_3}(x_3y_3) \end{aligned} \tag{6.55}$$

   with $\gamma_{x_1} = \lambda_{x_1}\zeta_{x_1}$, $\gamma_{x_2} = \lambda_{x_2}\eta_{x_2}$, and $\gamma_{x_3} = \eta_{x_3}\zeta_{x_3}$.

2. *count=0, decoder 2:*

   After a partial decoding with decoder 1, *outputs* of decoder 1 become *inputs* to decoder 2, thus the pieces of information fed into decoder 2 are: $P^{*}_{X_iY_i}(x_iy_i) = \alpha\,\gamma_{x_i} P^{e}_{X_iY_i}(x_iy_i)$ [9] and $P^{e}_{C_i'D_i'}(c_i'd_i')$ for $i = 1, 2, 3$. Repeat similar decoding as decoder 1, we obtain

$$\begin{aligned} P^{**}_{X_1X_2C_1'}(x_1x_2c_1') &= \alpha\,\lambda'_{x_1}\,\lambda'_{x_2}\,L_{x_1x_2c_1'} P^{e}_{X_1Y_1}(x_1) P^{e}_{X_2Y_2}(x_2) \\ P^{**}_{X_2X_3C_2'}(x_2x_3c_2') &= \alpha\,\eta'_{x_2}\,\eta'_{x_3}\,L_{x_2x_3c_2'} P^{e}_{X_2Y_2}(x_2) P^{e}_{X_3Y_3}(x_3) \\ P^{**}_{X_1X_3C_3'}(x_1x_3c_3') &= \alpha\,\zeta'_{x_1}\,\zeta'_{x_3}\,L_{x_1x_3c_3'} P^{e}_{X_1Y_1}(x_1) P^{e}_{X_3Y_3}(x_3) \end{aligned} \tag{6.56}$$

where λ', η' and ζ' are defined as follows:

$$\begin{aligned} \lambda'_{x_1} &= \sum_{x_3c_3'} \mathcal{K}_{x_1x_3c_3'} \gamma_{x_3} p^{e}_{X_3Y_3}(x_3) \\ \lambda'_{x_2} &= \sum_{x_3c_2'} \mathcal{K}_{x_2x_3c_2'} \gamma_{x_3} p^{e}_{X_3Y_3}(x_3) \\ \eta'_{x_2} &= \sum_{x_1c_1'} \mathcal{K}_{x_1x_2c_1'} \gamma_{x_1} p^{e}_{X_1Y_1}(x_1) \\ \eta'_{x_3} &= \sum_{x_1c_3'} \mathcal{K}_{x_1x_3c_3'} \gamma_{x_1} P^{e}_{X_1Y_1}(x_1) \\ \zeta'_{x_1} &= \sum_{x_2c_1'} \mathcal{K}_{x_1x_2c_1'} \gamma_{x_2} p^{e}_{X_2Y_2}(x_2) \\ \zeta'_{x_3} &= \sum_{x_2c_2'} \mathcal{K}_{x_2x_3c_2'} \gamma_{x_2} P^{e}_{X_2Y_2}(x_2) \end{aligned} \tag{6.57}$$

Further, the joint probabilities $P^{**}_{X_iY_i}(x_iy_i)$ are determined to be

$$\begin{aligned} P^{**}_{X_1X_2}(x_1y_1) &= \alpha\,\gamma_{x_1}\,\gamma'_{x_1}\,P^{e}_{X_1Y_1}(x_1y_1) \\ P^{**}_{X_2Y_2}(x_2y_2) &= \alpha\,\gamma_{x_2}\,\gamma'_{x_2}\,P^{e}_{X_2Y_2}(x_2y_2) \\ P^{**}_{X_3Y_3}(x_3y_3) &= \alpha\,\gamma_{x_3}\,\gamma'_{x_3}\,P^{e}_{X_3Y_3}(x_3y_3) \end{aligned} \tag{6.58}$$

with $\gamma'_{x_1} = \lambda'_{x_1}\zeta'_{x_1}$, $\gamma'_{x_2} = \lambda'_{x_2}\eta'_{x_2}$, and $\gamma'_{x_3} = \eta'_{x_3}\zeta'_{x_3}$. What significant here is that we start to see a coupling of both decoders through $\gamma\,\gamma'$.

3. *count=1, decoder 1:*

   Next, information from decoder 2 is fed back to decoder 1; the fed in evidence

[9] Remind that we rename $P^{**}_{X_iY_i}(x_iy_i)$ to $P^{*}_{X_iY_i}(x_iy_i)$ on feeding information from decoder 1 to decoder 2.

reads: $P^*_{X_iY_i}(x_iy_i) = \alpha\,\gamma_{x_i}\gamma'_{x_i}P^e_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$; by which we get

$$\begin{aligned}
P^{**}_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma'_{x_1}\,\gamma'_{x_2}\,L_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^{**}_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma'_{x_2}\,\gamma'_{x_3}\,L_{x_2x_3c_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^{**}_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma'_{x_1}\,\gamma'_{x_3}\,L_{x_1x_3c_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)
\end{aligned} \tag{6.59}$$

However, it follows immediately from (6.59) that

$$\begin{aligned}
P^{**}_{X_1X_2X_1}(x_1) &= \alpha\,\gamma_{x_1}\,\gamma'_{x_1}\,P^e_{X_1Y_1}(x_1)\\
&= P^*_{X_1Y_1}(x_1)\\
P^{**}_{X_1X_2X_1}(x_2) &= \alpha\,\gamma_{x_2}\,\gamma'_{x_2}\,P^e_{X_2Y_2}(x_2)\\
&= P^*_{X_2Y_2}(x_2)
\end{aligned} \tag{6.60}$$

and similarly for the marginalized probabilities for other schemata; this signifies the termination of iterations with decoder 1.

