Aus dem Departement für Informatik
Universität Freiburg (Schweiz)

Algebras of Gaussian Linear Information

Inaugural–Dissertation

zur Erlangung der Würde eines *Doctor scientiarum informaticarum*
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Freiburg in der Schweiz

vorgelegt von

Christian Eichenberger

aus

Lenzburg & Beinwil am See

Nr. 1640
UniPrint, Freiburg
2009

Von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Freiburg in der Schweiz angenommen, auf Antrag von

- Prof. Dr. Ulrich Ultes-Nitsche, Universität Freiburg, Schweiz (Jurypräsident)
- Prof. Dr. Jürg Kohlas, Universität Freiburg, Schweiz
- Prof. Dr. François Bavaud, Université de Lausanne, Suisse
- Dr. Paul-André Monney, Ely, USA

Freiburg, 23.06.2009

Der Leiter der Doktorarbeit: Prof. Dr. Jürg Kohlas

Der Dekan: Prof. Dr. Titus Jenny

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

Sir Arthur Conan Doyle, The Adventures of Sherlock Holmes

Acknowledgements

Before diving into the Gaussian world, I would like to thank everyone who has contributed to this thesis or has supported me morally during the work.

First, I would like to mention my collaborators at the Theoretical Computer Science group at the University of Fribourg (in alphabetical order): Jutta Langel, Norbert Lehmann (who adapted Abel for Gaussian linear models), Marc Pouly and Cesar Schneuwly (who proofread early drafts of parts of this thesis).

In particular, I am grateful to Prof. Jürg Kohlas, the director of this thesis, whose lectures and seminars initiated me to theoretical computer science and to algebraic information theory in particular. He also motivated me to work on the subject of Gaussian hints and has given the time and the freedom to work on the subject ever since. With his broad knowledge, he has also given invaluable input to the development and the final presentation of this thesis.

Furthermore, the referees of this thesis merit being acknowledged: Prof. François Bavaud and Dr. Paul-André Monney, who both pointed out lots of details which helped to improve the text at its final stage.

Also, I am greatly indepted to Christian Gebhard and Renzo Caduff for linguistic advice and proofreading (English and Romansh, respectively).

Last but not least, thanks to family, friends and tutors for everything beyond Gaussian hints: Ruth, Andrea, Ursina, Beat, Christian, Anna, Sara, Thomas, Jutta, Tricia, Guy.

Abstract (English)

Gaussian linear information arises in many real-world models of the natural and social sciences. The Gaussian distribution has turned out to appropriately represent uncertainty in many linear models. The main goal of this thesis is to describe and to compare different algebras of Gaussian linear information: Corresponding elements and operations in the various algebras are revealed and the respective computational advantages are highlighted.

In order to make large models computationally tractable, they have to be decomposed into independent factors by exploiting sparsity. For such factorisations, valuation algebras provide a general, abstract framework for local computations. A valuation algebra is a two-sorted algebra with three operations: valuations (which may be seen as pieces of information) refer to a domain of interest; valuations can be marginalised (focussed) to a domain of interest, and they may be combined (aggregated). Generic message-passing schemes can be used to answer projection problems. Many problems in applications can be reduced to a projection problem: diagnostic estimation, prediction, filtering, smoothing. For instance, Gaussian densities form a valuation algebra: marginalisation is integration, and combination is multiplication (plus renormalisation). Gaussian densities may be represented by Gaussian potentials or moment matrices, using either the concentration or the variance-covariance matrix, respectively. Here, marginalisation and combination are matrix operations.

A conditional Gaussian density is the family of Gaussian densities obtained on the head variables by fixing a value for the tail variables. A conditional Gaussian density corresponds to a Gaussian density on the head variables plus a linear regression on the tail variables. Conditional Gaussian densities can be analysed in three ways: geometric, algebraic and analytic.

- General Gaussian linear systems lead to Gaussian hints by assumption-based inference. Gaussian hints have focal sets which are parallel linear manifolds of the same dimension in the parameter space. Combination corresponds to intersection of focal sets and marginalisation to projection of focal sets. Gaussian potentials correspond to Gaussian hints whose focal sets are all singletons.

- Gaussian potentials can be extended to a valuation algebra of quotients which are represented by pairs of Gaussian potentials. Conditional Gaussian densities can be represented in the so-called separative extension of Gaussian potentials.

- Since a conditional Gaussian density is a quotient function of two Gaussian densities, the concentration matrix in the exponent of the denominator can be subtracted from the concentration matrix in the exponent of the numerator. This leads to a new representation of symmetric Gaussian potentials whose pseudo-concentration matrix is only symmetric but not necessarily positive definite.

The main result of these considerations is that different conditional Gaussian densities turn out to be linked to the same Gaussian hints (up to equivalence) if and only if the conditional Gaussian densities are equal up to a constant factor. In other

words, Gaussian likelihood functions bear the full information contained in Gaussian hints. This explains why assumption-based reasoning on (over-)determined Gaussian linear systems reproduces the estimation results based on the maximum-likelihood principle.

Variables may be linear combinations of other variables. This imposes linear restrictions on the parameter space. In the spirit of assumption-based reasoning, algorithms for inference, the combination and marginalisation are derived for symmetric Gaussian potentials with deterministic equations.

Finally, it is shown how Gaussian linear systems can be expressed in the language ABEL. Queries on a complex Gaussian linear system can be answered in the ABEL system. Several examples illustrate the new approach of symmetric Gaussian potentials.

Zusammenfassung (deutsch)

Gauss'sche lineare Information kommt in verschiedenen Modellen der reellen Welt vor, sowohl in den natur- wie auch in den sozialwissenschaftlichen Disziplinen. Mit der Gauss-Verteilung kann Unsicherheit oft adäquat dargestellt werden. Das Ziel dieser Dissertation ist es, verschiedene Algebren Gauss'scher Information zu beschreiben und zu vergleichen: Die einander entsprechenden Elemente und Operationen in den unterschiedlichen Algebren sollen herausgearbeitet und ihre jeweiligen rechentechnischen Vorteile hervorgehoben werden.

Damit grosse Modelle computertechnisch behandelt werden können, müssen sie in unabhängige Faktoren zerlegt werden. Dies ist möglich, falls die Modelle dünnbesetzt sind. Valuationsalgebren bieten ein allgemeines abstraktes Framework für lokales Rechnen mit solchen Faktorisierungen. Eine Valuationsalgebra ist eine zweisortige Struktur mit drei Operationen: Valuationen (die als Informationsstücke angesehen werden können) beziehen sich auf eine Domäne; Valuationen können auf eine Domäne marginalisiert (fokussiert) und kombiniert (aggregiert) werden. Generische Algorithmen mit Nachrichtenaustausch können angewendet werden, um ein Projektionsproblem zu lösen. Viele Anwendungsprobleme können auf ein Projektionsproblem zurückgeführt werden: diagnostische, prädiktive, Filter- und Smoothing-Probleme. Zum Beispiel bilden Gauss'sche Dichten eine Valuationsalgebra: Marginalisierung ist Integration und Kombination ist Multiplikation (plus Normalisierung). Gauss'sche Dichten können durch Gauss'sche Potentiale oder Moment-Matrizen dargestellt werden, wobei entweder die Konzentrationsmatrix oder die Varianz-Kovarianzmatrix verwendet wird. Hier sind Marginalisierung und Kombination Matrizenoperationen.

Eine bedingte Gauss'sche Dichte ist die Familie von Gauss'sche Dichten über die Kopfvariabeln für einen jeweils festen Wert der Rumpfvariabeln. Eine bedingte Gauss'sche Dichte entspricht einer Gauss'schen Dichte über die Kopfvariabeln mit linearer Regression von den Rumpfvariabeln. Bedingte Gauss'sche Dichten können auf drei Arten betrachtet werden: auf geometrische, algebraische und analytische.

- Allgemeine Gauss'sche lineare System führen zu Gauss'schen Hinweisen durch annahmen-basiertes Schliessen in Gauss'schen linearen Systemen. Die Fokalmengen Gauss'scher Hinweise sind parallele lineare Mannigfaltigkeiten derselben Dimension im Parameterraum. Gauss'sche Hinweise werden kombiniert, indem ihre Fokalmengen geschnitten werden, und marginalisert, indem ihre Fokalmengen projiziert werden. Gauss'sche Potentiale entsprechen Gauss'schen Hinweisen mit einelementigen Fokalmengen.

- Gauss'sche Potentiale können zu einer Valuationsalgebra von Quotienten erweitert werden, die durch Paare von Gauss'schen Potentialen repräsentiert werden. In dieser sogenannten separativen Erweiterung können auch bedingte Gauss'sche Dichten dargestellt werden.

- Da eine bedingte Gauss'sche Dichte eine Quotientenfunktion zweier Gauss'scher Dichten ist, entspricht diese der Subtraktion zweier Konzentrationsmatrizen.

> Dies führt zu symmetrischen Gauss'schen Potentialen, deren Pseudo-Konzentrationsmatrix nur symmetrisch, aber nicht notwendigerweise positiv definit ist.

Aus diesen Betrachtungen ergibt sich, dass bedingte Gauss'sche Dichten genau dann dem (bis auf Äquivalenz) gleichen Gauss'schen Hinweis entsprechen, falls die bedingten Gauss'schen Dichten bis auf einen konstanten Faktor gleich sind. In anderen Worten tragen Gauss'sche Likelihood-Funktionen dieselbe Information wie Gauss'sche Hinweise. Dies erklärt, wieso annahmen-basiertes Schliessen aus (über-)bestimmten Gauss'schen linearen Systemen zu denselben Schätzern führt wie die Maximum-Likelihood-Methode.

Variabeln können lineare Kombinationen anderer Variabeln sein. Dies erlegt lineare Einschränkungen auf den Parameterraum. Durch annahmen-basiertes Schliessen werden Algorithmen für Inferenz, Kombination und Marginalisierung für symmetrische Gauss'sche Potentiale mit linearen Gleichungen hergeleitet.

Schliesslich wird gezeigt, wie Gauss'sche lineare System in der Sprache ABEL ausgedrückt werden können. Anfragen über komplexe Gauss'sche lineare Systeme können durch ABEL beantwortet werden. Symmetrische Gauss'sche Potentiale werden anhand mehrerer Beispiele illustriert.

Abstract (romontsch)

Informaziun Gaussian-lineara vegn avon en biars models dil mund real dallas scienzias naturalas e socialas. La distribuziun Gaussiana ei semussada adequata per representar incertezia en numerus models linears. La finamira principala da questa dissertaziun eis ei da descriver e cumparegliar differentas algebras d'informaziun Gaussian-lineara: Ils elements e las operaziuns che corrispundan in a l'auter ellas diversas algebras vegnan fatgs resortir e lur avantatgs respectivs per las calculaziuns vegnan mess en evidenza.

Per tractar models gronds cul computer, eis ei necessari da decumponer quels en facturs independents. Quei ei pusseivel, sch'ils models ein spargliai. Algebras da valuaziun porschan in rom d'applicaziuns general ed abstract per la calculaziun locala cun da quellas factorisaziuns. In'algebra da valuaziuns ei ina structura en duas specias cun treis operaziuns: valuaziuns (che san gnir interpretadas sco tocs d'informaziun) serefereschan ad ina domena particulara; valuaziuns san gnir marginalisadas (ni focussadas) sin in'autra domena, e differentas valuaziuns san gnir cumbinadas (ni agregadas). Metodas genericas cun scomi d'informaziun san gnir duvradas per sligiar problems da projecziun. Biars problems en applicaziuns san gnir reduci ad in problem da projecziun: schazetg diagnostic, predicziun, filtraziun, smoothing. Aschia fuorman distribuziuns Gaussianas per exempel in'algebra da valuaziuns: marginalisaziun ei integraziun, e cumbinaziun ei multiplicaziun (plus renormalisaziun). Distribuziuns Gaussianas san gnir representadas tras potenzials Gaussians (culla matriza da concentraziun) ni tras matrizas dils muments (cullas varianzas e covarianzas). Cheu corrispundan marginalisaziun e cumbinaziun ad operaziuns da matrizas.

Ina distribuziun Gaussiana cundiziunala ei ina famiglia da distribuziuns Gaussianas ch'ins obtegn via las variablas da tgau tras fixar ina valur per las variablas da cua. Ella corrispunda ad ina distribuziun via las variablas da tgau plus ina regressiun lineara sillas variablas da cua. Distribuziuns Gaussianas cundiziunalas san gnir analisadas en treis modas e manieras: geometrica, algebraica ed analitica.

- L'inferenza che sebasa sillas interpretaziuns deriva in'indicaziun Gaussiana dad in sistem Gaussian-linear. Ils ensembels focals dad indicaziuns Gaussianas ein multiplicitads linearas parallelas dalla medema dimensiun el spazi dils parameters. La cumbinaziun corrispunda alla intersecziun dils ensembels focals e la marginalisaziun alla projecziun dils ensembels focals. Potenzials Gaussians corrispundan ad indicaziuns Gaussianas cun ensembels focals da mo in element.

- Potenzials Gaussians san gnir extendi ad in'algebra da valuaziuns da quozients che san gnir representai sco pèrs da potenzials Gaussians. Distribuziuns Gaussianas cundiziunalas san gnir representadas ella schinumnada extensiun separativa da potenzials Gaussians.

- Essend che distribuziuns Gaussianas cundiziunalas ein ina funcziun da quozient da duas distribuziuns Gaussianas, sa la matriza da concentraziun egl exponent dil numnader gnir subtrahada dalla matriza da concentraziun egl exponent dil

dumbrader. Quei meina ad ina nova representaziun da potenzials Gaussians simmetrics che han ina matriza da pseudo-concentraziun che ei mo simmetrica denton buc necessariamein positiv definita.

Il resultat principal da quellas consideraziuns ei quel che differentas distribuziuns Gaussianas cundiziunalas serefereschan alla medema indicaziun Gaussiana (tochen tier equivalenza) precis lu, sche las distribuziuns Gaussianas cundiziunalas ein identicas tochen tier in factur constant. Cun auters plaids, la funcziun da likelihood cuntegn l'entira informaziun dall'indicaziun Gaussiana. Quei explichescha pertgei l'inferenza che sebasa sillas interpretaziuns en sistems linears (sur-)determinai reproducescha ils resultats dils schazetgs che sebasan sil principi da maximum-likelihood. Variablas san esser cumbinaziuns linearas dad autras variablas. Quei inducescha restricziuns linearas sil spazi dils parameters. Metodas dad inferenza che sebasan sillas interpretaziuns, la cumbinaziun e la marginalisaziun ein derivadas per potenzials Gaussians simmetrics cun equaziuns deterministicas.

Alla fin vegn mussau co sistems Gaussians-linears san gnir exprimi el lungatg ABEL. Il sistem ABEL sa rispunder a damondas davart sistems Gaussians-linears cumplicai. Differents exempels illustreschan la representaziun nova da potenzials Gaussians simmetrics.

Contents

1 Introduction

Gaussian linear models have been extensively studied from various perspectives and in different fields. In this thesis, different structures of Gaussian linear information are compared from the *algebraic perspective*.

Chapter Outline

Gaussian linear models are illustrated in Section 1.1 by means of two simple introductory examples: noisy transmission over a Gaussian channel and Kalman filtering in a simple tracking problem. The algebraic perspective on Gaussian linear models is explained in Section 1.2. The principal goal of this thesis is to compare different algebras of Gaussian linear information, as discussed in Section 1.3. The basic results are sketched in Section 1.4. Finally, this thesis is outlined in Section 1.5.

1.1 Gaussian Linear Models

Gaussian linear models are presented by way of two introductory examples, followed by an outline of the general case.

A First Introductory Example: Noisy Transmission over a Gaussian Channel

Assume that Alice wants to send a message to her friend Bob, either 0 or 1. She may encode that piece of information as an input signal to a continuous physical channel going to Bob, i.e. the possible inputs are $\{x_0, x_1\} \subseteq \mathbb{R}$. However, the channel may be noisy or lossy: When the signal arrives at Bob, it may have changed on the way from Alice. This is shown in Figure 1.1: An error ω from a noise source (depicted as a cloud) is added to the input x. When the output arrives, Bob wants to know Alice's message. However, all he has got is the output signal $z = x + \omega$. He may also make some assumptions on the channel: For instance, he may assume that distortions are less likely the bigger they are. He assumes that the channel noise is Gaussian. Hence, the channel is specified as a *Gaussian linear model*: It is linear

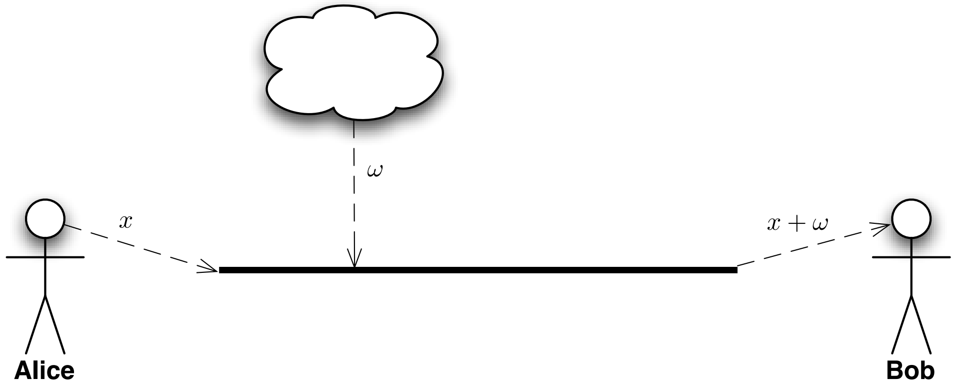


FIGURE 1.1: Gaussian Channel between Alice and Bob

since the output is a given linear combination of the input and a random noise, and it is Gaussian because of the given distribution of the additive noise term.