4. *count=1, decoder 2:*

   With respect to decoder 2, its output evidence to decoder 1: $P^*_{X_iY_i}(x_iy_i) = \alpha\,\gamma_{x_i}\gamma'_{x_i}P^e_{X_iY_i}(x_iy_i)$ is not altered by decoder 1; so decoder 2 itself looks essentially the same as error-control structure I, apart from the exceptions of certain notational differences and the modified corrective factors $\gamma\,\gamma'$; as a result, it follows at once

$$\begin{aligned}
P^{**}_{X_1X_2C'_1}(x_1x_2c'_1) &= \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma'_{x_1}\,\gamma'_{x_2}\,L_{x_1x_2c'_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^{**}_{X_2X_3C'_2}(x_2x_3c'_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma'_{x_2}\,\gamma'_{x_3}\,L_{x_2x_3c'_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^{**}_{X_1X_3C'_3}(x_1x_3c'_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma'_{x_1}\,\gamma'_{x_3}\,L_{x_1x_3c'_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)
\end{aligned} \tag{6.61}$$

   which are the terminated solutions for the concerned schemata.

5. *Output results:*

   As a summary, the solutions of the iteration dynamics of error-control structure III read as follows:

   (a) First, we have the joint probabilities of the schemata in the structure layer:

$$P^t_{X_1X_2C_1}(x_1x_2c_1) = \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma'_{x_1}\,\gamma'_{x_2}\,L_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)$$

$$
\begin{aligned}
P^t_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma'_{x_2}\,\gamma'_{x_3}\,L_{x_2x_3c_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma'_{x_1}\,\gamma'_{x_3}\,L_{x_1x_3x_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_2C'_1}(x_1x_2c'_1) &= \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma'_{x_1}\,\gamma'_{x_2}\,L_{x_1x_2c'_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^t_{X_2X_3C'_2}(x_2x_3c'_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma'_{x_2}\,\gamma'_{x_3}\,L_{x_2x_3c'_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C'_3}(x_1x_3c'_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma'_{x_1}\,\gamma'_{x_3}\,L_{x_1x_3c'_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)
\end{aligned}
\tag{6.62}
$$

(b) From (6.62), we obtain the belief for the information bits, namely

$$P^t(x_i) = \alpha\,\gamma_{x_i}\,\gamma'_{x_i}\,P^e_{X_iY_i}(x_i) \tag{6.63}$$

for $i = 1, 2, 3$.

(c) Belief of the error-control bits can be similarly obtained by marginalization from (6.62), so for C_1 say, we have

$$
\begin{aligned}
P^t(c_1) &= \sum_{x_1,x_2} P^t_{X_1X_2C_1}(x_1x_2c_1)\\
&= \sum_{x_1,x_2} \alpha\,L_{x_1x_2c_1}P^t(x_1)P^t(x_2)
\end{aligned}
\tag{6.64}
$$

The counterparts for the other error-control bits can be derived in the same manner.

Thus, by the same technique with error-control structure I, we derive the exact solutions for the iteration dynamics of error-control structure III; compare (6.62) to (6.64) with (6.35) to (6.37) we see the obvious analogy. Moreover, it is immediately appreciated why this code is literally called a turbo-code: decoding is achieved not just by γ, but by γ' which incorporate γ again. The terminated solutions are *asymmetric* with respect to both decoders, which is no mystery but a direct result of the decoding dynamics being *asymmetrically* arranged; this is in contrast with error-control structure IV to be attacked next.

6.4.2 Error-control structure IV

Error-control structure IV is error-control structure III with dynamical updating symmetrically implemented with both decoders; it is thus expected that the corrective factors of the present structure would show up symmetrically, contrasted with the asymmetric one previously. Without stepping through the detailed calculations again, we simply remark that the formal solutions are basically the same as those of error-control structure III, i.e. we have

$$\begin{aligned}
P^t_{X_1X_2C_1}(x_1x_2c_1) &= \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma''_{x_1}\,\gamma''_{x_2}\,L_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^t_{X_2X_3C_2}(x_2x_3c_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma''_{x_2}\,\gamma''_{x_3}\,L_{x_2x_3c_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C_3}(x_1x_3c_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma''_{x_1}\,\gamma''_{x_3}\,L_{x_1x_3x_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_2C'_1}(x_1x_2c'_1) &= \alpha\,\gamma_{x_1}\,\gamma_{x_2}\,\gamma''_{x_1}\,\gamma''_{x_2}\,L_{x_1x_2c'_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^t_{X_2X_3C'_2}(x_2x_3c'_2) &= \alpha\,\gamma_{x_2}\,\gamma_{x_3}\,\gamma''_{x_2}\,\gamma''_{x_3}\,L_{x_2x_3c'_2}P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C'_3}(x_1x_3c'_3) &= \alpha\,\gamma_{x_1}\,\gamma_{x_3}\,\gamma''_{x_1}\,\gamma''_{x_3}\,L_{x_1x_3c'_3}P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)
\end{aligned}\tag{6.65}$$

with γ defined identically with those of structure III; the only difference exhibited now lies in the definitions of the correctives γ''; the formal definitions of γ'' are similar to those of error-control structure III:

$$\begin{aligned}
\gamma''_{x_1} &= \lambda''_{x_1}\zeta''_{x_1}\\
\gamma''_{x_2} &= \lambda''_{x_2}\eta''_{x_2}\\
\gamma''_{x_3} &= \eta''_{x_3}\zeta''_{x_3}
\end{aligned}\tag{6.66}$$

where λ'', η'' and ζ'' are given by

$$\begin{aligned}
\lambda''_{x_1} &= \sum_{x_3c'_3}\mathcal{K}_{x_1x_3c'_3}p^e_{X_3Y_3}(x_3)\\
\lambda''_{x_2} &= \sum_{x_3c'_2}\mathcal{K}_{x_2x_3c'_2}p^e_{X_3Y_3}(x_3)\\
\eta''_{x_2} &= \sum_{x_1c'_1}\mathcal{K}_{x_1x_2c'_1}p^e_{X_1Y_1}(x_1)\\
\eta''_{x_3} &= \sum_{x_1c'_3}\mathcal{K}_{x_1x_3c'_3}P^e_{X_1Y_1}(x_1)\\
\zeta''_{x_1} &= \sum_{x_2c'_1}\mathcal{K}_{x_1x_2c'_1}p^e_{X_2Y_2}(x_2)\\
\zeta''_{x_3} &= \sum_{x_2c'_2}\mathcal{K}_{x_2x_3c'_2}P^e_{X_2Y_2}(x_2)
\end{aligned}\tag{6.67}$$

Notice that γ'' are completely *free* of γ, in contrast with the intimate coupling exhibited by γ' for error-control structure III. So by incorporating a *symmetric scheduling* of information updating, the updated information exhibits a different manifestation: *symmetric and completely decoupled correctives.* As a summary of the decoding results, we have

1. The joint probabilities of the schemata in the structure layer are given by (6.65).

2. From (6.65), we obtain the belief for the information bits, namely

$$P^t(x_i) = \alpha\, \gamma_{x_i}\, \gamma''_{x_i}\, P^e_{X_iY_i}(x_i) \tag{6.68}$$

   for $i = 1, 2, 3$.

3. Belief of the error-control bits can similarly be obtained by marginalization from (6.65), so for C_1 we have

$$\begin{aligned} P^t(c_1) &= \sum_{x_1,x_2} P^t_{X_1X_2C_1}(x_1x_2c_1) \\ &= \sum_{x_1,x_2} \alpha\, L_{x_1x_2c_1} P^t(x_1)P^t(x_2) \end{aligned} \tag{6.69}$$

   The counterparts for the other error-control bits can be derived in the same manner.

6.4.3 Structure preserving property and the maximum a posteriori solutions

Parallel analysis with that of error-control structure I can be implemented with error-control structures III and IV; the main difference arises from the structural difference of correctives exhibited with different updating schemes.

First, as what we have done before, we cast both structures III and IV to the *deterministic limit*, then it can be verified by direct computations that we have generally all L turned into Kronecker's δ-functions representing the codes. Further, we assume work in the *algebraic limit*. By these two assumptions, it is trivial to verify that theorem 6.1 hold in the present context; in other words, *the decoding mechanisms of both structures III and IV are structure preserving and the fixed points of iteration are and only are valid code-words defined by the codes.*

Next, we proceed to the error correcting capability of the two structures. It suffices to check only decoding of the information bits which read

$$\begin{aligned} P^t_{III}(x_i) &= \alpha\gamma_{x_i}\gamma'_{x_i}P^e_{X_iY_i}(x_i) \\ P^t_{IV}(x_i) &= \alpha\gamma_{x_i}\gamma''_{x_i}P^e_{X_iY_i}(x_i) \end{aligned} \tag{6.70}$$

where we have introduced explicit subscripts $_{III}$ and $_{IV}$ to distinguish the respective solutions for error-control structures III and IV respectively. Compare (6.70) with (6.36), we find discrepancies arise out of the presence of S_2 which manifests in the additional correctives: γ' for structure III and γ'' for structure IV; amongst which we further appreciate deviations due to series vs parallel implementation of decoding dynamics with respect to decoders: γ' incorporate γ in the definitions while γ'' do not. Anyway, no matter how the correctives are coupled, we expect such additional correctives γ' and γ'' should help enhance the error-correcting capability of both structures. To appreciate why and how it is going to be the case, we derive from (6.70) the counterparts to (6.49) in the *deterministic limit*:

$$\begin{aligned} R_{x_1} &\equiv \frac{P^t(x_1=0)}{P^t(x_1=1)} \\ &= \chi_{x_1}^{X_1X_2C_1}\chi_{x_1}^{X_1X_3C_3}\chi_{x_1}^{X_1X_2C_1'}\chi_{x_1}^{X_1X_3C_3'}r_{x_1} \end{aligned} \tag{6.71}$$

for both structures. While the correctives due to error-control bits of S_1: $\chi_{x_1}^{X_1X_2C_1}$ and $\chi_{x_1}^{X_1X_3C_3}$, are given by (6.48) and (6.47); those due to S_2 are for structure III

$$\begin{aligned} \chi_{x_1}^{X_1X_2C_1'} &= \frac{\lambda'_{x_1=0}}{\lambda'_{x_1=1}} \\ &= \frac{\sum_{x_2c_1'}\delta_{x_1=0,x_2c_1'}P_I^t(x_2)P^e_{C_1'D_1'}(c_1')}{\sum_{x_2c_1'}\delta_{x_1=1,x_2c_1'}P_I^t(x_2)P^e_{C_1'D_1'}(c_1')} \\ \chi_{x_1}^{X_1X_3C_3'} &= \frac{\zeta'_{x_1=0}}{\zeta'_{x_1=1}} \\ &= \frac{\sum_{x_3c_3'}\delta_{x_1=0,x_3c_3'}P_I^t(x_3)P^e_{C_3'D_3'}(c_3')}{\sum_{x_3c_3'}\delta_{x_1=1,x_3c_3'}P_I^t(x_3)P^e_{C_3'D_3'}(c_3')} \end{aligned} \tag{6.72}$$

where we have introduced $P_I^t(x_2) \equiv \gamma_{x_2}P^e_{X_2Y_2}(x_2)$, etc., to emphasize that the product on the right amounts to the terminated solution due to error-control structure I; upon these substitutions, their immediate resemblances to $\chi_{x_1}^{X_1X_2C_1}$ and $\chi_{x_1}^{X_1X_3C_3}$ are recovered. Similarly, with structure IV we have