Given the Gaussian linear model of the channel, what can Bob infer on Alice's message? This situation is shown in Figure 1.2: He knows the received message z and the Gaussian linear model of the channel.

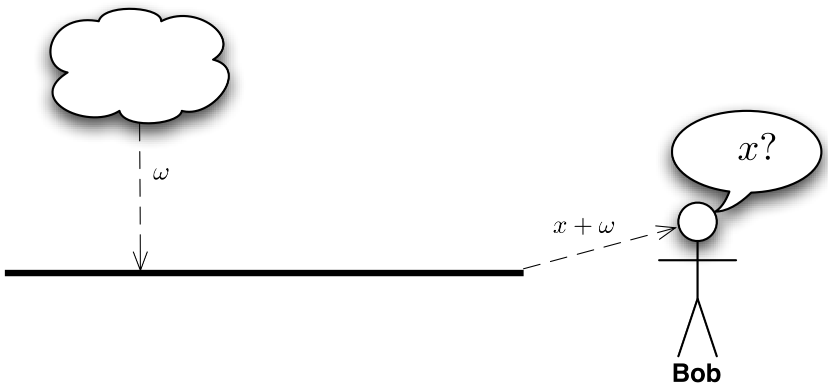


FIGURE 1.2: What can Bob infer on the input given the output and some knowledge about the channel?

A First Approach

When Alice sends a message to Bob, say the message 0, she can compute the probability of each outcome at Bob: The outcomes which are close to the input are more likely than those further away. Bob can use these conditional probabilities of the outcome given the input to choose the input under which the output is *more likely* or more *plausible* or *less surprising*. Therefore, this approach chooses the input of *maximum likelihood* given the output. However, it has to be emphasised that these conditional probabilities of the output given the input must not be confounded with a probability distribution on the input, even when the output is fixed.

A Second Approach

The first approach does not exploit all the information contained in the Gaussian linear model of the channel: If the actual error ω of the transmission were known, then Bob could compute the input $x = z - \omega$. However, Bob could not only consider the isolated likelihoods of the observation given either 0 or 1 was the input. He may also consider the probabilities in a *forward-looking*, *predicting* way: Given input x (0 or 1) and assuming an error ω, he is able to predict the outcome $z = x + \omega$. Given an observation z, only the two errors $\omega_0 = z - x_0$ and $\omega_1 = z - x_1$ are possible in light of the generating model. Thus, the Gaussian distribution has to be conditioned on the event $\{\omega_0, \omega_1\}$; the corresponding probabilities measure the strength of the hypotheses x_0 and x_1, respectively. Using this second approach, Bob gets the same numerical result with a different interpretation: The assumption ω_0 *proves* the hypothesis that x_0 was the input, and the assumption ω_1 proves the hypothesis x_1. Therefore, in this *assumption-based* approach, *probabilities of provability* are derived (Pearl, 1988).

In summary, the first approach only allows to derive probabilities of not disproving a hypothesis; the second approach also allows to derive probabilities of the evidence proving or supporting a hypothesis.

A Second Introductory Example: A Tracking Problem Using a Simple Kalman Filter

Consider the following example (Kalman, 1960): Particles leave the origin at time $t = 1$, each particle with a constant (unknown) velocity. Suppose that the position of one of these particles is measured repeatedly at the same interval $\Delta t = 1$. If the data is contaminated by stationary, additive, correlated noise, what can be inferred on the position and the velocity of the particle at the time $t = k$ of the last measurement? Let x_t be the position and $\vec{x}_t$ the velocity of the particle; $noise_t$ is the noise. The problem is then represented by the model

$$x_{t+1} = x_t + \vec{x}_t \tag{1.1}$$

$$\vec{x}_{t+1} = \vec{x}_t \tag{1.2}$$

$$noise_{t+1} = c_{t,t+1} \cdot noise_t + \omega_t \tag{1.3}$$

$$z_t = x_t + noise_t \tag{1.4}$$

for $t \in \{1, 2, \dots, k\}$. Assume that the disturbances ω_t are independent and normally distributed with mean 0 and variance σ_t^2. The unknown values in this problem are x_t, $\vec{x}_t$, $noise_t$, ω_t. The model, the measurements z_t and the correlation coefficients $c_{t,t+1}$ and the variances σ_t^2 are all known. A block diagram of the model is shown in Figure 1.3. Boxes represent functions, arrows labelled by variables refer to their in- and output, and black dots are branching points, i.e. equality constraints.

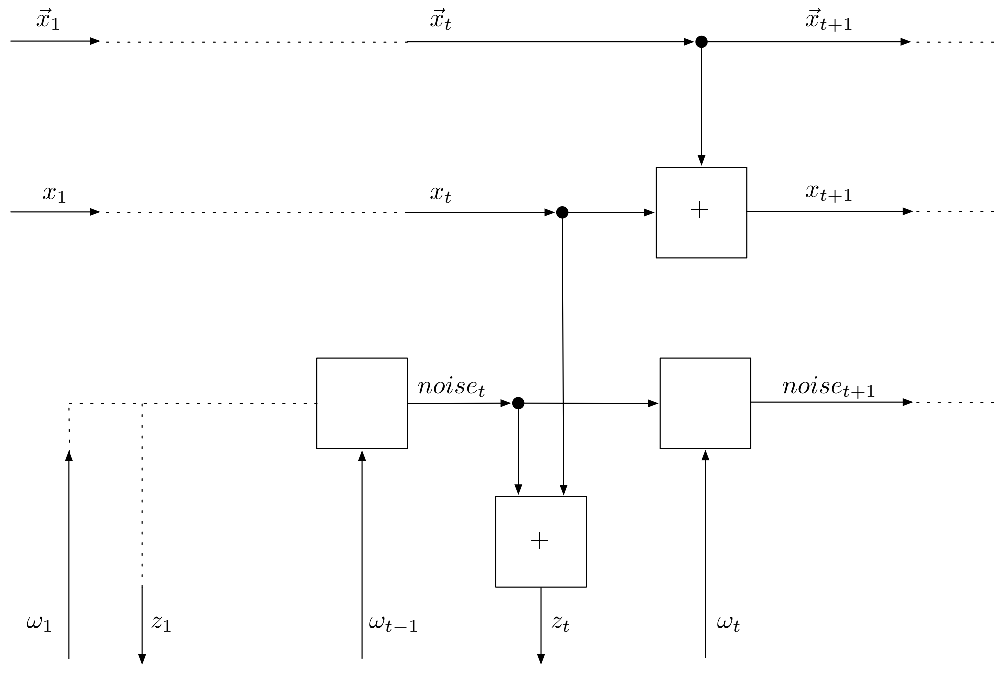


FIGURE 1.3: A block diagram for the tracking problem

The *filtering problem* consists in making inferences on the current state x_k. Again, the two methods can be applied. In order to apply the likelihood method, the probabilities of all sequences of unknown values x_t and $\vec{x}_t$ can be weighed. On the other hand, in the assumption-based approach, the conditional distribution of the sequences of $noise_t$ which are compatible with the observations z_t and which prove a hypothesis on x_k allow to derive the probability of the assumptions proving that hypothesis.

Gaussian Linear Systems

The foregoing two examples are instances of the more general form

$$A\mathbf{x} + \omega = z \tag{1.5}$$

for a vector $\mathbf{x}$ of unknown values, a known design matrix A, a known observation z and unknown disturbances ω whose Gaussian distribution is however known. This is a *functional model*, relating a parameter $\mathbf{x}$ and an assumption ω to the observation z: If the true parameter and the true disturbance were known, then the observation would necessarily be z. On the other hand, if the observation z is given and an ω is *assumed*, then only a subset of the parameters is compatible with the observation and the assumption. Thereby, the disturbances ω in the generating model (1.5) become assumptions in the reasoning process. The geometric picture of the assumption-based inference on Gaussian linear systems is very simple: For different ω, the solution sets or *focal sets*

$$\Gamma_z(\omega) = \{\mathbf{x} : A\mathbf{x} + \omega = z\} \tag{1.6}$$

are parallel linear manifolds. However, some ω are not possible (if $\Gamma_z(\omega) = \emptyset$) and the distribution of the disturbances has to be conditioned on the event of possible assumptions

$$v_z = \{\omega : \exists \mathbf{x} \text{ s.t. } \omega = z - A\mathbf{x}\} \tag{1.7}$$

induced by the model. The result of such *assumption-based reasoning* is captured in a structure called *Gaussian hint* (Monney, 2003). In contrast to classical regression analysis, neither have restrictions to be imposed on the rank of A nor on the form of the distribution of the ω in order to apply assumption-based reasoning.

Such a Gaussian hint allows to make both statements of support and of plausiblity (or possibility or likelihood) on a hypothesis H on the parameters $\mathbf{x}$: The assumptions under which a hypothesis is necessarily true are said to support H, whereas the assumptions under which the hypothesis is not disproved are only plausible. The supporting and plausible assumptions of a hypothesis induce a *subadditive measure* (i.e. $sp(H) + sp(H^c) \leq 1$) and a *superadditive measure* (i.e. $pl(H) + pl(H^c) \geq 1$), respectively.

Gaussian linear models arise in very different fields: *control theory* (Kalman, 1960) and *coding theory* (MacKay, 2003), in *regression analysis* (in econometrics), or *Bayesian networks* (expert systems). The different graphical representations of Gaussian linear information (such as by Bayesian networks or block diagrams) can all be reduced to a Gaussian linear system.

Gaussian linear models have been extensively studied in the past: On the one hand, many problems can be modelled appropriately by Gaussian linear information; on the other hand, Gaussian linear models are computationally tractable. This has often been pointed out in the literature:

> The fact that Kalman filter models have proven useful in many situations is partly a consequence of their mathematical and computational tractability, but a more fundamental reason is that the models portray with sufficient accuracy the phenomena under analysis. In the case of a physical system, the state equations often derive in part from well-studied physical phenomena, and similarly the observation equations derive from understanding of measurement processes. The associated Gaussian distribution assumptions [...] are sometimes made without much thought, but can often be empirically substantiated, and of course should be treated as potential sources of mistaken conclusions if their validity and consequences are not analyzed. In other situations with softer theoretical and empirical support, the assumed equations, distributions, and independence properties may be deliberately speculative, but still may improve substantially on nonstochastic thinking. (Dempster, 1990b; p.127)

> The popularity of linear gaussian models comes from two fortunate analytical properties of gaussian processes: the sum of two independent

> gaussian distributed quantities is also gaussian distributed, [...] and the output of a linear system whose input is gaussian distributed is again gaussian distributed. (Roweis and Ghahramani, 1999; p.309)

The Gaussian distribution is often used as a means of modelling random noise:

> It is a fairly generally accepted fact that primary macroscopic sources of random phenomena are independent gaussian processes[:] The probability distributions will be gaussian because macroscopic random effects may be thought of as the superposition of very many microscopic random effects; under very general conditions, such aggregate effects tend to be gaussian, regardless of the statistical properties of the microscopic effects. The assumption of independence in this context is motivated by the fact that microscopic phenomena tend to take place much more rapidly than macroscopic phenomena; thus primary random sources would appear to be independent on a macroscopic time scale. (Kalman, 1960; p.39)

1.2 Algebras of Gaussian Linear Information

The foregoing discussion shows the basic operations on Gaussian linear information.

- Given Gaussian linear information, the basic interest is *inference* on variables whose values may be unknown. An algebra of Gaussian linear information has therefore to provide an operator for *extracting the relevant information*.

- Since the problem structure is most often *modular*, the problem can be split into independent factors. Therefore, an algebra of Gaussian linear information has to provide an operator for *combining* pieces of Gaussian linear information, each of which refering to a given set of variables.

- When solving the inference problem, the modular structure should be exploited for efficient *local* computations, i.e. computations on domains not larger than the individual factors.

In mathematical terms, the first two points correspond to an *algebra of Gaussian linear information*. It turns out that algebras of Gaussian linear information fit into a more general framework of *valuation algebras* for local computation. This relates Gaussian linear information to other structures representing probabilistic, uncertain or logical knowledge. In this framework, information is defined implicitly by a set of axioms imposed on the operations. Furthermore, these axioms are sufficient for local computation.

- Pieces of information are called *valuations*. They refer to a *domain* of interest, which is a set of *variables*.

- There is a *combination* operator for valuations.

- Valuations can be *projected* or *marginalised* to a domain of interest.

- A *projection problem* is given by a *factorisation* of the combined model and the *queries*, i.e. the domains of interest.
- Algorithms for *local computation* work on *join trees*. Each node in the tree contains a factor. Join trees have the *running intersection property*, i.e., if a variable occurs in two different nodes, it has to be in every node on the unique path between those two nodes.

1.3 Motivation & Purpose

On a practical level, Gaussian linear information often contains *redundancy*. Redundancy can be treated and described in different ways, for instance by *sparse matrix techniques* or by *conditional independences* as reflected by *Bayesian networks*. This thesis advocates the following point of view: Redundancy in the model is reflected by independent modules referring only to a subset of all variables, i.e. by a *factorisation* of the overall model. The overall model is the (virtual) combination of all these factors. When the inference problem regarding a particular domain of interest has to be computed, the *domain structure* of the factorisation should be taken into account in order to make computations tractable and to avoid inefficient computations. For this purpose, the framework of valuation algebras will be used, since it provides generic message-passing algorithms for local computation.

In fact, there are many different representations of Gaussian linear information:

- likelihood functions or conditional Gaussian densities;
- Gaussian densities plus regression equations;
- Gaussian linear systems (including Gaussian linear regression models);
- quotient functions of a Gaussian density and a marginal (Kohlas, 2003);
- Gaussian hints (Monney, 2003; Kohlas and Monney, 2008) derived by assumption-based reasoning (Kohlas and Monney, 1995);
- Gaussian belief functions (Dempster, 1990a; Liu, 1996a; 1999).

However, it is not well understood how these different representations are related. Therefore, the main purpose of this thesis is

- to present different algebras of Gaussian linear information,
- to show how they are related, and
- to discuss their respective advantages on the computational level.

In addition, a new representation of Gaussian linear information, *symmetric Gaussian potentials*, is developed.

Furthermore, some of these representations do not include *deterministic linear knowledge*. Adding deterministic linear equations imposes crisp restrictions on the values of the parameter; if there is logically no explanation for the observation, the Gaussian linear model even becomes *contradictory*. Therefore, a second purpose of this thesis is to

- analyse Gaussian linear systems with deterministic equations using assumption-based reasoning and to
- study the effect of deterministic knowledge on Gaussian hints and symmetric Gaussian potentials.

1.4 Results & Validation

Algebras of Gaussian Linear Information

A Gaussian linear functional model (1.5) induces different families of conditional Gaussian distributions for different conditioning variables. In general, such a parametric setting only allows to make probabilistic statements of plausibility or likelihood. In contrast, in the Gaussian linear case, it turns out that *the induced conditional Gaussian densities contain the full information of the Gaussian linear model.* This remarkable result is derived from three different perspectives.

- *Geometrically*, the points of the same conditional Gaussian density value form parallel linear manifolds. If the corresponding Gaussian linear system of regression equations is interpreted in a prescriptive way of how observations are generated, this leads to the focal sets $\Gamma(\omega)$ of (1.6). Hence, a conditional Gaussian density represents a *distribution over these parallel linear manifolds* $\Gamma(\omega)$. Remarkably, it turns out that different conditional Gaussian densities (with different conditioning variables) are related to the same Gaussian hint if and only they are the same up to a constant factor; in this case, they are called *equivalent*. The geometric interpretation of this equivalence is simple: The constant factor accounts for different axes of integration for the distribution over the focal sets $\Gamma(\omega)$.
- *Algebraically*, a conditional Gaussian density can be described as a quotient function of a Gaussian density and the marginal density for the conditioning variables. Therefore, a conditional Gaussian density can be represented by a pair of two Gaussian densities. Using purely algebraic arguments, it turns out that *separativity* is a sufficient criterion for embedding a valuation algebra into a valuation algebra of pairs (Kohlas, 2003). It turns out that the same equivalence of conditional Gaussian densities is induced by purely algebraic arguments instead of geometric arguments.
- *Analytically*, the quotient of two Gaussian densities corresponds to the matrix difference of their concentration matrices. The operations from the geometric and the analytic approach can then be carried over to this new representation.