$$\begin{aligned} \chi_{x_1}^{X_1X_2C_1'} &= \frac{\lambda''_{x_1=0}}{\lambda''_{x_1=1}} \\ &= \frac{\sum_{x_2c_1'}\delta_{x_1=0,x_2c_1'}P^e_{X_2Y_2}(x_2)P^e_{C_1'D_1'}(c_1')}{\sum_{x_2c_1'}\delta_{x_1=1,x_2c_1'}P^e_{X_2Y_2}(x_2)P^e_{C_1'D_1'}(c_1')} \\ \chi_{x_1}^{X_1X_3C_3'} &= \frac{\zeta''_{x_1=0}}{\zeta''_{x_1=1}} \\ &= \frac{\sum_{x_3c_3'}\delta_{x_1=0,x_3c_3'}P^e_{X_3Y_3}(x_3)P^e_{C_3'D_3'}(c_3')}{\sum_{x_3c_3'}\delta_{x_1=1,x_3c_3'}P^e_{X_3Y_3}(x_3)P^e_{C_3'D_3'}(c_3')} \end{aligned} \tag{6.73}$$

which are in exact parallel with their S_1's counterparts. Similar though the decoding formulae are to those of structure I; however, we are on the alert that decoding for structure III and IV is *asymmetric* with respect to the information and the error-control bits: for instance, information on X_2 appears in both $\chi^{X_1X_2C_1}$ [10] and $\chi^{X_1X_2C_1'}$ while information on C_1 exhibits only in $\chi^{X_1X_2C_1}$; in fact, it will be interesting to observe later that it is precisely due to such asymmetry that the error correcting behaviour of the present structures varies **not** simply with the number of errors but with the *error patterns* showing up. This triggers the following discussion.

It is trivial to verify from (6.71) that all single-error codes can be corrected; so, we proceed to decoding with two-error codes. Regarding two-error tainted codes, it might be taken as a guess from theorem 6.2 that they might be decodable with the present structures; the actual picture is *close* to this estimate but not exactly, since factor due to *error patterns* is coming into play. In the following, without loss of generality we shall investigate the decoding of X_1 unless otherwise specified.

1. Suppose two information bits, say X_2 and X_3, are tainted, then by (6.72) and (6.73) all χ are erred and the resulting decoding of X_1 is incorrect.

2. Suppose two error-control bits of the same set, say C_1 and C_3, are corrupted; then we shall have two χ factors erred, namely $\chi^{X_1X_2C_1}$ and $\chi^{X_1X_3C_3}$, but then the remaining χ factors from S_2 together with r will be sufficient to retrieve the correct X_1.

3. Suppose two error-control bits belonging to different sets of error-control bits, say C_1 and C_3', are corrupted, then two χ factors will be erred, namely $\chi^{X_1X_2C_1}$ and $\chi^{X_1X_3C_3'}$; but then the remaining χ factors together with r will be sufficient to retrieve the correct X_1.

4. Suppose there are two errors, one in the information set say X_2, and the other in the error-control set say C_1 or C_3; we are going to observe the different error correcting capabilities resulted from this apparently insignificant combinations. We distinguish the two cases as follows:

   (a) Tainted bits: X_2 and C_1

   Notice that both tainted bits come up in $\chi^{X_1X_2C_1}$, this particular error

[10] We suppress the subscript x_1 when the context is clear.

combination render $\chi^{X_1X_2C_1}$ *correct* and leave only $\chi^{X_1X_2C_1'}$ erred, hence subsequent decoding of X_1 is correct.

(b) Tainted bits: X_2 and C_3
In this case, we shall have $\chi^{X_1X_2C_1}$, $\chi^{X_1X_3C_3}$ and $\chi^{X_1X_2C_1'}$ erred and the decoded X_1 is incorrect.

5. Lastly, apart from the common behaviour aforementioned, we want to point out an important difference between structures III and IV as have already suggested in section 5.7.1. In table 6.1 we contrast the decoding details of X_2 on the assumption that the bits X_1, together with C_2' are tainted; here the symbol Y stands for a correct χ factor while N an incorrect one. From the table, we see that the hinge to the problem lies on the $\chi^{X_2X_3C_2'}$,: differences in the make up of the terms due to different decoding dynamics explain the different decoding capabilities observed.

χ	$X_1X_2C_1$	$X_2X_3C_2$ or	$X_1X_2C_1'$	$X_2X_3C_2'$	r_2	decoding status
Structure III	N	Y	N	Y	Y	Yes
Structure IV	N	Y	N	N	Y	No

Table 6.1: A comparison on the decoding capability of error-control structures III and IV

By the above studies, we conclude that decoding behaviour of our proposed structures depends heavily on the combination of errors: the number of errors and the error-patterns associated! These theoretical deductions testifies to the observations in the previous computational experiments.

Error-control structures III and IV deploy a decoder by decoder decoding dynamics and the resulting solutions show up in a simple factor form, our next task is to go back to structure II which employs a decoding dynamics on the basis of individual sub-universes. It has been suggested in the computational experiments that the decoding capability of structure II is weaker than those of structures III and IV, in the next section we shall see why.

6.5 Error-control structure II – exact solutions

We have not been successful in pinning down any general dynamical constants with error-control structure II; however, we can still proceed analytically to derive the exact solutions.

6.5.1 Iteration dynamics

Suppose evidence has been absorbed to the schemata in the evidence layers, we have, as an initialization to the decoding dynamics, the following updated probabilities: $P^*_{X_iY_i}(x_iy_i) = P^e_{X_iY_i}(x_iy_i)$, $P^*_{C_iD_i}(c_id_i) = P^e_{C_iD_i}(c_id_i)$ and $P^*_{C'_iD'_i}(c'_id'_i) = P^e_{C'_iD'_i}(c'_id'_i)$ for $i = 1, 2, 3$. Further, although not conducing any dynamical constant, we define the following quantities in parallel with (6.1) and (6.2) for future convenience:

$$\mathcal{K}_{x_1x_2c_1} = P_{X_1X_2C_1}(c_1|x_1x_2)\frac{P^*_{C_1D_1}(c_1)}{P_{C_1D_1}(c_1)} \tag{6.74}$$

and based on which we further have

$$\begin{aligned}
\Omega^{X_1X_2C_1}_{x_1x_2} &= \sum_{c_1}\mathcal{K}_{x_1x_2c_1} \\
L'_{x_1x_2c_1} &= \frac{\mathcal{K}_{x_1x_2c_1}}{\left(\Omega^{X_1X_2C_1}_{x_1x_2}\right)^3} \\
L_{x_1x_2c_1} &= \frac{\mathcal{K}_{x_1x_2c_1}}{\Omega^{X_1X_2C_1}_{x_1x_2}}
\end{aligned} \tag{6.75}$$