Symmetric Gaussian Potentials

The analytic approach leads to symmetric Gaussian potentials. This new representation inherits full marginalisation from Gaussian hints and division from the separative extension of Gaussian potentials. Furthermore, it provides a unique representation of equivalent Gaussian hints and conditional Gaussian potentials.

Deterministic knowledge

Deterministic linear equations impose restrictions on the parameter space. This induces an event in a Gaussian hint. Algorithms are developed for the application of deterministic knowledge to Gaussian hints as well as for combination and marginalisation.

Validation

The validity of the algorithms developed for symmetric Gaussian potentials and deterministic knowledge is shown in an implementation and in various examples from the literature.

1.5 Thesis Outline

It follows a more detailed outline of the four main parts of this thesis.

Part I: Local Computation in Valuation Algebras

In the first part, local computation in valuation algebras is introduced by using Gaussian potentials as the basic example.

In Chapter 2, the framework of valuation algebras is introduced. Different *axiomatics* of valuation algebras – in terms of marginalisation or variable elimination – are discussed, and the *algebraic theory* is presented on an abstract level. Further, *stable* valuation algebras can be defined in a *domain-free* way. Finally, valuation algebras which decompose into a union of groups are discussed.

In Chapter 3, the valuation algebra of Gaussian potentials is introduced. A Gaussian potential represents a Gaussian density by its mean vector and its concentration matrix. Combination and marginalisation of Gaussian potentials will be defined in terms of matrix operations: combination corresponds to multiplication (plus renormalisation) and marginalisation to integration of Gaussian densities. Alternatively, a Gaussian density can be represented by a moment matrix, i.e. by its mean vector and its variance-covariance matrix. Further, the complexity of the operations is compared in both representations.

In Chapter 4, *local computation* in *join trees* and the basic algorithms (Collect, Shenoy-Shafer, Lauritzen-Spiegelhalter) are reviewed.

Part II: Conditional Gaussian Densities

In Chapter 5, conditional Gaussian densities are defined, and three approaches are introduced: A *geometric* approach linking conditional Gaussian densities to Gaussian hints, an *algebraic* approach linking conditional Gaussian densities to pairs of Gaussian potentials, and an *analytic* approach leading to symmetric Gaussian potentials. These three approaches are then discussed more thoroughly in the following chapters.

In Chapter 6, systems of linear equations with Gaussian disturbances are discussed by using *assumption-based reasoning* and the theory of *hints*. Gaussian hints, the result of the assumption-based inference, represent a distribution over parallel linear manifolds. Gaussian hints equally form a valuation algebra. Gaussian potentials can be embedded in this valuation algebra and are represented by precise Gaussian hints, which represent a distribution over points (parallel linear manifolds of dimension 0).

In Chapter 7, it is shown that different conditional Gaussian densities (with different head and tail variables) are related to the same Gaussian hint (up to equivalence). It is shown that the combination of Gaussian hints corresponds to the union of the heads and (up to renormalisation) to the multiplication of conditional Gaussian densities. Marginalisation corresponds to integration over a head variable or the reduction of *vacuous variables*, which are always in the tail.

In Chapter 8, an algebraic theory is given for embedding a *separative* valuation algebra into a union of groups whose elements are represented by pairs of valuations. Although conditional Gaussian densities can be represented in the separative extension of Gaussian potentials, only marginalisation corresponding to integration is possible; vacuous variables cannot be reduced.

In Chapter 9, it is shown that the symmetric matrix obtained by subtracting the numerator and the denominator concentration matrices of a quotient of Gaussian densities can be used for a new unique representation. These *symmetric Gaussian potentials* consist of a vector and a symmetric matrix of the same domain. Furthermore, any symmetric matrix is the difference of two symmetric positive definite matrices whose domains overlap. Therefore, symmetric Gaussian potentials and the elements of the separative extension are in one-to-one correspondence. Combination and marginalisation can be carried over from the separative extension and from Gaussian hints. Thereby, the valuation algebra of symmetric Gaussian potentials inherits full marginalisation from Gaussian hints and division from the separative extension. Conditional Gaussian densities correspond to a symmetric Gaussian potential whose pseudo-concentration matrix is non-negative definite.

Part III: Deterministic Knowledge

In Chapter 10, deterministic variables and deterministic linear systems are captured by a deterministic hint. Such deterministic linear hints induce an event in a Gaussian hint or the corresponding symmetric Gaussian potential. Algorithms for the application of deterministic knowledge as well as for the combination and the marginalisation of these generalised symmetric Gaussian potentials are derived.

Part IV: Applications and Implementation

In Chapter 11, the *Kalman filter model* as well as the filtering, smoothing and prediction problems can easily be translated into the messaging-passing schemes of local computation. The different parts of the Kalman filter model are represented by symmetric Gaussian potentials.

In Chapter 12, an implementation of the algorithms for symmetric Gaussian potentials is shown. Gaussian linear systems and queries can be expressed in the Abel language and are then handled by the Gauss solver. Queries can be answered by using the Nenok framework for local computation.

In Chapter 13, some concrete applications of Gaussian linear systems are formulated and computed in Abel.

Finally, a conclusion will be given in Chapter 14. In Appendix A, some results from matrix algebra used in the text are listed; in Appendix B, the Gaussian distribution is reviewed.

Chapter Structure

Every chapter consists of the following sections:

1. an *introduction*, where context and motivation are given in an informal way;
2. a *chapter outline*, where the thread of the argumentation in the overall chapter is outlined;
3. the serially numbered *sections* figuring in the table of contents, where the theory is developed or exposed formally;
4. *chapter synopsis and discussion*, recapitulating the most important points in technical terms, relating the exposition to previous work in the field and suggesting alternatives and open questions.

Proofs are terminated by the symbol □, examples, definitions, lemmata and theorems by the symbol ⊘. Terms defined in the text are highlighted using this font.

Part I

Local Computation in Valuation Algebras

2

Valuation Algebras

The algebraic framework of valuation algebras unifies information processing of seemingly disparate domains such as relational databases, belief functions, probability potentials etc.; see (Shafer, 1991; Kohlas, 2003) for more details and references. *Query* answering in these knowledge representation systems is often computationally expensive. Therefore, efficient algorithms have been developed in various domains. It has turned out that many of them are instances of generic algorithms in the abstract framework of valuation algebras; these algorithms will be the topic of Chapter 4.

Chapter Outline

The following four topics are discussed in this chapter.

1. *Axiomatics*: In Section 2.1, valuation algebras are formally defined in terms of marginalisation. In Section 2.2, valuation algebras are defined in terms of variable elimination. This second approach is equivalent with the first one if domains are restricted to be finite.

2. *Algebraic theory*: In Section 2.3, neutral elements for each domain as well as null elements are discussed. In Section 2.4, algebraic notions such as congruences, homomorphisms and quotient algebras are carried over to valuation algebras. Furthermore, if the combination of a valuation with a part of the same valuation yields nothing new with respect to a congruence, the congruence is called idempotent. Idempotent congruences induce a partial order of equivalence classes into valuation algebras. Such idempotent congruences are studied in Section 2.5.

3. *Stability*: If valuations may represent the same information with respect to different domains, the valuation algebra is called stable as discussed in Section 2.6. From such a stable valuation algebra, a domain-free valuation algebra can be derived; conversely, a domain-free valuation algebra induces a stable valuation algebra. This is the topic of Section 2.7.

4. *Division*: Division in a valuation algebra can be defined if it is composed of disjoint groups with inverses. This notion will be formally defined in Section 2.8. On the one hand, division is important for advanced local computation techniques, see Chapter 4. On the other hand, division may also be relevant for the semantical point of view, see Chapter 5.

2.1 The Algebraic Framework of Valuation Algebras

The following definition captures an algebraic structure of information pieces called valuations. A piece of information refers to a domain of interest consisting of a set of variables. The sets of domains considered are supposed to be closed under union $\cup$ and intersection $\cap$, i.e. they are supposed to form a lattice. For instance, the lattice of all subsets of a set r is the triplet $(2^r, \cup, \cap)$ where

$$2^r = \{s : s \subseteq r\}, \tag{2.1}$$

$\cup$ is set union and $\cap$ is intersection. The set 2^r of all subsets is called the powerset of r. Another example is the set lattice (with union and intersection) over the set D_{finite} of finite subsets of a set r,

$$D_{finite} = \{s : s \subseteq r, s \text{ finite}\}. \tag{2.2}$$

DEFINITION **2.1**. *Let Φ be a set of elements called* valuations *and let $(D, \cup, \cap)$ be a lattice of subsets of a set r with partial order $\subseteq$. The elements of r are called* variables *and the elements of D are called* domains. *Suppose that there are four operations defined:*

1. labelling: $d : \Phi \to D$, $\phi \mapsto d(\phi)$,
2. combination: $\otimes : \Phi \times \Phi \to \Phi$, $(\phi, \psi) \mapsto \phi \otimes \psi$,
3. domain operator: $\mathcal{M} : \Phi \to 2^D$, $\phi \mapsto \mathcal{M}(\phi)$, *and*
4. marginalisation: $\downarrow: \Phi \times D \to \Phi$, $(\phi, s) \mapsto \phi^{\downarrow s}$ *defined for* $s \in \mathcal{M}(\phi)$.

The set $\mathcal{M}(\phi)$ contains all domains $s \in D$ such that the marginal $\phi^{\downarrow s}$ of ϕ is defined relative to the domain s.

Let the following set of axioms be imposed on Φ and D.

(A1) Commutative Semigroup: *Φ is associative and commutative under combination, i.e. for $\phi, \psi, \chi \in \Phi$,*

$$\phi \otimes (\psi \otimes \chi) = (\phi \otimes \psi) \otimes \chi, \qquad \text{and}$$
$$\phi \otimes \psi = \psi \otimes \phi.$$

(A2) Labelling: *For $\phi, \psi \in \Phi$,*

$$d(\phi \otimes \psi) = d(\phi) \cup d(\psi). \tag{2.3}$$

(A3) Marginalisation: *For* $\phi \in \Phi$ *and* $s \in \mathcal{M}(\phi)$,

$$d(\phi^{\downarrow s}) = s. \tag{2.4}$$

(A4) Transitivity: *If* $\phi \in \Phi$, $s \subseteq t \subseteq d(\phi)$, *it follows that*

$$s \in \mathcal{M}(\phi) \iff t \in \mathcal{M}(\phi),\, s \in \mathcal{M}(\phi^{\downarrow t}); \tag{2.5}$$

then, it also holds that

$$(\phi^{\downarrow t})^{\downarrow s} = \phi^{\downarrow s}. \tag{2.6}$$

(A5) Combination: *If* $\phi, \psi \in \Phi$ *with* $d(\phi) = x$, $d(\psi) = y$ *and* $z \in D$ *such that* $x \subseteq z \subseteq x \cup y$, *then* $z \cap y \in \mathcal{M}(\psi)$ *implies* $z \in \mathcal{M}(\phi \otimes \psi)$ *and*

$$(\phi \otimes \psi)^{\downarrow z} = \phi \otimes \psi^{\downarrow z \cap y}.$$

(A6) Domain: $\phi \in \Phi$ *with* $d(\phi) = x$ *implies that* $x \in \mathcal{M}(\phi)$ *and*

$$\phi^{\downarrow x} = \phi. \tag{2.7}$$

(A7) Identity Element: *There is an element* $e \in \Phi$, $d(e) = \emptyset$, $\mathcal{M}(e) = \{\emptyset\}$ *such that for any* $\phi \in \Phi$

$$\phi \otimes e \quad = \quad \phi \quad = \quad e \otimes \phi. \tag{2.8}$$

A sextuple $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ *satisfying these axioms is called a* labelled valuation algebra. *If* $\mathcal{M}(\phi) = 2^{d(\phi)}$ *for all* $\phi \in \Phi$, *then* $\mathfrak{A}$ *is called* valuation algebra with full marginalisation, *abbreviated* $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$. ⊘

The axioms of a valuation algebra represent natural properties of information processing. The first axiom indicates that, if information comes in pieces, the sequence does not influence the overall information. The labelling axiom says that the combination of valuations gives information over the union of the domains involved; neither do variables vanish, nor do new ones appear. The marginalisation axiom says that marginalisation yields an element of the target domain. The transitivity axioms says that the direct and the two-step marginalisation lead to the same result. The combination axiom is essential for local computation since marginalisation of a product can be performed in the factors if one the factor domains is a subset of the marginal domain. The domain axiom says that information is not influenced by projecting it to its own domain. Without the domain axiom, this is not always the case (Shafer, 1991). The identity element axiom is used for technical reasons in the local computation algorithms as discussed in Chapter 4. In (Schneuwly et al., 2004), it is shown that such an element can be assumed without loss of generality.

Remark **2.2**. The basic model for the valuation algebra axioms are quotient functions $f(\mathbf{x}) = \frac{f_1(\mathbf{x}^{\downarrow d(f_1)})}{f_2(\mathbf{x}^{\downarrow d(f_2)})}$ of domain $x = d(f_1) \cup d(f_2)$. Here, variables in the numerator f_1 can be integrated out as long as they do not occur in the denominator f_2. Therefore, the quotient function can be marginalised to any domain s such that $d(f_2) \subseteq s \subseteq x$. ⊘

Remark **2.3**. In the case of a valuation algebra with full marginalisation, the transitivity and the combination axioms can be simplified as follows.

(A4)′ *Transitivity* (full marginalisation): If $\phi \in \Phi$, $s \subseteq t \subseteq d(\phi)$, then

$$(\phi^{\downarrow t})^{\downarrow s} = \phi^{\downarrow s}. \tag{2.9}$$

(A5)′ *Combination* (full marginalisation): If $\phi, \psi \in \Phi$ with domains $x = d(\phi)$ and $y = d(\psi)$, $x \subseteq s \subseteq x \cup y$, then

$$(\phi \otimes \psi)^{\downarrow s} = \phi \otimes \psi^{\downarrow s \cap y}. \tag{2.10}$$

⊘

The following lemma gives an important property of valuation algebras.

Lemma **2.4**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a valuation algebra, and $\phi, \psi \in \Phi$, $x = d(\phi)$, $y = d(\psi)$, and $s \in D$. If $x \cap y \subseteq s \subseteq x \cup y$, $s \cap x \in \mathcal{M}(\phi)$ and $s \cap y \in \mathcal{M}(\psi)$, then $s \in \mathcal{M}(\phi \otimes \psi)$ and*

$$(\phi \otimes \psi)^{\downarrow s} = \phi^{\downarrow x \cap s} \otimes \psi^{\downarrow y \cap s}. \tag{2.11}$$

⊘

Proof. Since $x \cap y \subseteq s \cap y \subseteq s$ and $s \cap y = (s \cup x) \cap y$, by the transitivity and the combination axioms,

$$\begin{aligned} \phi^{\downarrow s \cap x} \otimes \psi^{\downarrow s \cap y} &= (\phi \otimes \psi^{\downarrow s \cap y})^{\downarrow s} \\ &= (\phi \otimes \psi^{\downarrow (s \cup x) \cap y})^{\downarrow s} \\ &= ((\phi \otimes \psi)^{\downarrow s \cup x})^{\downarrow s} \\ &= (\phi \otimes \psi)^{\downarrow s}. \end{aligned}$$

□

It says that the marginalisation of a product can be done on the factors if the intersection of the factor labels is smaller than the domain of the marginalisation (provided that the necessary marginals are defined).

2.2 Variable Elimination

The transitivity axiom allows to eliminate variables in a valuation one by one in any order instead of directly marginalising the valuation to the corresponding subdomain. The notation

$$\phi^{-X} = \phi^{\downarrow d(\phi) - \{X\}} \tag{2.12}$$

is used for the elimination of X in ϕ provided that $d(\phi) - \{X\} \in \mathcal{M}(\phi)$. More generally, the definition of a valuation algebra may be reformulated in terms of variable elimination instead of marginalisation if D consists only of *finite* sets.

Definition **2.5**. *Let Φ be a set of valuations and let D be a lattice of* finite *subsets of a set r of variables. Suppose that there are four operations defined:*

1. *Labelling:* $d : \Phi \to D$; $\phi \mapsto d(\phi)$,

2. *Combination:* $\otimes : \Phi \times \Phi \to \Phi$; $(\phi, \psi) \mapsto \phi \otimes \psi$,

3. *Variable Operator:* $\mathcal{V} : \Phi \to D$; $\phi \mapsto \mathcal{V}(\phi)$ *where* $\mathcal{V}(\phi) \subseteq d(\phi)$,

4. *Variable Elimination:* $- : \Phi \times D \to \Phi$; $(\phi, X) \mapsto \phi^{-X}$ *defined for* $X \in \mathcal{V}(\phi)$.

The set $\mathcal{V}(\phi)$ *contains all variables* $s \subseteq d(\phi)$ *which can be eliminated in* ϕ, *i.e. such that* ϕ^{-X} *is defined relative to the valuation* $\phi \in \Phi$.

Let the following set of axioms be imposed on Φ *and* D.

(E1) Commutative Semigroup: *Φ is associative and commutative under combination, i.e. for* $\phi, \psi, \chi \in \Phi$,

$$\phi \otimes (\psi \otimes \chi) = (\phi \otimes \psi) \otimes \chi, \qquad \textit{and}$$
$$\phi \otimes \psi = \psi \otimes \phi.$$

(E2) Labelling: *For* $\phi, \psi \in \Phi$,

$$d(\phi \otimes \psi) = d(\phi) \cup d(\psi). \tag{2.13}$$

(E3) Variable Elimination: *For* $\phi \in \Phi$ *and* $X \in \mathcal{V}(\phi)$,

$$d(\phi^{-X}) = d(\phi) - \{X\}. \tag{2.14}$$

(E4) Transitivity: *For* $\phi \in \Phi$,

$$X \in \mathcal{V}(\phi) \text{ and } Y \in \mathcal{V}(\phi^{-X}) \implies Y \in \mathcal{V}(\phi) \text{ and } X \in \mathcal{V}(\phi^{-Y}); \tag{2.15}$$

then, it also holds that

$$\left(\phi^{-X}\right)^{-Y} = \left(\phi^{-Y}\right)^{-X}. \tag{2.16}$$

(E5) Combination: *If* $\phi, \psi \in \Phi$ *with* $d(\phi) = x$, $d(\psi) = y$ *and* $X \notin d(\phi)$, *then* $X \in \mathcal{V}(\psi)$ *implies* $X \in \mathcal{V}(\phi \otimes \psi)$, *and*

$$(\phi \otimes \psi)^{-X} = \phi \otimes \psi^{-X}. \tag{2.17}$$

(E6) Identity Element: *There is an element* $e \in \Phi$, $d(e) = \emptyset$, $\mathcal{V}(e) = \emptyset$ *such that for any* $\phi \in \Phi$

$$\phi \otimes e \quad = \quad e \otimes \phi \quad = \quad \phi. \tag{2.18}$$

A sextuple $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{V}, -)$ *satisfying these axioms is called a* *labelled valuation algebra with variable elimination.*

⊘

Define $(X_1, X_2, \ldots, X_n) \in \mathcal{V}(\phi)$ if

- $X_i \in \mathcal{V}(\phi_{i-1})$, $\phi_0 = \phi$, $\phi_i = \left((\phi^{-X_1})^{\cdots}\right)^{-X_i}$, $i \in \{1, \ldots, n-1\}$.

The following lemma shows that variables can be eliminated in any order.

LEMMA **2.6**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{V}, -)$ be a valuation algebra with partial variable elimination. Let $(X_1, X_2, \ldots, X_n) \in \mathcal{V}(\phi)$ for some $\phi \in \Phi$. Then, $(X_{i_1}, \ldots, X_{i_n}) \in \mathcal{V}(\phi)$ and*

$$((\phi^{-X_1})^{\cdots})^{-X_n} = ((\phi^{-X_{i_1}})^{\cdots})^{-X_{i_n}} \tag{2.19}$$

for every permuation $i_1, \ldots, i_n$ of $1, \ldots, n$. ⊘

PROOF. Any two consecutive variables X_i, X_{i+1} can be swapped without affecting the result: Since $X_{i+1} \in \mathcal{V}(\phi_i) = \mathcal{V}({\phi_{i-1}}^{-X_i})$, it follows that by the transitivity axiom that $X_{i+1} \in \mathcal{V}(\phi_{i-1})$, $X_i \in \mathcal{V}({\phi_{i-1}}^{-X_{i+1}})$ and $({\phi_{i-1}}^{-X_{i+1}})^{-X_i} = ({\phi_{i-1}}^{-X_i})^{-X_{i+1}} = \phi_{i+1}$. Therefore, the equality can be established in the following way: For j from 1 to n, bring variable X_{i_j} in front by "bubbling" variable X_{i_j} up to position j by at most $n - j$ swappings. □

In light of the previous lemma, define

$$\phi^{-\{X_1,\ldots,X_n\}} = ((\phi^{-X_{i_1}})^{\cdots})^{-X_{i_n}} \tag{2.20}$$

if $(X_1, \ldots, X_n) \in \mathcal{V}(\phi)$ and define $\phi^{-\emptyset} = \phi$.

REMARK **2.7**. $(X_1, \ldots, X_n) \in \mathcal{V}(\phi)$ implies $\{X_1, \ldots, X_n\} \subseteq \mathcal{V}(\phi)$, whereas the converse is not necessarily true. ⊘

THEOREM **2.8**. *Let D be the lattice of all finite subsets of a set r of variables.*

(1) *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a valuation algebra. Then, the algebraic structure $\mathfrak{A}_{\mathcal{M}} = (\Phi, D, d, \otimes, \mathcal{V}_{\mathcal{M}}, -_{\mathcal{M}})$ is a valuation algebra with variable elimination where $\mathcal{V}_{\mathcal{M}}$ and $-_M$ are defined by*

$$\mathcal{V}_{\mathcal{M}}(\phi) = \{X : d(\phi) - \{X\} \in \mathcal{M}(\phi)\} \tag{2.21}$$

and for such an X

$$\phi^{-_{\mathcal{M}}X} = \phi^{\downarrow d(\phi)-\{X\}} \tag{2.22}$$

for $\phi \in \Phi$.

(2) *If $\mathfrak{A}' = (\Phi, D, d, \otimes, \mathcal{V}, -)$ is a valuation algebra with variable elimination, then $\mathfrak{A}_{\mathcal{V}} = (\Phi, D, d, \otimes, \mathcal{M}_{\mathcal{V}}, \downarrow_{\mathcal{V}})$ is a valuation algebra where, for $\phi \in \Phi$, $\mathcal{M}_{\mathcal{V}}$ and $\downarrow_{\mathcal{V}}$ are defined by*

$$\mathcal{M}_{\mathcal{V}}(\phi) = d(\phi) \cup \{s = d(\phi) - \{X_1, \ldots, X_n\} : (X_1, \ldots, X_n) \in \mathcal{V}(\phi), \tag{2.23}$$

and then

$$\phi^{\downarrow_{\mathcal{V}} s} = \phi^{-\{X_1,\ldots,X_n\}}. \tag{2.24}$$

⊘

PROOF. Marginalisation derived from variable elimination is well defined in light of Lemma 2.6.
Axioms (A1) and (E1) are the same, as well as (A2) and (E2). On the one hand, the marginalisation axiom (A3) in $\mathfrak{A}_{\mathcal{V}}$ follows from the definition (2.24) and the variable elimination axiom (E3). On the other hand, the variable elimination axiom (E3) in $\mathfrak{A}_{\mathcal{M}}$ follows from the definition (2.22) and the marginalisation axiom (A3). The domain axiom (A6) in $\mathfrak{A}_{\mathcal{V}}$ follows from the definition (2.24). Observing that $\mathcal{V}_{\mathcal{M}}(e) = \emptyset$ and $\mathcal{M}_{\mathcal{V}}(e) = \emptyset$, the identity axioms hold in $\mathfrak{A}_{\mathcal{M}}$ and $\mathfrak{A}_{\mathcal{V}}$.

(A4) Let $s \subseteq t \subseteq d(\phi)$. Define $x_1 = d(\phi) - t = \{X_1, \ldots, X_m\}$ and $x_2 = t - s = \{X_{m+1}, \ldots, X_{m+n}\}$. In light of Lemma 2.6, $\phi^{-x_1-x_2}$ is defined if and only if ϕ^{-x} is defined, and then also $(\phi^{-x_1})^{-x_2} = \phi^{-x}$. Therefore, $s \in \mathcal{M}_{\mathcal{V}}(\phi)$ if and only if $t \in \mathcal{M}_{\mathcal{V}}(\phi)$ and $s \in \mathcal{M}_{\mathcal{V}}(\phi^{-x_1}) = \mathcal{M}_{\mathcal{V}}(\phi^{\downarrow_{\mathcal{V}} t})$. Further, $(\phi^{\downarrow_{\mathcal{V}} t})^{\downarrow_{\mathcal{V}} s} = \phi^{-x_1-x_2} = \phi^{-x} = \phi^{\downarrow_{\mathcal{V}} s}$.

(A5) Assume $d(\phi) \subseteq z \subseteq d(\phi) \cup d(\psi)$ and $z \in \mathcal{M}_{\mathcal{V}}(\psi)$. Let $\{X_1, \ldots, X_m\} = d(\psi) - z$. Then, $(X_1, \ldots, X_m) \in \mathcal{V}(\phi)$. Hence, successive application of the combination axiom shows that $(X_1, \ldots, X_m) \in \mathcal{V}(\phi \otimes \psi)$. Hence, $d(\psi) - z \in \mathcal{M}_{\mathcal{V}}(\psi)$. Further, $(\phi \otimes \psi)^{\downarrow_{\mathcal{V}} z} = (\phi \otimes \psi)^{-\{X_1,\ldots,X_m\}} = \phi \otimes \psi^{-\{X_1,\ldots,X_m\}} = \phi \otimes \psi^{\downarrow_{\mathcal{V}} z \cap d(\psi)}$.

(E4) Assume $X \in \mathcal{V}_{\mathcal{M}}(\phi)$ and $Y \in \mathcal{V}_{\mathcal{M}}(\phi^{-_{\mathcal{M}} X})$. This implies $s \in \mathcal{M}(\phi)$ for $s = d(\phi) - \{X, Y\}$. Hence, using the transitivity axiom, $t \in \mathcal{M}(\phi)$ and $s \in \mathcal{M}(\phi^{\downarrow t})$ for $t = d(\phi) - \{X\}$. Therefore, $Y \in \mathcal{V}_{\mathcal{M}}(\phi)$ and $X \in \mathcal{V}_{\mathcal{M}}(\phi^{-_{\mathcal{M}} Y})$ since $\phi^{-_{\mathcal{M}} X} = \phi^{\downarrow t}$. Furthermore, $(\phi^{-_{\mathcal{M}} X})^{-_{\mathcal{M}} Y} = \phi^{\downarrow s} = (\phi^{-_{\mathcal{M}} Y})^{-_{\mathcal{M}} X}$.

(E5) Assume $X \in \mathcal{V}_{\mathcal{M}}(\psi)$ and $X \notin d(\phi)$. Define $z = (d(\phi) \cup d(\psi)) - X$. Then, the combination axiom implies that $z \in \mathcal{M}(\phi \otimes \psi)$, i.e. $X \in \mathcal{V}_{\mathcal{M}}(\phi \otimes \psi)$. Further, $(\phi \otimes \psi)^{-_{\mathcal{M}} X} = (\phi \otimes \psi)^{\downarrow z} = \phi \otimes \psi^{\downarrow z \cap d(\psi)} = \phi \otimes \psi^{-_{\mathcal{M}} X}$. □

If the elements of D are not finite, then marginalisation in a labelled valuation algebra cannot be expressed as a finite sequence of variable eliminations; so marginalisation is more general than variable elimination.

REMARK **2.9**. $\mathcal{V}(\phi) = d(\phi)$ does not imply that $\mathcal{M}_{\mathcal{V}} = 2^{d(\phi)}$, whereas $\emptyset \in \mathcal{M}(\phi)$ implies $\mathcal{V}_{\mathcal{M}}(\phi) = d(\phi)$. ⊘

2.3 Neutral and Null Elements in Valuation Algebras

Many valuation algebras $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ have a neutral element with respect to combination for every domain $s \in D$.

DEFINITION **2.10**. *An element $e_s \in \Phi_s = \{\phi \in \Phi : d(\phi) = s\}$ is called neutral element of the subsemigroup Φ_s if*

$$e_s \otimes \phi = \phi \tag{2.25}$$

for all valuations $\phi \in \Phi_s$. ⊘

Since neutral elements do not add any information, they represent *empty* or *vacuous information*. However, there are important examples where such elements can either not be represented explicitly or do not exist at all. For instance, in relational database theory, the neutral element for a domain is the relation that contains all – possibly infinitely many – tuples for some domain, see (Kohlas, 2003). If there are neutral elements for all domains, it is postulated that the neutrality axiom holds:

(A8) *Neutrality:* For $s, t \in D$, there are neutral elements e_s and e_t, and

$$e_s \otimes e_t \quad = \quad e_{s \cup t}. \tag{2.26}$$

DEFINITION **2.11**. *A valuation algebra satisfying the neutrality axiom is called valuation algebra with neutral elements.* ⊘

LEMMA **2.12**. (1) *Neutral elements are* unique *if they exist, i.e. if e_s, e'_s are neutral elements of Φ_s, then $e_s = e'_s$. In particular, the identity element is the neutral element of the empty domain,*

$$e_\emptyset = e. \tag{2.27}$$

(2) *Neutral elements are* idempotent*, i.e. $e_s \otimes e_s = e_s$.*

(3) *In a valuation algebra with neutral elements, if $d(\phi) = x$ and $y \subseteq x$, then*

$$\phi \otimes e_y = \phi. \tag{2.28}$$

⊘

PROOF. (1) By the definition of neutral elements and the commutativity of combination,

$$e_s = e'_s \otimes e_s = e_s \otimes e'_s = e'_s.$$

(2) By the definition of neutral elements, $e_s \otimes e_s = e_s$.

(3) By the definition of neutral elements and the neutrality axiom,

$$\phi \otimes e_y = \phi \otimes e_x \otimes e_y = \phi \otimes e_{x \cup y} = \phi \otimes e_x = \phi.$$ □

Some valuation algebras also have a null or absorbing element $z_s \in \Phi_s$ for every domain $s \in D$, i.e. an element z_s such that $z_s \otimes \phi = z_s$ for all valuations $\phi \in \Phi_s$. It represents contradictory information. It is postulated that the nullity axiom holds, which says that the null information cannot arise from non-contradictory information by marginalisation.

(A10) *Nullity:* For $\phi \in \Phi$, $x = d(\phi)$, $s \in \mathcal{M}(\phi)$, it holds that

$$\phi^{\downarrow s} = z_s \iff \phi = z_x. \tag{2.29}$$

DEFINITION **2.13**. *A valuation algebra with a null element z_s for every domain $s \in D$ satisfying the nullity axiom is called valuation algebra with null elements.* ⊘

LEMMA **2.14**. (1) *Null elements are* unique *if they exist, i.e. if* $z_s, z'_s \in \Phi_s$, *then* $z_s = z'_s$.

(2) *Null elements are* idempotent, *i.e.* $z_s \otimes z_s = z_s$ ⊘

PROOF. (1) By the definition of null elements and the commutativity of $\otimes$, it holds that $z_s = z_s \otimes z'_s = z'_s \otimes z_s = z'_s$.

(2) Follows by definition. □

2.4 Algebraic Theory

Some concepts of universal algebra (e.g. (Burris and Sankappanavar, 1981)) will now be applied to the two-sorted special case of valuation algebras. Particular attention has to be paid to the set lattice of domains. The results of (Kohlas, 2003) for full marginalisation are generalised to the case of partial marginalisation.

Homomorphisms and Embeddings

Let $\mathfrak{A}_1 = (\Phi_1, D, d_1, \otimes_1, \mathcal{M}_1, \downarrow^1)$ and $\mathfrak{A}_2 = (\Phi_2, D, d_2, \otimes_2, \mathcal{M}_2, \downarrow^2)$ be valuation algebras. Then, a mapping $h : \Phi_1 \to \Phi_2$ is called homomorphism if it is

1. compatible with labelling, i.e. $d_1(\phi) = d_2(h(\phi))$ for all $\phi \in \Phi_1$,
2. compatible with combination, i.e.
$$h(\phi \otimes_1 \psi) = h(\phi) \otimes_2 h(\psi)$$
for all $\phi, \psi \in \Phi_1$,
3. compatible with marginalisation, i.e.
$$\eta = h(\phi), \quad x \in \mathcal{M}_1(\phi) \quad \Longrightarrow \quad x \in \mathcal{M}_2(\eta), \quad h(\phi^{\downarrow x}) = (h(\phi))^{\downarrow^2 x}$$
for all $\phi \in \Phi_1$, and
4. the identity element e_1 of $\mathfrak{A}_1$ is mapped to the identity element e_2 of $\mathfrak{A}_2$, $h(e_1) = e_2$.