Similar terms are defined with other schemata in the structure layer. We start from step 4 onwards:

The first iteration, *count* $= 0$

1. *Absorption of evidence II:*
   By (6.74), we reformulate (5.17) as

   $$P^*_{X_1X_2C_1}(x_1x_2c_1) = \alpha\mathcal{K}_{x_1x_2c_1}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \tag{6.76}$$

   and similarly for other schemata in the structure layer.

2. *Checking evidence against the code structure:*
   After an initial updating by the observed evidence, schemata inside the structure layer update one another; without going through the algebra, we simply state

   $$P^{**}_{X_1X_2C_1}(x_1x_2c_1) = \alpha L'_{x_1x_2c_1}\lambda_{x_1}\lambda''_{x_1}\lambda_{x_2}\lambda''_{x_2}P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \tag{6.77}$$

   where L' is given by (6.75) and λ and λ'' are given by (6.33) and (6.67) respectively. Similar expressions hold for other schemata.

3. *Output of evidence:*

   Next, information is fed back to the evidence layers; in parallel with previous discussion we focus on $P^{**}_{X_iY_i}(x_iy_i)$ only, for which we have

$$P^{**}_{X_iY_i}(x_iy_i) = \alpha\omega_{x_i}\left(\gamma_{x_i}\gamma''_{x_i}\right)^2 P^e_{X_iY_i}(x_iy_i) \tag{6.78}$$

   for $i = 1, 2, 3$; here, γ and γ'' are given by (6.34) and (6.66) respectively. Notice here a special factor ω_{x_i} shows up which is a highly entangled term; for instance, ω_{x_1} is given by

$$\omega_{x_1} = t^1_{x_1}t^2_{x_1}t^3_{x_1}t^4_{x_1} \tag{6.79}$$

   where

$$\begin{aligned}
t^1_{x_1} &= \sum_{x_2c_1} L'_{x_1x_2c_1}\lambda_{x_2}\lambda''_{x_2}\Omega^{X_1X_2C'_1}_{x_1x_2}P^e_{X_2Y_2}(x_2) \\
t^2_{x_1} &= \sum_{x_2c'_1} L'_{x_1x_2c'_1}\lambda_{x_2}\lambda''_{x_2}\Omega^{X_1X_2C_1}_{x_1x_2}P^e_{X_2Y_2}(x_2) \\
t^3_{x_1} &= \sum_{x_3c_3} L'_{x_1x_3c_3}\zeta_{x_3}\zeta''_{x_3}\Omega^{X_1X_3C'_3}_{x_1x_3}P^e_{X_3Y_3}(x_3) \\
t^4_{x_1} &= \sum_{x_3c'_3} L'_{x_1x_3c'_3}\zeta_{x_3}\zeta''_{x_3}\Omega^{X_1X_3C_3}_{x_1x_3}P^e_{X_3Y_3}(x_3)
\end{aligned} \tag{6.80}$$

   The term ω_{x_1}, together with ω_{x_2} and ω_{x_3} derived similarly, share the same characteristic structure: it incorporates contributions from both decoders in a complicated manner. Compared with structures III and IV, we see that such ω are the peculiar contributions due to the global dynamics of structure II, which allows updating amongst *all* elements in the structure layer.

4. *Renaming items:* To proceed with further iterations, we rename items:

   (a) $P^{**}_{X_iY_i}(x_iy_i) \to P^{*}_{X_iY_i}(x_iy_i)$ for $i = 1, 2, 3$.

   (b) $P^{**}_{X_1X_2C_1}(x_1x_2c_1) \to P_{X_1X_2c_1}(x_1x_2c_1)$, $P^{**}_{X_1X_2C'_1}(x_1x_2c'_1) \to P_{X_1X_2c'_1}(x_1x_2c'_1)$, and similary for other schemata in the structure layer.

This concludes the first iteration, next we proceed to the second.

The second iteration, *count* $= 1$

5. *Absorption of evidence II :*
Upon feedback of information which has been *mediated* by the schemata X_iY_i, we obtain

$$P^*_{X_1X_2C_1}(x_1x_2c_1) = \alpha L_{x_1x_2c_1}\omega_{x_1}\omega_{x_2}\left(\gamma_{x_1}\gamma''_{x_1}\gamma_{x_2}\gamma''_{x_2}\right)^2 P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2) \tag{6.81}$$

and obvious counterparts to other schemata. Heed that (6.81) carries L instead of L'; without showing further details, we just remark that terms like (6.81) are the terminated solutions for error-control structure II. The resemblances with the counterparts of the previously studied structures are immediate.