If a homomorphism h is injective, i.e. if
$$h(\phi) = h(\psi) \quad \Longrightarrow \quad \phi = \psi,$$
then h is called an embedding. Here, $\mathcal{M}_1(\phi) \subseteq \mathcal{M}_2(h(\phi))$. If there is an embedding $e : \Phi_1 \to \Phi_2$, then $\mathfrak{A}_2$ is called an extension of $\mathfrak{A}_1$. An embedding (and the corresponding extension) are called weak if neutral elements in $\mathfrak{A}_1$ (if they exist and

except for the identity element) are not mapped to the corresponding neutral elements of $\mathfrak{A}_2$. A homomorphism is called surjective if for all $\eta \in \Phi_2$ there is a $\phi \in \Phi_1$ such that

$$h(\phi) = \eta.$$

A bijective (i.e. injective and surjective) embedding h is called an isomorphism if $\mathcal{M}_1(\phi) = \mathcal{M}_2(h(\phi))$ for all $\phi \in \Phi_1$. Then, the inverse mapping $h^{-1} : \Phi_2 \to \Phi_1, h^{-1}(h(\phi)) \mapsto \phi$ exists and is also an embedding. The valuation algebra $\mathfrak{A}_1$ is called a subalgebra of $\mathfrak{A}_2$ if

- $\Phi_1 \subseteq \Phi_2$,
- $d_1(\phi) = d_2(\phi)$ for $\phi \in \Phi_1$,
- $\phi_1 \otimes_1 \phi_2 = \phi_1 \otimes_2 \phi_2$ for $\phi_1, \phi_2 \in \Phi_1$,
- $M_1(\phi) \subseteq \mathcal{M}_2(\phi)$ for $\phi \in \Phi_1$ and $\phi^{\downarrow_1 s} = \phi^{\downarrow_2 s}$ for $s \in \mathcal{M}_1(\phi)$.

In these terms, an embedding is always an isomorphism with a subalgebra.

Congruences

A binary relation θ in a set A, i.e. a set $\theta \subseteq A \times A$, is called an equivalence relation if it is

1. *reflexive*: $a \in A$ implies $(a, a) \in \theta$,
2. *symmetric*: $(a, b) \in \theta$ implies $(b, a) \in \theta$, and
3. *transitive*: $(a, b), (b, c) \in \theta$ imply $(a, c) \in \theta$.

The equivalence class of an element a modulo θ is denoted

$$[a]_\theta = \{b \in A : (a, b) \in \theta\}. \tag{2.30}$$

If $(a, b) \in \theta$, then a is said to be equivalent b modulo θ; the following equivalent notations will be used:

$$(a, b) \in \theta \iff a \equiv b \pmod{\theta} \iff a \equiv_\theta b \iff a \in [b]_\theta \iff b \in [a]_\theta. \tag{2.31}$$

The family of equivalence classes modulo θ is called quotient set and is denoted

$$A/\theta = \{[a]_\theta : a \in A\}. \tag{2.32}$$

It is well known that A/θ is a partition of A, i.e. that the classes in A/θ are disjoint and cover A. An equivalence relation θ in a valuation algebra $(\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ is called congruence if it is compatible with marginalisation and combination, i.e. if it is

1. compatible with combination: $\phi_1 \equiv \psi_1 \pmod{\theta}$, $\phi_2 \equiv \psi_2 \pmod{\theta}$ imply $\phi_1 \otimes \phi_2 \equiv \psi_1 \otimes \psi_2 \pmod{\theta}$;

2. compatible with marginalisation: if $\phi \equiv \psi \pmod{\theta}$ and $s \in \mathcal{M}(\phi), \mathcal{M}(\psi)$, then

$$[\phi^{\downarrow s}]_\theta = [\psi^{\downarrow s}]_\theta, \tag{2.33}$$

and

3. complete under marginalisation:

$$\phi^{\downarrow t} \equiv_\theta \psi,\ s \in \mathcal{M}(\psi) \implies \exists \psi' \in [\phi]_\theta \text{ s.t. } t, s \cap t \in \mathcal{M}(\psi'). \tag{2.34}$$

Remark **2.15**. A congruence θ carries the transitivity of marginalisation over to equivalence classes modulo θ. Equation (2.34) carries transitivity upwards from $[\phi^{\downarrow t}]_\theta$ to $[\phi]_\theta$. Let $\phi \equiv_\theta \psi \pmod{\theta}$, $t \in \mathcal{M}(\phi)$, $t' \in \mathcal{M}(\psi)$, $\phi' \in [\phi^{\downarrow t}]_\theta$, $\psi' \in [\phi^{\downarrow t'}]_\theta$, $s \in \mathcal{M}(\phi'), \mathcal{M}(\psi')$ and assume $s \subseteq t, t'$. If θ were not complete under marginalisation, it could not be proved that $\phi'^{\downarrow s} \equiv_\theta \psi'^{\downarrow s}$. This situation is shown in Figure 2.1 where equivalent valuations appear in the same box. However, since θ is complete under marginalisation, there are $\phi'' \in [\phi]_\theta$ and $\psi'' \in [\psi]_\theta$ with $s = s \cap t \in \mathcal{M}(\phi'')$ and $s = s \cap t' \in \mathcal{M}(\psi'')$. By the transitivity axiom, it also holds that $t \in \mathcal{M}(\phi'')$ and $t' \in \mathcal{M}(\psi'')$. Hence, since θ is compatible with marginalisation, it follows that

$$\phi'^{\downarrow s} \equiv_\theta \left(\phi''^{\downarrow t}\right)^{\downarrow s} = \phi''^{\downarrow s} \equiv_\theta \psi''^{\downarrow s} = \left(\psi''^{\downarrow t'}\right)^{\downarrow s} \equiv_\theta \psi'^{\downarrow s}.$$

This gives the situation in Figure 2.2. ⊘

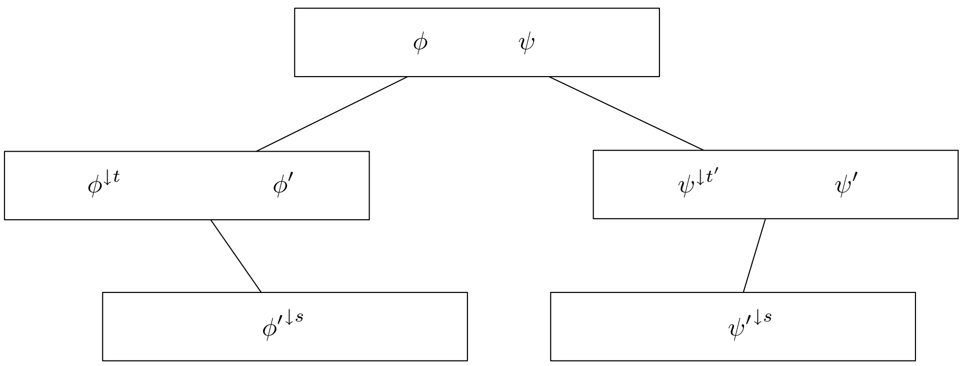


Figure 2.1: If θ is not complete under marginalisation, does it hold that $\phi'^{\downarrow s} \equiv_\theta \psi'^{\downarrow s}$?

According to this definition of congruence, equivalent valuations are not required to have the same domain. If, however,

$$\phi \equiv \psi \pmod{\theta} \implies d(\phi) = d(\psi), \tag{2.35}$$

then θ is called domain-contained. If θ is domain-contained, define

$$d_\theta([\phi]_\theta) = d(\phi), \tag{2.36}$$

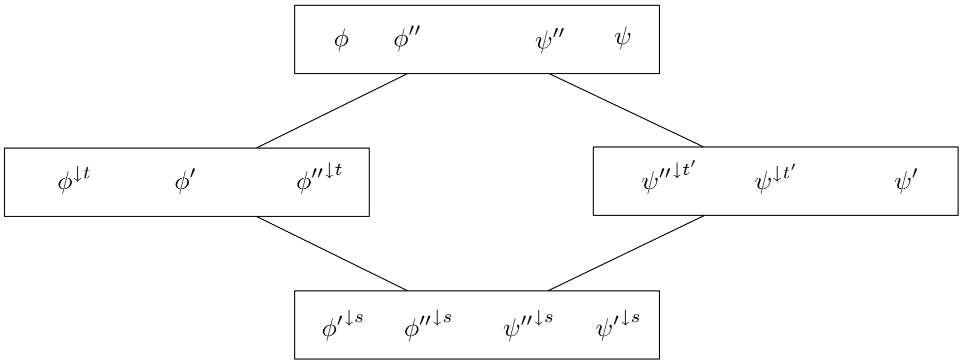


FIGURE 2.2: A congruence θ carries the transitivity of marginalisation over to equivalence classes modulo θ.

$$[\phi]_\theta \otimes_\theta [\psi]_\theta = [\phi \otimes \psi]_\theta, \tag{2.37}$$

and

$$[\phi]_\theta^{\downarrow_\theta s} = [\phi^{\downarrow s}]_\theta \tag{2.38}$$

if $s \in \mathcal{M}(\phi)$, i.e. the marginalisation is defined for domains

$$\mathcal{M}_\theta([\phi]_\theta) = \{s : \exists \phi' \in [\phi]_\theta \text{ s.t. } s \in \mathcal{M}(\phi')\}. \tag{2.39}$$

These definitions are well defined since θ is compatible with combination and marginalisation. These operations define a quotient valuation algebra as shown by the next theorem.

THEOREM **2.16**. *Let θ be a* domain-contained *congruence. Then,*

- $(\Phi/\theta, D, d_\theta, \otimes_\theta, \mathcal{M}_\theta, \downarrow_\theta)$ *is a valuation algebra;*
- *the mapping $\phi \mapsto [\phi]_\theta$ is a homomorphism.* ⊘

PROOF. The axioms are verified in turn.

(A1) The commutative semigroup property is inherited as is easily verifiable.

(A2) The labelling axiom is also inherited since

$$d_\theta([\phi]_\theta \otimes_\theta [\psi]_\theta) = d_\theta([\phi \otimes \psi]_\theta) = d(\phi \otimes \psi) = d(\phi) \cup d(\psi) = d_\theta([\phi]_\theta) \cup d_\theta([\psi]_\theta).$$

(A3) The marginalisation axiom holds since

$$d_\theta([\phi]_\theta^{\downarrow_\theta s}) = d_\theta([\phi^{\downarrow s}]_\theta) = d(\phi^{\downarrow s}) = s$$

for $s \in \mathcal{M}(\phi) \subseteq \mathcal{M}_\theta([\phi]_\theta)$.

(A4) Let $s \subseteq t \subseteq d_\theta([\phi]_\theta)$. On the one hand, assume $s \in \mathcal{M}(\phi) \subseteq \mathcal{M}_\theta([\phi_\theta]$. Then, the transitivity axiom implies that $t \in \mathcal{M}(\phi) \subseteq \mathcal{M}_\theta([\phi_\theta])$. On the other hand, assume $t \in \mathcal{M}(\phi) \subseteq \mathcal{M}_\theta([\phi]_\theta)$ and $s \in \mathcal{M}(\psi)$ for a $\psi \in [\phi^{\downarrow t}]_\theta = [\phi]^{\downarrow_\theta t}$. Using (2.34), there is a $\psi' \in [\phi]_\theta$ such that $s, t \in \mathcal{M}(\psi') \subseteq \mathcal{M}_\theta([\phi]_\theta)$. Then, using the transitivity axiom,

$$[\phi]_\theta^{\downarrow_\theta s} = [\psi'^{\downarrow s}]_\theta = [\psi'^{\downarrow t \downarrow s}]_\theta = [\psi'^{\downarrow t}]_\theta^{\downarrow_\theta s} = [\phi]_\theta^{\downarrow_\theta t \downarrow_\theta s}.$$

(A5) Let $[\phi]_\theta$ and $[\psi]_\theta$ with domains $x = d_\theta([\phi]_\theta) = d(\phi)$ and $y = d_\theta([\psi]_\theta) = d(\psi)$. Further let s such that $x \subseteq s \subseteq x \cup y$ and assume $s \cap d(\psi) \in \mathcal{M}(\psi) \subseteq \mathcal{M}_\theta([\psi]_\theta)$. Then, the combination axiom implies that $s \in \mathcal{M}(\phi \otimes \psi)$. Therefore, $s \in \mathcal{M}_\theta([\phi \otimes \psi]_\theta) = \mathcal{M}_\theta([\phi]_\theta \otimes_\theta [\psi]_\theta)$, and

$$\begin{aligned}([\phi]_\theta \otimes_\theta [\psi]_\theta)^{\downarrow_\theta s} &= [\phi \otimes \psi]_\theta^{\downarrow_\theta s} = [(\phi \otimes \psi)^{\downarrow s}]_\theta = [\phi \otimes \psi^{\downarrow s \cap d(\psi)}]_\theta \\ &= [\phi]_\theta \otimes_\theta [\psi^{\downarrow s \cap d(\psi)}]_\theta = [\phi]_\theta \otimes_\theta [\psi]_\theta^{\downarrow_\theta s \cap d_\theta([\psi]_\theta)}.\end{aligned}$$

(A6) The domain axiom is also inherited since $d(\phi) \in \mathcal{M}(\phi)$ implies $d_\theta([\phi]_\theta) = d(\phi) \in \mathcal{M}_\theta([\phi]_\theta)$ and

$$[\phi]_\theta^{\downarrow_\theta d_\theta([\phi]_\theta)} = [\phi]_\theta^{\downarrow_\theta d(\phi)} = [\phi^{\downarrow d(\phi)}]_\theta = [\phi]_\theta.$$

(A7) The element $e_\theta = [e]_\theta$ is an identity element since

$$[\phi]_\theta \otimes_\theta e_\theta = [\phi \otimes e]_\theta = [\phi]_\theta = [e \otimes \phi]_\theta = e_\theta \otimes_\theta [\phi]_\theta$$

and $d_\theta(e_\theta) = d(e) = \emptyset$.

It is readily verified that $\phi \mapsto [\phi]_\theta$ is a homomorphism. □

The following homomorphism theorem is a sort of dual of the previous quotient valuation algebra theorem. However, if marginalisation is only partially defined, then the homomorphism is required to satisfy an additional property in order to induce a congruence.

Theorem **2.17**. *Let $h : \Phi_1 \to \Phi_2$ be a homomorphism from the valuation algebra $\mathfrak{A}_1 = (\Phi_1, D, d_1, \otimes_1, \mathcal{M}_1, \downarrow^1)$ to $\mathfrak{A}_2 = (\Phi_2, D, d_2, \otimes_2, \mathcal{M}_2, \downarrow^2)$. Assume that*

$$s \subseteq t,\ s, t \in \mathcal{M}_2(h(\phi)) \implies \exists \phi' \text{ s.t. } h(\phi') = h(\phi), \text{ and } s, t \in \mathcal{M}_1(\phi'). \tag{2.40}$$

Then, the relation θ,

$$\phi \equiv_\theta \psi \iff h(\phi) = h(\psi), \tag{2.41}$$

is a domain-contained congruence in $\mathfrak{A}_1$. Furthermore, the mapping $h_\theta : \Phi_1/\theta \to \Phi_2, [\phi]_\theta \mapsto h(\phi)$ is an embedding. ⊘

Proof. The relation θ is an equivalence relation since it is

- reflexive: $h(\phi) = h(\phi)$ implies $\phi \equiv_\theta \phi$;

- symmetric: $\phi \equiv_\theta \psi \implies h(\phi) = h(\psi) \implies \psi \equiv_\theta \psi$; and
- transitive $\phi \equiv_\theta \psi, \psi \equiv_\theta \chi, \implies h(\phi) = h(\psi) = h(\chi) \implies \phi \equiv_\theta \chi$.

Furthermore, θ is compatible with combination since $h(\phi) = h(\phi')$ and $h(\psi) = h(\psi')$ imply

$$h(\phi \otimes_1 \psi) = h(\phi) \otimes_2 h(\psi) = h(\phi') \otimes_2 h(\psi') = h(\phi' \otimes_1 \psi').$$

It is also compatible with marginalisation since $h(\phi) = h(\phi')$ and $s \in \mathcal{M}_1(\phi), \mathcal{M}_1(\phi')$ imply that $s \in \mathcal{M}_2(h(\phi)) = \mathcal{M}_2(h(\phi'))$ and

$$h(\phi^{\downarrow^1 s}) = h(\phi)^{\downarrow^2 s} = h(\phi')^{\downarrow^2 s} = h(\phi'^{\downarrow^1 s}).$$

Since $h(\phi) = h(\phi')$ implies that $d_1(\phi) = d_2(h(\phi)) = d_2(h(\phi')) = d_1(\phi')$, θ is domain-contained.
It will now be proved that θ is complete under marginalisation. Assume $h(\phi^{\downarrow^1 t}) = h(\psi)$. Since θ is compatible with marginalisation, $t \in \mathcal{M}_2(h(\phi))$ and $h(\phi^{\downarrow^1 t}) = (h(\phi))^{\downarrow^2 t}$. If $s \in \mathcal{M}_1(\psi)$, then compatibility with marginalisation implies $s \in \mathcal{M}_2(h(\psi)) = \mathcal{M}_2(h(\phi^{\downarrow_1 t}))$. Hence, $s, t \in \mathcal{M}_2(h(\phi))$. Therefore, using the condition (2.40), there is a $\psi' \in \Phi_1$ such that $h(\psi') = h(\phi)$ and $s, t \in \mathcal{M}_1(\psi')$. Since θ is domain-contained, this shows that θ satisfies (2.34).
Finally, h_θ is injective since $h_\theta([\phi_1]_\theta) = h_\theta([\phi_2]_\theta)$ implies $h(\phi_1) = h(\phi_2)$, i.e. $[\phi_1]_\theta = [\phi_2]_\theta$. □

The situation of the Theorem is shown in Figure 2.3. If marginalisation in $\mathfrak{A}_2$ is fully defined, then the condition (2.40) can of course be dropped.

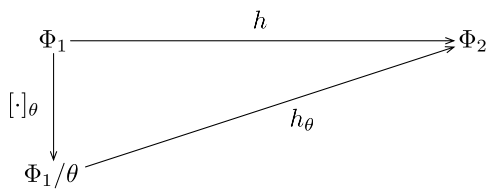


FIGURE 2.3: The homomorphism theorem

In order to keep the notation simple, the conventions of Table 2.1 are used for quotient valuation algebras induced by a congruence in an underlying valuation algebra.

2.5 Partial Order Induced by an Idempotent Congruence

A congruence θ in a valuation algebra $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ is called idempotent if

$$[\phi \otimes \phi^{\downarrow s}]_\theta = [\phi]_\theta \tag{2.42}$$

for all $\phi \in \Phi$, $s \in \mathcal{M}(\phi)$.

abbreviation	for
$(\Phi/\theta, D, d, \otimes, \mathcal{M}, \downarrow)$	$(\Phi/\theta, D, d_\theta, \otimes_\theta, \mathcal{M}_\theta, \downarrow_\theta)$
$d([\phi]_\theta)$	$d_\theta([\phi]_\theta)$
$[\phi]_\theta \otimes [\psi]_\theta$	$[\phi]_\theta \otimes_\theta [\psi]_\theta$
$\mathcal{M}([\phi]_\theta)$	$\mathcal{M}_\theta([\phi]_\theta)$
$[\phi]_\theta{}^{\downarrow s}$	$[\phi]_\theta{}^{\downarrow_\theta s}$

TABLE 2.1: Abbreviations for induced quotient valuation algebras working on quotients

REMARK **2.18**. For all $\phi \in \Phi$ the domain axiom shows that $d(\phi) \in \mathcal{M}(\phi)$ and $\phi^{\downarrow d(\phi)} = \phi$. Hence,

$$[\phi \otimes \phi]_\theta = [\phi \otimes \phi^{\downarrow d(\phi)}]_\theta = [\phi]_\theta.$$

Furthermore, the equivalence classes are semigroups, as $\phi \equiv \psi \pmod{\gamma}$ implies $\phi \otimes \psi \equiv \phi \otimes \phi \equiv \phi \pmod{\gamma}$. ⊘

Such a congruence induces a partial order $\leq$ between the equivalence classes by

$$[\phi]_\theta \leq [\psi]_\theta \quad \Longleftrightarrow \quad [\phi \otimes \psi]_\theta = [\psi]_\theta. \tag{2.43}$$

It has to be verified that $\leq$ is a partial order. Let $\phi, \psi \in \Phi$.

1. *Reflexivity:* $[\phi]_\theta \leq [\phi]_\theta$ since θ being idempotent implies that $[\phi \otimes \phi]_\theta = [\phi]_\theta$.
2. *Antisymmetry*: $[\psi]_\theta \leq [\phi]_\theta$ and $[\phi]_\theta \leq [\psi]_\theta$ imply $[\phi]_\theta = [\psi]_\theta$ since the two conditions and commutativity of combination imply
$$[\phi]_\theta = [\psi \otimes \phi]_\theta = [\phi \otimes \psi]_\theta = [\psi]_\theta.$$
3. *Transitivity:* $[\phi]_\theta \leq [\psi]_\theta$ and $[\psi]_\theta \leq [\zeta]_\theta$ imply $[\phi]_\theta \leq [\zeta]_\theta$. This holds since the two conditions and θ being compatible with marginalisation imply
$$[\zeta]_\theta = [\psi \otimes \zeta]_\theta = [(\phi \otimes \psi) \otimes \zeta]_\theta = [\phi \otimes (\psi \otimes \zeta)]_\theta = [\phi \otimes \zeta]_\theta,$$
thus indeed $[\phi]_\theta \leq [\zeta]_\theta$.

This partial order has the following properties.

LEMMA **2.19**. *Let θ be an idempotent congruence in a valuation algebra*

$$\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow),$$

and let $\phi, \psi, \psi' \in \Phi$, $x = d(\phi)$, $y = d(\psi)$, $y' = d(\psi')$.

(1) $[\phi]_\theta, [\psi]_\theta \leq [\phi \otimes \psi]_\theta$.

(2) *If $[\phi]_\theta \leq [\psi]_\theta$, then $[\phi \otimes \psi]_\theta = [\psi]_\theta$.*

(3) $(\Phi/\theta, \leq)$ *is a* join-semilattice *(i.e. there is a least upper bound* sup *or* $\wedge$ *for all pairs of elements) with*

$$\sup\{[\phi]_\theta, [\psi]_\theta\} = [\phi]_\theta \wedge [\psi]_\theta = [\phi \otimes \psi]_\theta. \tag{2.44}$$

(4) *If* $[\phi]_\theta \leq [\psi]_\theta = [\psi']_\theta$, *then*

$$[\phi \otimes \psi]_\theta = [\phi \otimes \psi']_\theta. \tag{2.45}$$

(5) *If* $[\phi]_\theta \leq [\psi]_\theta$, *then for all* $\psi' \in \Phi$

$$[\phi]_\theta \otimes [\psi']_\theta = [\phi \otimes \psi']_\theta \leq [\psi \otimes \psi']_\theta = [\psi]_\theta \otimes [\psi']_\theta. \tag{2.46}$$

(6) *If* $s \in \mathcal{M}(\phi)$, *then*

$$[\phi^{\downarrow s}]_\theta \leq [\phi]_\theta. \tag{2.47}$$

(7) *If* $s \in \mathcal{M}(\phi)$ *and* $[\phi^{\downarrow s}]_\theta = [\phi]_\theta$, $t \in \mathcal{M}(\phi)$, $s \subseteq t \subseteq x$, *then*

$$[\phi^{\downarrow t}]_\theta = [\phi]_\theta. \tag{2.48}$$

⊘

PROOF. (1), (2) are reformulations of the definition of $\leq$.

(3) By (1) $[\phi]_\theta, [\psi]_\theta \leq [\phi \otimes \psi]_\theta$. Assume there is a $\zeta \in \Phi$ such that $[\phi]_\theta, [\psi]_\theta \leq [\zeta]_\theta$. Then, by successive application of (2)

$$[\phi \otimes (\psi \otimes \zeta)]_\theta \quad = \quad [\phi \otimes \zeta]_\theta \quad = \quad [\zeta]_\theta,$$

which implies $[\phi \otimes \psi]_\theta \leq [\zeta]_\theta$. This shows that $[\phi \otimes \psi]_\theta$ is indeed the supremum of $[\phi]_\theta$ and $[\psi]_\theta$.

(4) By the definition of $\leq$, $[\phi \otimes \psi]_\theta = [\psi]_\theta = [\psi']_\theta = [\phi \otimes \psi']_\theta$.

(5) Since $[\psi]_\theta = [\phi \otimes \psi]_\theta$, $[\psi']_\theta = [\psi' \otimes \psi']_\theta$ and since θ is compatible with $\otimes$,

$$[\psi \otimes \psi']_\theta = [(\phi \otimes \psi) \otimes (\psi' \otimes \psi')]_\theta = [(\phi \otimes \psi') \otimes (\psi \otimes \psi')]_\theta,$$

i.e. $[\phi \otimes \psi']_\theta \leq [\psi \otimes \psi']_\theta$.

(6) Since θ is idempotent, $[\phi]_\theta = [\phi^{\downarrow s} \otimes \phi]_\theta$, i.e. $[\phi^{\downarrow s}]_\theta \leq [\phi]_\theta$.

(7) By the transitivity axiom, $s \in \mathcal{M}(\phi^{\downarrow t})$. Then, by (6),

$$[\phi]_\theta = [\phi^{\downarrow s}]_\theta = [(\phi^{\downarrow t})^{\downarrow s}]_\theta \leq [\phi^{\downarrow t}]_\theta \leq [\phi]_\theta.$$

Hence, $[\phi^{\downarrow t}]_\theta = [\phi]_\theta$ since $\leq$ is a partial order.

2.6 Stable Valuation Algebras

In many valuation algebras, marginals of neutral elements are again neutral elements, i.e. the marginal of a piece of vacuous information is vacuous again. This property is called *stability* (Kohlas, 2003).

(A9) *Stability:* For $s, t \in D$, $s \subseteq t$ there are neutral elements e_s and e_t such that $s \in \mathcal{M}(e_t)$ and

$$e_t^{\downarrow s} = e_s. \tag{2.49}$$

Definition **2.20**. *A valuation algebra that satisfies the neutrality and the stability axioms is called stable.* ⊘

Lemma **2.21**. *In a stable valuation algebra satisfying the nullity axiom, it holds that*

$$z_s \otimes e_t = z_{s \cup t}, \tag{2.50}$$

$$z_s \otimes z_t = z_{s \cup t}, \tag{2.51}$$

$$\phi \otimes z_t = z_{x \cup t}, \qquad x = d(\phi). \tag{2.52}$$

⊘

Proof. By stability, $s \cap t \in \mathcal{M}(e_t)$, thus by the combination axiom $s \in \mathcal{M}(z_s \otimes e_t)$ and

$$(z_s \otimes e_t)^{\downarrow s} = z_s \otimes e_t^{\downarrow s \cap t} = z_s \otimes e_{s \cap t} = z_s,$$

using equation (2.28). Hence, it follows by the nullity axiom that $z_s \otimes e_t = z_{s \cup t}$. Using equation (2.50) just proved and the idempotency of null elements, $z_s \otimes z_t = (z_s \otimes e_{s \cup t}) \otimes (z_t \otimes e_{s \cup t}) = z_{s \cup t} \otimes z_{s \cup t} = z_{s \cup t}$.
Finally, using equation (2.51), $\phi \otimes z_t = \phi \otimes z_t \otimes e_{x \cup t} = \phi \otimes z_{t \cup x} = \phi \otimes z_x \otimes z_t = z_x \otimes z_t = z_{x \cup t}$. □

Not every valuation algebra with neutral elements is stable, as has been noted by (Kohlas, 2003; p.21). However, if a valuation algebra does not contain neutral elements, it can be extended to a stable valuation algebra.

Lemma **2.22**. *Let* $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ *be a valuation algebra. Define*

- $\Phi' = \{(\phi, s) : \phi \in \Phi, s \in D, d(\phi) \subseteq s\}$,
- $d'(\phi, s) = s$,
- $(\phi, s) \otimes' (\psi, t) = (\phi \otimes \psi, s \cup t)$,
- $\mathcal{M}(\phi, s) = \{t \subseteq s : t \cap d(\phi) \in \mathcal{M}(\phi)\}$,
- $(\phi, s)^{\downarrow' t} = (\phi^{\downarrow t \cap d(\phi)}, t)$ *for* $t \cap d(\phi) \in \mathcal{M}(\phi)$.

Then, $\mathfrak{A}' = (\Phi', D, d', \otimes', \mathcal{M}', \downarrow')$ *is a stable valuation algebra with neutral elements* $e'_s = (e, s)$ *and it is a weak extension of* $\mathfrak{A}$ *by the (weak) embedding* $\phi \mapsto (\phi, d(\phi))$. ⊘

Proof. (A1) Combination in $\mathfrak{A}'$ is clearly associative and commutative.

(A2) The labelling axiom holds in $\mathfrak{A}'$ since

$$d'((\phi, x) \otimes' (\psi, y)) = x \cup y = d'(\phi, x) \cup d'(\psi, y)$$

for $(\phi, x), (\psi, y) \in \Phi'$.

(A3) The marginalisation axiom holds in $\mathfrak{A}'$ since $d'((\phi, x)^{\downarrow' s}) = d'(\phi^{\downarrow s \cap d(\phi)}, s) = s$ for $s \in \mathcal{M}'(\phi, x) \iff s \cap d(\phi) \in \mathcal{M}(\phi)$.

(A4) Let $(\phi, x) \in \Phi'$ and let $s \subseteq t \subseteq x$. Observe that $s \cap d(\phi) = s \cap (t \cap d(\phi))$, hence

$$\left[t \cap d(\phi) \in \mathcal{M}(\phi) \text{ and } s \cap d(\phi) \in \mathcal{M}(\phi^{\downarrow t \cap d(\phi)})\right] \iff s \cap d(\phi) \in \mathcal{M}(\phi)$$

by the transitivity axiom in $\mathfrak{A}$, thus

$$\left[t \in \mathcal{M}'(\phi, x) \text{ and } s \in \mathcal{M}'((\phi, x)^{\downarrow' t}) = \mathcal{M}'(\phi^{\downarrow t \cap d(\phi)}, t)\right] \iff s \in \mathcal{M}'(\phi, x).$$

Therefore, if $s \in \mathcal{M}'(\phi, x)$,

$$((\phi, x)^{\downarrow' t})^{\downarrow' s} = (\phi^{\downarrow t \cap d(\phi)}, t)^{\downarrow' s} = (\phi^{\downarrow s \cap (t \cap d(\phi))}, s) = (\phi^{\downarrow s \cap d(\phi)}, s) = (\phi, x)^{\downarrow' s}.$$

(A5) Let $(\phi, x), (\psi, y) \in \Phi'$ and $x \subseteq s \subseteq x \cup y$, and assume $s \in \mathcal{M}'(\psi)$, i.e. $s \cap d(\psi) \in \mathcal{M}(\psi)$. Observe that $d(\phi) \subseteq x \subseteq s$ implies $d(\phi) \subseteq s \cap (d(\phi) \cup d(\psi))$. Hence, by the combination axiom in $\mathfrak{A}$, $s \cap (d(\phi) \cup d(\psi)) \in \mathcal{M}(\phi \otimes \psi)$, i.e. $s \in \mathcal{M}'(\phi \otimes \psi, x \cup y)$. Then,

$$\begin{aligned}
((\phi, x) \otimes' (\psi, y))^{\downarrow' s} &= (\phi \otimes \psi, x \cup y)^{\downarrow' s} \\
&= ((\phi \otimes \psi)^{\downarrow s \cap (d(\phi) \cup d(\psi))}, s) \\
&= (\phi \otimes \psi^{\downarrow s \cap d(\psi)}, x \cup (s \cap y)) \\
&= (\phi, x) \otimes' (\psi^{\downarrow s \cap d(\psi)}, s \cap y) \\
&= (\phi, x) \otimes' (\psi, y)^{\downarrow' s}.
\end{aligned}$$

(A6) Let $(\phi, x) \in \Phi'$. Then, the domain axiom in $\mathfrak{A}$ shows that $x \cap d(\phi) = d(\phi) \in \mathcal{M}(\phi)$, hence $x \in \mathcal{M}'(\phi, x)$. Further, $(\phi, x)^{\downarrow' x} = (\phi^{\downarrow x \cap d(\phi)}, x) = (\phi^{\downarrow d(\phi)}, x) = (\phi, x)$. This shows that the domain axiom also holds in $\mathfrak{A}'$.

(A7) The identity element in $\mathfrak{A}'$ is $e' = e_\emptyset = (e, \emptyset)$ since for $(\phi, x) \in \Phi'$ $(\phi, x) \otimes' e' = (\phi \otimes e, x \cup \emptyset) = (\phi, x) = (\phi \otimes e, x \cup \emptyset) = e' \otimes' (\phi, x)$.

(A8) The elements $e'_s = (e, s)$ are neutral elements for the domain s since $(\phi, s) \otimes' (e, s) = (\phi, s)$. The neutrality axiom follows since $e'_s \otimes' e'_t = (e, s) \otimes' (e, t) = (e \otimes e, s \cup t) = (e, s \cup t) = e'_{s \cup t}$ for $s, t \in D$.

(A9) Let $e'_s = (e, s) \in \Phi'$ and let $t \subseteq s$. Then, using the identity axiom in $\mathfrak{A}$, $t \cap d(e) = \emptyset \in \mathcal{M}(e)$ shows that $t \in \mathcal{M}'(e'_s)$ and ${e'_s}^{\downarrow' t} = (e, s)^{\downarrow' t} = (e^{\downarrow t \cap \emptyset}, t) = (e, t) = e'_t$.

Finally, $\mathfrak{A}'$ is a weak extension of $\mathfrak{A}$ since

- $d'(\phi, d(\phi)) = d(\phi)$,
- $(\phi, d(\phi)) \otimes' (\psi, d(\psi)) = (\phi \otimes \psi, d(\phi) \cup d(\psi))$,
- $s = s \cap d(\phi) \in \mathcal{M}(\phi) \iff s \in \mathcal{M}'(\phi, d(\phi))$,
- $(\phi, d(\phi))^{\downarrow' s} = (\phi^{\downarrow s}, s)$, and,
- if there are neutral elements $e_s \in \Phi$, then $(e_s, s) \otimes' e'_s = (e_s, s) \otimes' (e, s) = (e_s, s) \neq e'_s$ for $s \neq \emptyset$. □

In a stable valuation algebra, multiplying a valuation by a neutral element of a larger domain is called vacuous extension to that domain,

$$\phi^{\uparrow y} = \phi \otimes e_y \tag{2.53}$$

for $d(\phi) \subseteq y \in D$. An illustration of vacuous extension will be given in Section 3.4 in the context of Gaussian potentials. Vacuous extension has the folllowing properties.