6. *Output results:*
Let us summarize the results for error-control structure II: by incorporating (6.33) and (6.34), together with (6.66) and (6.67), we have the following:

(a) For the schemata in the structure layer, we have

$$\begin{aligned}
P^t_{X_1X_2C_1}(x_1x_2c_1) &= \alpha L_{x_1x_2c_1}\omega_{x_1}\omega_{x_2}\left(\gamma_{x_1}\gamma''_{x_1}\gamma_{x_2}\gamma''_{x_2}\right)^2 P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^t_{X_2X_3C_2}(x_2x_3c_2) &= \alpha L_{x_2x_3c_2}\omega_{x_2}\omega_{x_3}\left(\gamma_{x_2}\gamma''_{x_2}\gamma_{x_3}\gamma''_{x_3}\right)^2 P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C_3}(x_1x_3c_3) &= \alpha L_{x_1x_3c_3}\omega_{x_1}\omega_{x_3}\left(\gamma_{x_1}\gamma''_{x_1}\gamma_{x_3}\gamma''_{x_3}\right)^2 P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_2C'_1}(x_1x_2c'_1) &= \alpha L_{x_1x_2c'_1}\omega_{x_1}\omega_{x_2}\left(\gamma_{x_1}\gamma''_{x_1}\gamma_{x_2}\gamma''_{x_2}\right)^2 P^e_{X_1Y_1}(x_1)P^e_{X_2Y_2}(x_2)\\
P^t_{X_2X_3C'_2}(x_2x_3c'_2) &= \alpha L_{x_2x_3c'_2}\omega_{x_2}\omega_{x_3}\left(\gamma_{x_2}\gamma''_{x_2}\gamma_{x_3}\gamma''_{x_3}\right)^2 P^e_{X_2Y_2}(x_2)P^e_{X_3Y_3}(x_3)\\
P^t_{X_1X_3C'_3}(x_1x_3c'_3) &= \alpha L_{x_1x_3c'_3}\omega_{x_1}\omega_{x_3}\left(\gamma_{x_1}\gamma''_{x_1}\gamma_{x_3}\gamma''_{x_3}\right)^2 P^e_{X_1Y_1}(x_1)P^e_{X_3Y_3}(x_3)
\end{aligned} \tag{6.82}$$

(b) For individual information bits, we have

$$P^t(x_i) = \alpha\omega_{x_i}\left(\gamma_{x_i}\gamma''_{x_i}\right)^2 P^e_{X_iY_i}(x_i) \tag{6.83}$$

for $i = 1, 2, 3$.

(c) Belief of the error-control bits is obtained by marginalizing the joint probabilities of schemata in the structure layer; from which we obtain for C_1 say,

$$P^t(c_1) = \sum_{x_1x_2} P^t_{X_1X_2C_1}(x_1x_2c_1)$$

$$= \sum_{x_1 x_2} \alpha L_{x_1 x_2 c_1} P^t(x_1) P^t(x_2) \tag{6.84}$$

and similarly for the other error-control bits.

6.5.2 Structure preserving property and the maximum a posteriori solutions

In parallel with previous analysis, we let L approach the *deterministic limit* so that $L = \delta$ where the Kronecker's delta-functions are defined with respect to the code assignment enlisted in (5.5) and (5.15). By further assuming work in the *algebraic* limit, it is an easy task to verify that theorem 6.1 holds well in the present context, and we reach our first conclusion that *all fixed points of the iteration dynamics are and only are valid code-words permitted by the code.*

We shall not drill into the nature of solutions of structure II but shall simply take heed to the intrinsic peculiarities. Compare structure II with both structures III and IV; while all three involve two sets of error-control bits, structure II decodes with regard to all sub-universes as a totality *without* differentiating individual decoders, which is however the basis according to which the decoding dynamics underlying structures III and IV are constructed. The resulting differences show up in their respective terminated solutions studied before. To investigate say the decoded information bits, we contrast amongst (6.63), (6.68) and (6.83), the peculiarities of structure II manifest in the higher-power correctives $(\gamma\gamma')^2$ and the connection terms ω. The discrepancy between $(\gamma\gamma')^2$ and $(\gamma\gamma')$ is significant in that it *boosts* the effect of $(\gamma\gamma')$ over that of r. On the other hand, ω stand for higher order *indirect-evidence* terms; for instance, in (6.80) the λ and ζ kick in information from $X_2X_3C_2$ and $X_2X_3C_2''$ which are **not** immediate neighbours of X_1. We shall not attempt at undergoing the tedious deductions here, but simply remark that the results conduced from such peculiarities reflect in the computational experiments studied last chapter: in terms of the qualitative behaviour they resemble those of structure IV, however, the quantitative deviations are vast; actually we experimented that decoding of two-error codes fails if we increase err to say 0.1.

6.6 A comparison on the four error-control structures

Although the four error-control structures are proposed separately, they are nonetheless intimately related: First, we propose error-control structure I which serves as the prototype; then we propose structures II to IV, all of which are constructed on the basis of structure I with the addition of one more set of error-control bits. Further, we differentiate structures II to IV by rendering each of them distinct decoding dynamics: structure II employs a naive adaptation of that of structure I, and it decodes simply with regard to the sub-universes as a whole without considering a higher decoder-by-decoder basis; structure III and IV both take regard of individual decoders, but structure III employs an asymmetric global scheduling of information flow while structure IV a symmetric one. It is first suggested by computational experiments, and later explicated by exact analytical treatments, that different decoding capabilities are pertinent to each structure. As we know exactly how the decoding mechanism of each structure works, we are entitled to ask the following question: what is the probability that a decoded code-word is correct[11]? This is obviously an important question in all practical applications and we study case by case.

Prior to commencing the algebra, recall the following six cases in decoding two-error codes:

1. both errors occur in S_0;
2. both errors occur in S_1;
3. both errors occur in S_2;
4. one error in S_0 and the other in S_1;
5. one error in S_0 and the other in S_2;
6. one error in S_1 and the other in S_2;

Let us denote the probability that a decoded code-word be correct by P_{cor}, and the probability that a bit be correctly transmitted by $t = 1 - err$; then for

[11]Notice that the probability here involves the combinatorial properties of error patterns; it has nothing immediate to do with the probabilities invoked in constructing the probabilistic reasoning networks.