LEMMA **2.23**. *Let $(\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a stable valuation algebra and let ϕ, ψ be valuations on domains x and y, respectively, such that*

$$\phi^{\uparrow x \cup y} = \psi^{\uparrow x \cup y}. \tag{2.54}$$

Then, the following properties hold.

(1) *Let $s, u \in D$ such that $u \supseteq x$ and $s \subseteq u$. Then, $s \cap x \in \mathcal{M}(\phi)$ implies $u \in \mathcal{M}(\phi^{\uparrow u})$ and*

$$\phi^{\uparrow u \downarrow s} = \phi^{\downarrow s \cap x \uparrow s}. \tag{2.55}$$

In particular, for $s = x$,

$$\phi^{\uparrow u \downarrow x} = \phi. \tag{2.56}$$

(2)

$$\phi^{\uparrow x \cup y} \otimes \psi^{\uparrow x \cup y} = \phi \otimes \psi \tag{2.57}$$

(3) *If $x \cap y \in \mathcal{M}(\phi), \mathcal{M}(\psi)$, then*

$$\phi^{\downarrow x \cap y} = \psi^{\downarrow x \cap y} \tag{2.58}$$

and

$$\phi^{\downarrow x \cap y \uparrow x} = \phi. \tag{2.59}$$

(4) *For $x \subseteq u \subseteq v \in D$,*

$$\phi^{\uparrow u \uparrow v} = \phi^{\uparrow v}. \tag{2.60}$$

(5) *For $x, y \subseteq u \in D$,*

$$\phi^{\uparrow u} = \psi^{\uparrow u} \implies \phi^{\uparrow x \cup y} = \psi^{\uparrow x \cup y}. \tag{2.61}$$

⊘

PROOF. (1) Using the neutrality axiom, $\phi \otimes e_u = \phi \otimes e_x \otimes e_{u-x} = \phi \otimes e_{u-x}$. Since $(s \cup x) \cap (u - x) = s \cap (u - x) \in \mathcal{M}(e_{u-x})$, the combination axiom shows that $s \cup x \in \mathcal{M}(\phi \otimes e_{u-x})$ and $(\phi \otimes e_{u-x})^{\downarrow s \cup x} = \phi \otimes {e_{u-x}}^{\downarrow s \cap (u-x)} = \phi \otimes e_{s \cap (u-x)}$, using stability. Furthermore, by the combination axiom, $s \cap x \in \mathcal{M}(\phi)$ implies $s \in \mathcal{M}(\phi \otimes e_{s \cap (u-x)})$ and $(\phi \otimes e_{s \cap (u-x)})^{\downarrow s} = \phi^{\downarrow x \cap s} \otimes e_{s \cap (u-x)} = \phi^{\downarrow x \cap s} \otimes e_s$. Hence, by the transitivity axiom, $s \in \mathcal{M}(\phi \otimes e_{u-x})$ and $(\phi \otimes e_u)^{\downarrow s} = \phi^{\downarrow x \cap s} \otimes e_s$. If $s = x$, then the domain axiom shows that $\phi^{\downarrow x \cap s} \otimes e_s = \phi \otimes e_x = \phi$.

(2) Using the neutrality axiom, $(\phi \otimes e_{x \cup y}) \otimes (\psi \otimes e_{x \cup y}) = (\phi \otimes \psi) \otimes e_{x \cup y} = \phi \otimes \psi$.

(3) By (1), $x \cap y \in \mathcal{M}(\phi^{\uparrow x \cup y}), \mathcal{M}(\psi^{\uparrow x \cup y})$ and $\phi^{\downarrow x \cap y} = \phi^{\uparrow x \cup y \downarrow x \cap y} = \psi^{\uparrow x \cup y \downarrow x \cap y} = \psi^{\downarrow x \cap y}$. Furthermore, using (1), the neutrality and the combination axioms,

$$\phi = \phi^{\uparrow x \cup y \downarrow x} = (\phi \otimes e_{x \cup y})^{\downarrow x} = (\phi \otimes e_{x-y})^{\downarrow x} = \phi^{\downarrow x \cap y} \otimes e_{x-y} = \phi^{\downarrow x \cap y} \otimes e_x.$$

(4) By the neutrality axiom $(\phi \otimes e_u) \otimes e_v = \phi \otimes e_v$.

(5) Since $(x \cup y) \cap u = x \cup y \in \mathcal{M}(e_u)$, the combination axiom implies that $x \cup y \in \mathcal{M}(\phi \otimes e_u)$ and $(\phi \otimes e_u)^{\downarrow x \cup y} = \phi \otimes {e_u}^{\downarrow x \cup y} = \phi \otimes e_{x \cup y}$ using the stability axiom. Similarly, it can be proved that $(\psi \otimes e_u)^{\downarrow x \cup y} = \psi \otimes e_{x \cup y}$. Hence, $\phi \otimes e_{x \cup y} = (\phi \otimes e_u)^{\downarrow x \cup y} = (\psi \otimes e_u)^{\downarrow x \cup y} = \psi \otimes e_{x \cup y}$. □

Property (3) states that the relevant information is contained in the smaller domain $x \cap y$ and that ϕ and ψ represent the same knowledge vacuously extended to their respective label. Therefore, define

$$\phi \equiv_\sigma \psi \iff \phi^{\uparrow x \cup y} = \psi^{\uparrow x \cup y}. \tag{2.62}$$

LEMMA **2.24**. *The relation σ is an equivalence relation in a stable valuation algebra. Furthermore, it is compatible with combination and marginalisation.* ⊘

PROOF. Let ϕ, ψ, χ be valuations on domains x, y, and z, respectively. Reflexivity follows since $\phi^{\uparrow x} = \phi^{\uparrow x}$. Symmetry follows since $\phi \equiv_\sigma \psi \implies \phi^{\uparrow x \cup y} = \psi^{\uparrow x \cup y} \implies \psi \equiv_\sigma \phi$. Assume $\phi \equiv_\sigma \psi$ and $\psi \equiv_\sigma \chi$. Then, by the transitivity of vacuous extension, Lemma 2.23 (4),

$$\left(\phi^{\uparrow x \cup y}\right)^{\uparrow x \cup y \cup z} = \left(\psi^{\uparrow x \cup y}\right)^{\uparrow x \cup y \cup z} = \left(\psi^{\uparrow y \cup z}\right)^{\uparrow x \cup y \cup z} = \left(\chi^{\uparrow y \cup z}\right)^{\uparrow x \cup y \cup z},$$

and by Lemma 2.23 (5) it follows that $\phi \equiv_\sigma \chi$. Hence, σ is also transitive. In order to prove that σ is compatible with combination, let $\phi \equiv_\sigma \phi'$ and $\psi \equiv_\sigma \psi'$ with $x = d(\phi)$, $x' = d(\phi')$, $y = d(\psi)$, and $y' = d(\psi')$. Then, using the transitivity of vacuous extension (Lemma 2.23 (4)),

$$\begin{aligned}
\phi^{\uparrow x \cup y \cup x' \cup y'} \otimes \psi^{\uparrow x \cup y \cup x' \cup y'} &= \phi^{\uparrow x \cup x' \uparrow x \cup y \cup x' \cup y'} \otimes \psi^{\uparrow y \cup y' \uparrow x \cup y \cup x' \cup y'} \\
&= \phi'^{\uparrow x \cup x' \uparrow x \cup y \cup x' \cup y'} \otimes \psi'^{\uparrow y \cup y' \uparrow x \cup y \cup x' \cup y'} \\
&= \phi'^{\uparrow x \cup y \cup x' \cup y'} \otimes \psi'^{\uparrow x \cup y \cup x' \cup y'}.
\end{aligned}$$

Hence, it follows from (5) of the same lemma that $\phi \otimes \psi \equiv_\sigma \phi' \otimes \psi'$.
In order to prove that σ is compatible with marginalisation, let $\phi \equiv_\sigma \phi'$, $x = d(\phi)$ and $y = d(\phi')$ and $s \in \mathcal{M}(\phi), \mathcal{M}(\phi')$. By Lemma 2.23, $s \in \mathcal{M}(\phi^{\uparrow x \cup y})$ and $s \in \mathcal{M}(\psi^{\uparrow x \cup y})$ and

$$\phi^{\downarrow s} = \phi^{\uparrow x \cup y \downarrow s} = \psi^{\uparrow x \cup y \downarrow s} = \psi^{\downarrow s}. \qquad \square$$

In order for σ to be a congruence, an additional property is required.

DEFINITION **2.25**. *The congruence σ in a stable valuation algebra is called* closed under vacuous reduction *if $\phi \equiv_\sigma \psi$, $x = d(\phi)$, $y = d(\psi)$ imply that*

- *$x \cap y \in \mathcal{M}(\phi)$ and*
- *$s \in \mathcal{M}(\phi) \implies s \cap y \in \mathcal{M}(\phi^{\downarrow x \cap y})$.* ⊘

LEMMA **2.26**. *Assume σ is closed under vacuous reduction. Let $\phi \equiv_\sigma \psi$ and $x = d(\phi)$ and $y = d(\psi)$. Then, the following properties hold:*

(1) *$x \cap y \in \mathcal{M}(\phi), \mathcal{M}(\psi)$ and*

$$\phi \equiv_\sigma \phi^{\downarrow x \cap y} = \psi^{\downarrow x \cap y} \equiv_\sigma \psi. \tag{2.63}$$

(2) *$s \cap x \in \mathcal{M}(\phi) \iff s \cap y \in \mathcal{M}(\psi)$ and*

$$\phi^{\downarrow s \cap x} \equiv_\sigma \psi^{\downarrow s \cap y}. \tag{2.64}$$

⊘

PROOF. (1) Since σ is closed under vacuous reduction, $x \cap y \in \mathcal{M}(\phi), \mathcal{M}(\psi)$. Then, (1) follows from Lemma 2.23 (3).

(2) Assume $s \cap x \in \mathcal{M}(\phi)$. Since σ is closed under vacuous reduction, $(s \cap x) \cap y \in \mathcal{M}(\phi^{\downarrow x \cap y}) = \mathcal{M}(\psi^{\downarrow x \cap y})$. Hence, by the transitivity axiom, $s \cap y \in \mathcal{M}(\psi)$. Since $\phi = \phi^{\downarrow x \cap y} \otimes e_{x-y}$, Lemma 2.4 and stability imply that $\phi^{\downarrow s \cap x} = \phi^{\downarrow s \cap x \cap y} \otimes e_{(x-y) \cap s}$. Similarly, $\psi^{\downarrow s \cap y} = \psi^{\downarrow s \cap x \cap y} \otimes e_{(y-x) \cap s}$. Since $\phi^{\downarrow x \cap y} = \psi^{\downarrow x \cap y}$, the transitivity axiom implies that $\phi^{\downarrow s \cap x \cap y} = \psi^{\downarrow s \cap x \cap y}$. Hence, indeed $\phi^{\downarrow s \cap x} \equiv_\sigma \psi^{\downarrow s \cap y}$. $\square$

LEMMA **2.27**. *If σ is closed under vacuous reduction, it is a congruence.* ⊘

PROOF. It remains to be proved that σ is complete under marginalisation. Let $\phi^{\downarrow t} \equiv_\sigma \psi$ and $s \in \mathcal{M}(\psi)$. By Lemma 2.26 (2), $s = s \cap t \in \mathcal{M}(\phi^{\downarrow t})$, and hence $s = s \cap t \in \mathcal{M}(\phi)$ by the transitivity axiom. $\square$

2.7 Domain-Free Valuation Algebras

The congruence σ in a stable valuation algebra groups elements which represent the same information with respect to different domains. This motivates the derivation of a "quotient valuation algebra" of "domain-free" equivalence classes.

A domain $s \in D$ such that $s \cap d(\phi) \in \mathcal{M}(\phi)$ and $[\phi]_\sigma = [\phi^{\downarrow s \cap d(\phi)}]_\sigma$ is called a support of $[\phi]_\sigma$. In light of equation (2.63), define the least support of $[\phi]_\sigma$ by

$$\Delta[\phi]_\sigma = \bigcap_{\phi' \in [\phi]_\sigma} d(\phi') = \bigcap_{s \text{ support of } [\phi]_\sigma} s. \tag{2.65}$$

A variable $X \in d(\phi)$ is called vacuous in ϕ if $X \notin \Delta[\phi]_\sigma$. Define the reduct of ϕ as

$$\Delta\phi = \phi^{\downarrow \Delta[\phi]_\sigma}. \tag{2.66}$$

The reduct of ϕ is the marginal of ϕ where all vacuous variables are eliminated and ϕ is the vacuous extension of its reduct to x,

$$\phi = \Delta\phi \otimes e_x = \Delta\phi \otimes e_{x-d(\Delta\phi)}. \tag{2.67}$$

LEMMA **2.28**. *For x' such that $\Delta[\phi]_\sigma \subseteq x' \subseteq d(\phi) = x$, it holds that*

(1) *$x' \in \mathcal{M}(\phi)$ and $\phi^{\downarrow x'} \equiv_\sigma \phi$,*

(2) *$s \cap x \in \mathcal{M}(\phi) \iff s \cap x' \in \mathcal{M}(\phi^{\downarrow x'})$, and then*

$$\phi^{\downarrow s \cap x} \equiv_\sigma \phi^{\downarrow s \cap x'} = \phi^{\downarrow x' \downarrow s \cap x'}. \tag{2.68}$$

⊘

PROOF. (1) Using (2.67), the combination, stability and neutrality axioms, $\phi^{\downarrow x'} = \Delta\phi \otimes e_{x-x'} = \Delta\phi \otimes e_{x'} \equiv_\sigma \Delta\phi \equiv_\sigma \phi$.

(2) Since $\phi \equiv_\sigma \phi^{\downarrow x'}$ by (1), the claim follows from Lemma 2.26 (2) and the transitivity axiom. ☐

The quotient construction for the combination is the same as in the domain-contained case. Define

$$[\phi]_\sigma \otimes_\sigma [\psi]_\sigma = [\phi \otimes \psi]_\sigma. \tag{2.69}$$

This is well defined since σ is compatible with $\otimes$. Define focussing of $[\phi]_\sigma$ for domains

$$\mathcal{M}_\sigma([\phi]_\sigma) = \{s \in D : s \cap d(\phi) \in \mathcal{M}(\phi)\} \tag{2.70}$$

by

$$([\phi]_\sigma)^{\Rightarrow_\sigma s} = [\phi^{\downarrow s \cap d(\phi)}]_\sigma. \tag{2.71}$$

These definitions are sound in light of Lemma 2.26 (2).

The following lemma shows that the least support is non-increasing under focussing, i.e. that vacuous variables remain vacuous (if they are not eliminated).