1. Structure I:

   Decoding succeed if there is at most one error, so

$$\begin{aligned} P^I_{cor}(t) &= t^6 + {}_6C_1 t^5 (1-t) \\ &= t^5 (6-5t) \\ P^I_{cor}(0.9) &= 0.886 \end{aligned} \tag{6.85}$$

2. Structure II:

   Decoding succeed if there are at most two errors, with the exception of cases (1), (4) and (5), so

$$\begin{aligned} P^{II}_{cor}(t) &= t^9 + {}_9C_8 t^8 (1-t) + 2\,{}_3C_1 t (1-t)^2 t^6 + \left[{}_3C_2 t^2 (1-t)\right]^2 t^3 \\ &= t^7 \left(7t^2 - 21t + 15\right) \\ P^{II}_{cor}(0.9) &= 0.847 \end{aligned} \tag{6.86}$$

3. Structure IV:

   It is similar to that of structure II but with some instances out of cases (4) and (5) succeeded. We have

$$\begin{aligned} P^{IV}_{cor}(t) &= P^{II}_{cor} + {}_3C_2 t^2 (1-t)\ {}_2C_1 t (1-t)\, t^4 + {}_3C_2 t^2 (1-t)\ {}_2C_1 t (1-t)\, t^4 \\ &= t^7 \left(19t^2 - 45t + 27\right) \\ P^{IV}_{cor}(0.9) &= 0.904 \end{aligned} \tag{6.87}$$

4. Structure III:

   It is similar to that of structure IV but with all instances of case (5) succeeded. We have thus

$$\begin{aligned} P^{IV}_{cor}(t) &= P^{II}_{cor} + {}_3C_2 t^2 (1-t)\ {}_2C_1 t (1-t)\, t^4 + {}_3C_2 t^2 (1-t)\ {}_3C_2 t (1-t)\, t^3 \\ &= t^7 \left(22t^2 - 51t + 30\right) \\ P^{II}_{cor}(0.9) &= 0.918 \end{aligned} \tag{6.88}$$

By comparing (6.85) to (6.88), we see the appreciable differences at first due to the addition of error-control bits, and then due to different implementations of decoding algorithms.

6.7 Conclusion

In this chapter, we cast the computational results obtained in chapter 5 to a solid footing by actually computing the iteration results explicitly and analytically. By so doing, we can assess exactly how various decoding dynamics give rise to different decoding capabilities. Apart from the particular results concerning individual structures, we emphasize again our underlying theme of *cyclic probabilistic reasoning networks*; these coding examples serve us the first vindication towards the insisted *bottom-up* approach on constructing the information structures. By the way, we can appreciate differences due to coupling of global topology of the SLs and the imposed decoding dynamics; this serves another testimony to the theme of complexity: couplings between local dynamics and global constraints give rise to various possibilities. In particular, with these examples we are blessed to come up with something easily interpretable and useful.

Chapter 7
Conclusion

7.1 Our thesis

The problem of reasoning is undoubtedly one of the most fundamental themes of human civilization; it is a matter of concern in virtually all aspects of life. Precisely due to its ubiquitous occurrence, it has been impossible to accommodate the concern in a single paradigm. While there are attempts aspire at reducing all questions to a single, unique and universal *Being*, however, as long as we are attacking the problems on the basis of *experience*, and expressing our thoughts and desires through various humanly *devised* languages, we are actually in a world of *Becoming* which admits varieties and differences. It is under this spirit of *Becoming* that we undertake analysis in this project.

Amongst all possible methodologies in the realm of *Becoming* which are directed specifically to the reasoning problem, we attacked in particular the probability based reasoning networks which form one of the contemporary approaches of artificial intelligence. Probability is fundamentally a devised mathematical concept to *represent* the human concern of *likelihood*; however, the very concept of likelihood plays in so diverse contexts that probability is rendered a status far removed from being concretely and uniquely defined; this is in contrast with other human conceptions like length which can be settled by a calibrated ruler together with an agreement on simultaneous measurement. Under such circumstances, any proper use of probability over situations which embrace a variety of situations is usually intriguing. To circumvent various pitfalls lurking behind the intricacy, our use of probability is confined to descriptions on *local* measurements and *local* beliefs studied within an *immediately* related set of random variables, and render *global* behaviour treatments in the form of dynamical information updating amongst the constituting pieces of local information. This picture is natural since human reasoning is indeed based on first knowledge on closely related phenomena, and then an establishment of a global knowledge base by conjoining local pieces of information consistently at their common elements; this is what we call the *bottom-up* process. On this basis, our intended probability based reasoning approach is more properly deemed the *bottom-up probability reasoning scheme.*

Following the *bottom-up probability reasoning scheme* is our first set of concrete examples on a series of error-control structures. It has been found in previous years that probability based reasoning networks (classical Bayesian networks primarily) play an efficient role to deal with error-control problems; however, the iterative character underlying the decoding dynamics often render further analysis difficult, if not impossible. Our models are constructed on a different footing from the classical models, and by resorting to the modern theme of complexity we devise a set of decoding dynamics pertinent to our problems. The resulted beautiful functioning of all these structures enabled us new and simple insights to problems unnoticed before. Therefore, in terms of the *verifiability* perspective of justifying the value of the *bottom-up probability reasoning* approach, it is not exaggerated to claim that a pretty good beginning is at hand.

7.2 Hind-sights and foresights

An important methodological shift underlying our proposed reasoning schemes is to attack problems in a hierarchical order: instead of focusing on individual events as in the BN, we shift our view first to clusters of events, and then to global arrangement of such clusters. This is reminiscent of how physical theories are constructed to deal with natural phenomena: though we believe that our world is fundamentally made up of atoms, electrons and so on, we never deal with say the atmosphere atoms by atoms; instead we invoke averaged quantities like pressure, density and temperature which view a whole collection of air particles. Our treatment of probability in this project is unconventional, but it is natural. On the other hand, to take the view of a BN is like dissecting the world into two sets: skeleton (arrows) and flesh (nodes); our schemes, in the contrary, have treated the two just as different ingredients of an organism without dissection. This is again reminiscent of how the theory of relativity and quantum mechanics have shifted our world-views from classical Newtonian mechanical universe: while in Newton's world we have the observer, the observed, space and time separated without direct correlation (other than possible functional roles in between), Einstein tells us that space and time are really in one piece – the spacetime, later quantum mechanics reveals the correlation between the observer and the observed as expressed in say the Heisenberg's interpretation of the uncertainty principle.

The examples presented in this thesis is nothing but a tiny facet that we need to concern. There are undoubtedly many interesting problems to follow, to name a few:

1. With regard to the "turbo"-type code structures we attacked, what will happen after incorporating a random interleaver? This might help in explaining the widely verified, but poorly understood near-Shannon performances of the codes?

2. With regard to the theme on *bottom-up probability reasoning scheme*, can we construct other meaningful examples to study; in particular, what are the general relationships between global topology and updating dynamics?

3. The whole approach we have set forth is very general and so is not limited to error-control applications, what are other directions to attack? By the way, we should be cautioned that every step of the reasoning dynamics proposed is subject to change: say the Jeffrey's rule of updating is no more than a convention and there are indeed situations which do not respect it; under such circumstances, what can we do to suitably adapt the structures?

7.3 Concluding remark

As with every newly ventured direction, ours is no exception from being fragile, limited and immature. It is the hope that the effort presented here will be able to give us some ideas on some meaningful directions to go in probabilistic reasoning problems. Lastly, as our approach is essentially one particular manifestation of all possible *Becoming* towards the reasoning problem, it is no doubt the greatest query to know its relationship to other instances sharing the same theme of concern. Perhaps in a complicated world as what we are having today, it may be no exaggeration to conjecture that it might hold the key to the well-*Being* to have very careful *communication* betweem various sorts of *Becoming*!

Appendix A
An alternative derivation of the local updating formula

The local updating formula discussed in 3.4.2 was derived locally. In this appendix we study an example on an alternative derivation of it for an acyclic SL based on the associated global probability, which is the original proposal in Lauritzen and Spiegelhalter [5]. As explained in chapter 4, *cyclic* SLs lack a global probability and so this derivation will not be adequate for those cases.

Suppose we have the information structure depicted in Fig:3.6 for which the global probability reads

$$\begin{aligned} P(abcdefg) &= \frac{P(abcdef)P(fg)}{P(f)} && \text{(A.1)} \\ &= \frac{P(abcde)P(bcf)P(fg)}{P(bc)P(f)} && \text{(A.2)} \\ &= \frac{P(abce)P(abd)P(bcf)P(fg)}{P(ab)P(bc)P(f)} && \text{(A.3)} \end{aligned}$$

on successively decomposing the structures according to the AJD conditions present. Next, suppose information on G is updated from $P(g)$ to $P^*(g)$; we would like to know how individual schemata update themselves. First, from (A.1) we know how the joint probability $P(abcdef)$ updates itself, namely

$$\begin{aligned} P^*(abcdef) &= \sum_g P^*(abcdefg) \\ &= \sum_g P(abcdefg)\frac{P^*(g)}{P(g)} \\ &= \sum_g \frac{P(abcdef)P(fg)}{P(f)}\frac{P^*(g)}{P(g)} \\ &= \frac{P(abcdef)P^*(f)}{P(f)} && \text{(A.4)} \end{aligned}$$

where the second equality follows by the usual *absorption map* and the third equality by (A.1); besides, we have introduced $P^*(f) \equiv \sum_g P^*(fg) = \sum_g P(fg)P^*(g)/P(g)$

to denote the updated information to F. Equation (A.4) has the peculiar property that it is essentially just an *absorption map* to the sub-universe $ABCDEF$ by the updated information of F, which is in turn induced by the *absorption map* due to G. From (A.4) it is trivial to verify that

$$\begin{aligned} P^*(bcf) &= \sum_{ade} P^*(abcdef) \\ &= \sum_{ade} \frac{P(abcdef)P^*(f)}{P(f)} \\ &= P(bcf)\frac{P^*(f)}{P(f)} \end{aligned} \tag{A.5}$$

which suggests that local updating to schema BCF *alone* is again simply some *absorption map* by some suitable separator (F in this case). In general, it is straight forward to see that all local schemata are updated by similar *absorption maps*. By this example, we see that for *acyclic* SLs, global updating is simply equivalent to a finite series of local updating; this equivalence is however generally not true for *cyclic* SLs.

Bibliography

[1] Jensen F. V., "*An introduction to Bayesian networks*", UCL Press, 1996.

[2] Kant I., "*Critique of Pure Reason*", translated by W. S. Pluhar, Hackett Publishing Company, Inc.

[3] Khrennikov A., "*Interpretations of Probability*", VSP Press, 1999.

[4] Kschischang R. and Frey J., "*Iterative Decoding of Compound Codes by Probability Propagation in Graphical Models*", IEEE Journal on Selected Areas in Communications, vol. 16, No. 2, Feb. 1998, P.219-230.

[5] Lauritzen S. L. and Spiegelhalter D. J., "*Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems*", J. R. Statist. Soc. B, 50, No.2, 1988, P.157-224.

[6] Lee T. T., "*An information-theoretic analysis of relational databases – Part I: data dependencies and information metric*", IEEE Trans. on Software Engineering, vol. SE-13, No. 10, Oct. 1987, P.1049 - 1060.

[7] Lee T. T., "*An information-theoretic analysis of relational databases – Part II: information structures of database schemas*", IEEE Trans. on Software Engineering, vol. SE-13, No. 10, Oct. 1987, P.1062 - 1071.

[8] McEliece R. J., Mackay J. C., Cheng J. F., "*Turbo Decoding as an Instance of Pearl's "Belief Propagation" Algorithm*", IEEE Journal on Selected Areas in Communications, vol. 16, No. 2, Feb. 1998, P.140-152.

[9] Pearl J., "*Probabilistic reasoning in intelligent systems: networks of plausible inference*", Morgan Kaufmann Publishers, 1988.

[10] Richardson T., "*The Geometry of Turbo-Decoding Dynamics*", IEEE Trans. on Information Theory, vol 46, No. 1, Jan. 2000, P.9-23.

[11] Vorob'ev N. N., "*Consistent families of measures and their extensions*", Theory of Probability and Applications, 7, 147-163, 1962.