LEMMA **2.29**. *For $s \in \mathcal{M}_\sigma([\phi]_\sigma)$, it holds that*

$$\Delta([\phi]_\sigma{}^{\Rightarrow s}) \subseteq \Delta[\phi]_\sigma. \tag{2.72}$$

⊘

PROOF. It follows from Lemma 2.28 that $s \cap d(\Delta\phi) \in \mathcal{M}(\Delta\phi)$ and $\phi^{\downarrow s} \equiv_\sigma \Delta\phi^{\downarrow s\cap d(\Delta\phi)}$. Hence, indeed

$$\Delta([\phi]_\sigma{}^{\Rightarrow s}) = \Delta[\phi^{\downarrow s}]_\sigma = \Delta[\Delta\phi^{\downarrow s\cap d(\Delta\phi)}]_\sigma \subseteq s \cap d(\Delta\phi) \subseteq d(\Delta\phi) = \Delta[\phi]_\sigma.$$ □

DEFINITION **2.30**. *Let Ψ be a set of valuations and D be a lattice of subsets of a set r of variables. Suppose that there are three operations defined:*

1. *Combination $\otimes : \Psi \times \Psi \to \Psi$; $(\eta, \zeta) \mapsto \eta \otimes \zeta$,*
2. *Domain $\mathcal{M} : \Psi \to 2^D$; $\eta \mapsto \mathcal{M}(\eta)$,*
3. *Focussing $\Rightarrow : \Psi \times D \to \Psi$; $(\eta, x) \mapsto \eta^{\Rightarrow x}$ defined for $x \in \mathcal{M}(\eta)$.*

The set $\mathcal{M}(\eta)$ contains all domains $x \in D$ which $\eta \in \Psi$ can be focussed on, i.e. all domains $x \in D$ for which $\eta^{\Rightarrow x}$ is well defined. Let the following set of axioms be imposed on Ψ and D:

(U1) *Commutative Semigroup: Ψ is associative and commutative under combination, i.e. for $\eta, \zeta, \chi \in \Psi$,*

$$\eta \otimes (\zeta \otimes \chi) = (\eta \otimes \zeta) \otimes \chi, \qquad \textit{and} \tag{2.73}$$

$$\eta \otimes \zeta = \zeta \otimes \eta. \tag{2.74}$$

(U2) *Transitivity: For $s, t \in D$, $\eta \in \Psi$,*

$$s \cap t \in \mathcal{M}(\eta) \iff t \in \mathcal{M}(\eta),\ s \in \mathcal{M}(\eta^{\Rightarrow t}); \tag{2.75}$$

then, it also holds that

$$(\eta^{\Rightarrow t})^{\Rightarrow s} = \eta^{\Rightarrow s\cap t}. \tag{2.76}$$

(U3) *Combination: If $\eta, \zeta \in \Psi$ with $s \in \mathcal{M}(\eta)$, $s \in \mathcal{M}(\zeta)$, then $s \in \mathcal{M}(\eta^{\Rightarrow s} \otimes \zeta)$ and*

$$(\eta^{\Rightarrow s} \otimes \zeta)^{\Rightarrow s} = \eta^{\Rightarrow s} \otimes \zeta^{\Rightarrow s}. \tag{2.77}$$

(U4) *Identity Element: There is a neutral element $e \in \Psi$ such that $\eta \otimes e = \eta = e \otimes \eta$ for all $\eta \in \Psi$. Furthermore, $\mathcal{M}(e) = D$ and for all $s \in D$*

$$e^{\Rightarrow s} = e. \tag{2.78}$$

(U5) *Support: All $\eta \in \Psi$ have a* support *$s \in D$, i.e. an $s \in \mathcal{M}(\eta)$ such that*

$$\eta^{\Rightarrow s} = \eta. \tag{2.79}$$

A quintuple $(\Psi, D, \otimes, \mathcal{M}, \Rightarrow)$ *satisfying these axioms is called a* **domain-free valuation algebra**. *If every* $\eta \in \Psi$ *can be focussed to any domain* $x \in D$, *then* $(\Psi, D, \otimes, \mathcal{M}, \Rightarrow)$ *is called a* **domain-free valuation algebra with full focussing**. ⊘

REMARK **2.31**. The identity element in a domain-free valuation algebra is *unique*. Assume e, e' are identity elements of Ψ. Then, $e = e \otimes e' = e' \otimes e = e'$. ⊘

The combination axiom of domain-free valuation algebras can be generalised in the following way.

LEMMA **2.32**. *Let* $(\Psi, D, \otimes, \mathcal{M}, \Rightarrow)$ *be a domain-free valuation algebra.*

(U5)′ *If* $\eta, \zeta \in \Psi$ *with* $s \in \mathcal{M}(\eta)$, $t \in \mathcal{M}(\zeta)$, *then* $s \subseteq t$ *implies that* $t \in \mathcal{M}(\eta^{\Rightarrow s} \otimes \zeta)$ *and*

$$(\eta^{\Rightarrow s} \otimes \zeta)^{\Rightarrow t} = \eta^{\Rightarrow s} \otimes \zeta^{\Rightarrow t}. \tag{2.80}$$

⊘

PROOF. The transitivity axiom (U2) implies that $t \in \mathcal{M}(\eta)$, $s \in \mathcal{M}(\eta^{\Rightarrow t})$ and $(\eta^{\Rightarrow s})^{\Rightarrow t} = \eta^{\Rightarrow s}$. Further, the combination axiom shows that $t \in \mathcal{M}((\eta^{\Rightarrow s})^{\Rightarrow t} \otimes \zeta)$ and

$$(\eta^{\Rightarrow s} \otimes \zeta)^{\Rightarrow t} = \left((\eta^{\Rightarrow s})^{\Rightarrow t} \otimes \zeta\right)^{\Rightarrow t} = (\eta^{\Rightarrow s})^{\Rightarrow t} \otimes \zeta^{\Rightarrow t} = \eta^{\Rightarrow s} \otimes \zeta^{\Rightarrow t}.$$ □

THEOREM **2.33**. *If* $(\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ *is a stable valuation algebra and if* σ *is closed under vacuous reduction, then* $(\Phi/\sigma, D, \otimes_\sigma, \mathcal{M}_\sigma, \Rightarrow_\sigma)$ *is a domain-free valuation algebra.* ⊘

PROOF. (U1) It is readily verified that $(\Phi/\sigma, \otimes_\sigma)$ inherits associativity and commutativity from $(\Phi, \otimes)$.

(U2) Let $x = d(\phi)$.
On the one hand, assume $t \in \mathcal{M}_\sigma([\phi]_\sigma)$ and $s \in \mathcal{M}_\sigma([\phi]_\sigma{}^{\Rightarrow t})$. This implies that $t \cap x \in \mathcal{M}(\phi)$, $[\phi]_\sigma^{\Rightarrow_\sigma t} = [\phi^{\downarrow t\cap x}]_\sigma$ and that there is a $\psi \in [\phi]_\sigma{}^{\Rightarrow t}$ such that $s \in \mathcal{M}(\psi)$. Using Lemma 2.26 (2), it follows that $s \cap (t \cap x) \in \mathcal{M}(\phi^{\downarrow t\cap x})$. By the transitivity axiom, $s \cap t \cap x \in \mathcal{M}(\phi)$, and thus $s \cap t \in \mathcal{M}_\sigma([\phi]_\sigma)$.
On the other hand, assume $s \cap t \in \mathcal{M}_\sigma([\phi]_\sigma)$. This implies that $s \cap t \cap x \in \mathcal{M}(\phi)$. By the transitivity axiom, $t \cap x \in \mathcal{M}(\phi)$ and $s \cap t \cap x \in \mathcal{M}(\phi^{\downarrow t\cap x})$. Hence, $t \in \mathcal{M}_\sigma([\phi]_\sigma)$ and $s \cap t \in \mathcal{M}_\sigma([\phi^{\downarrow t\cap x}]_\sigma) = \mathcal{M}_\sigma([\phi]_\sigma^{\Rightarrow_\sigma t})$.
Then, it holds that

$$([\phi]_\sigma^{\Rightarrow_\sigma t})^{\Rightarrow_\sigma s} = [\phi^{\downarrow t\cap x \downarrow s\cap t\cap x}]_\sigma = [\phi^{\downarrow s\cap t\cap x}]_\sigma = [\phi]_\sigma^{\Rightarrow_\sigma s\cap t}.$$

(U3) Let $[\phi]_\sigma$ and $[\psi]_\sigma$ and $s \subseteq t$ such that $s \in \mathcal{M}_\sigma([\phi]_\sigma)$ and $t \cap d(\psi) \in \mathcal{M}(\psi) \subseteq \mathcal{M}_\sigma([\psi]_\sigma)$. Then, $[\phi]_\sigma^{\Rightarrow_\sigma s} \otimes_\sigma [\psi]_\sigma = [\phi^{\downarrow s\cap d(\phi)} \otimes \psi]_\sigma$. By the combination axiom, it holds that $t \cap ((s \cap d(\phi)) \cup d(\psi)) \in \mathcal{M}(\phi^{\downarrow s\cap d(\phi)} \otimes \psi)$, hence $t \in \mathcal{M}_\sigma([\phi^{\downarrow s\cap d(\phi)} \otimes \psi]_\sigma) = \mathcal{M}_\sigma([\phi]_\sigma^{\Rightarrow_\sigma s} \otimes_\sigma [\psi]_\sigma)$. Further,

$$\begin{aligned}([\phi]_\sigma^{\Rightarrow_\sigma s} \otimes_\sigma [\psi]_\sigma)^{\Rightarrow_\sigma t} &= [(\phi^{\downarrow s\cap d(\phi)} \otimes \psi)^{\downarrow t\cap((s\cap d(\phi))\cup d(\psi))}]_\sigma \\ &= [\phi^{\downarrow s\cap d(\phi)} \otimes \psi^{\downarrow t\cap d(\psi)}]_\sigma \\ &= [\phi]_\sigma^{\Rightarrow_\sigma s} \otimes_\sigma [\psi]_\sigma^{\Rightarrow_\sigma t}.\end{aligned}$$

(U4) $[e]_\sigma$ is an identity element since $[\phi]_\sigma \otimes [e]_\sigma = [\phi \otimes e]_\sigma = [\phi]_\sigma = [e \otimes \phi]_\sigma = [e]_\sigma \otimes_\sigma [\phi]_\sigma$. Furthermore, $s \cap d(e) = \emptyset \in \mathcal{M}(e)$ and $\Delta[e]_\sigma = \emptyset$ imply that $s \in \mathcal{M}_\sigma[e]_\sigma$ for all $s \in D$.

(U5) For all $\phi \in \Phi$, it holds that $d(\phi) \in \mathcal{M}(\phi)$ by the domain axiom. Therefore, $d(\phi) \in \mathcal{M}_\sigma([\phi]_\sigma)$ and $[\phi]_\sigma^{\Rightarrow_\sigma d(\phi)} = [\phi^{\downarrow d(\phi)}]_\sigma = [\phi]_\sigma$. This shows that $d(\phi)$ is a support of $[\phi]_\sigma$. □

From a domain-free valuation algebra $(\Psi, D, \otimes, \mathcal{M}, \Rightarrow)$, a labelled valuation algebra can be constructed. For this purpose, consider the set of pairs

$$\Psi^* = \{(\eta, x) : \quad \eta \in \Psi, \quad x \in \mathcal{M}(\eta), \quad \eta^{\Rightarrow x} = \eta\}. \tag{2.81}$$

These pairs can be considered as valuations, labelled by their support. The following lemma holds for these pairs.

Lemma **2.34**. (1) *If x is a support of $\eta \in \Psi$ and $y \supseteq x$, $y \in D$, then y is a support of η.*

(2) *If $y \in \mathcal{M}(\eta)$, $\eta \in \Psi$, then y is a support of $\eta^{\Rightarrow y}$.*

(3) *For $(\eta, x), (\zeta, y) \in \Psi^*$, it holds that $(\eta \otimes \zeta, x \cup y) \in \Psi^*$, i.e. $x \cup y$ is a support of $\eta \otimes \zeta$.* ⊘

Proof. (1) Assume x is a support of $\eta \in \Psi$ and $y \supseteq x$. Using the transitivity axiom, $y \cap x = x \in \mathcal{M}(\eta)$ implies that $y \in \mathcal{M}(\eta^{\Rightarrow x})$ and

$$\eta^{\Rightarrow y} = (\eta^{\Rightarrow x})^{\Rightarrow y} = \eta^{\Rightarrow x \cap y} = \eta^{\Rightarrow x} = \eta.$$

This shows that y is a support of η.

(2) Assume $y \in \mathcal{M}(\eta)$, $\eta \in \Psi$. Then, by (U2), $y \cap y = y \in \mathcal{M}(\eta^{\Rightarrow y})$ and

$$(\eta^{\Rightarrow y})^{\Rightarrow y} = \eta^{\Rightarrow y \cap y} = \eta^{\Rightarrow y},$$

so y is indeed a support of $\eta^{\Rightarrow y}$.

(3) Assume $(\eta, x), (\zeta, y) \in \Psi^*$. By (1), it follows that $x \cup y$ is a support of ζ. Hence, by (U3), it follows that $x \cup y \in \mathcal{M}(\eta^{\Rightarrow x} \otimes \zeta)$ and

$$(\eta^{\Rightarrow x} \otimes \zeta)^{\Rightarrow x \cup y} = \eta^{\Rightarrow x} \otimes \zeta^{\Rightarrow x \cup y} = \eta \otimes \zeta,$$

i.e. $x \cup y$ is a support of $\eta \otimes \zeta$. Therefore, $(\eta \otimes \zeta, x \cup y) \in \Psi^*$. □

Therefore, the following operations are well defined in (Ψ^*, D).

1. *Labelling*: For $(\eta, x) \in \Psi^*$ define

$$d(\eta, x) = x. \tag{2.82}$$

2. *Combination*: For $(\eta, x), (\zeta, y) \in \Psi^*$ define

$$(\eta, x) \otimes^* (\zeta, y) = (\eta \otimes \zeta, x \cup y) \tag{2.83}$$

3. *Marginalisation*: For $(\eta, x) \in \Psi^*$, define

$$\mathcal{M}^*(\eta, x) = \mathcal{M}(\eta) \cap 2^x \tag{2.84}$$

and for $y \in \mathcal{M}(\eta, x)$ define

$$(\eta, x)^{\downarrow^* y} = (\eta^{\Rightarrow y}, y). \tag{2.85}$$

THEOREM **2.35**. *$\mathfrak{A}^* = (\Psi^*, d, D, \otimes, \mathcal{M}, \downarrow)$ as defined by equations (2.82)-(2.85) is a stable labelled valuation algebra.* ⊘

PROOF. The axioms are verified in turn.

(A1) For $(\eta, x), (\zeta, y), (\chi, z) \in \Psi^*$ it holds that

$$(\eta, x) \otimes^* ((\zeta, y) \otimes^* (\chi, z)) = (\eta \otimes \zeta \otimes \chi, x \cup y \cup z) = ((\eta, x) \otimes (\zeta, y)) \otimes (\chi, z)$$

and

$$(\eta, x) \otimes^* (\zeta, y) = (\eta \otimes \zeta, x \cup y) = (\zeta \otimes \eta, y \cup x) = (\zeta, y) \otimes^* (\eta, x).$$

This shows that $(\Psi^*, \otimes^*)$ is a commutative semigroup.

(A2) The labelling axiom is satisfied by the definition of labelling and combination in $\mathfrak{A}$.

(A3) The marginalisation axiom is satisfied by the definition of marginalisation and labelling in $\mathfrak{A}$.

(A4) Let $s \subseteq t \subseteq x = d^*(\eta, x)$. On the one hand, assume $t \in \mathcal{M}^*(\eta, x)$ and $s \in \mathcal{M}^*((\eta, x)^{\downarrow^* t}) = \mathcal{M}^*(\eta^{\Rightarrow t}, t)$. Using the transitivity axiom, $s \cap t = s \in \mathcal{M}(\eta)$ implies that $s \in \mathcal{M}(\eta^{\Rightarrow t})$. This shows that $s \in \mathcal{M}^*(\eta, x)$. On the other hand, assume $s \in \mathcal{M}^*(\eta, x)$. Since $s = s \cap t \in \mathcal{M}(\eta)$, the transitivity axiom implies that $t \in \mathcal{M}(\eta)$ and $s \in \mathcal{M}(\eta^{\Rightarrow t})$. Thus, $t \in \mathcal{M}^*(\eta^{\Rightarrow t}, t) = \mathcal{M}^*((\eta, x)^{\downarrow^* t})$. Then, it holds that $(\eta, x)^{\downarrow^* t \downarrow^* s} = (\eta^{\Rightarrow t \Rightarrow s}, s) = (\eta^{\Rightarrow s \cap t}, s) = (\eta^{\Rightarrow s}, s) = (\eta, x)^{\downarrow^* s}$. This verifies the transitivity axiom in $\mathfrak{A}^*$.

(A5) Let $(\eta, x), (\zeta, y) \in \Psi^*$ with $x \subseteq s \subseteq x \cup y$ and assume $s \cap y \in \mathcal{M}^*(\zeta, y) \subseteq \mathcal{M}(\zeta)$. By (U2), $s \in \mathcal{M}(\zeta)$. By (U3), $s \in \mathcal{M}(\eta^{\Rightarrow x} \otimes \zeta)$. Hence,

$$(\eta^{\Rightarrow x} \otimes \zeta)^{\Rightarrow s} = (\eta^{\Rightarrow x} \otimes \zeta^{\Rightarrow y})^{\Rightarrow s} = \eta^{\Rightarrow x} \otimes \zeta^{\Rightarrow s \cap y}.$$

Therefore, observing that $s \cap y$ is a support of $\zeta^{\Rightarrow s \cap y}$ in light of Lemma 2.34 (2), the following is well defined:

$$\begin{aligned}((\eta, x) \otimes^* (\zeta, y))^{\downarrow^* s} &= (\eta^{\Rightarrow x} \otimes \zeta^{\Rightarrow y}, x \cup y)^{\downarrow^* s} = (\eta^{\Rightarrow x} \otimes \zeta^{\Rightarrow s \cap y}, s) \\ &= (\eta^{\Rightarrow x}, x) \otimes (\zeta^{\Rightarrow s \cap y}, s \cap y) = (\eta, x) \otimes^* (\zeta, y)^{\downarrow^* s \cap y}.\end{aligned}$$

This shows that the combination axiom holds in $\mathfrak{A}^*$.

(A6) By definition, $(\eta, x) \in \Psi^*$ implies that $x \in \mathcal{M}(\eta)$ and $\eta^{\Rightarrow x} = \eta$. This implies that $(\eta, x)^{\downarrow^* x} = (\eta^{\Rightarrow x}, x) = (\eta, x)$, so the domain axiom holds in $\mathfrak{A}^*$.

(A7) Define $e^* = (e, \emptyset) \in \Psi^*$. Then, $d^*(e^*) = \emptyset$ and $\mathcal{M}^*(e^*) = \mathcal{M}(e) \cap \{\emptyset\} = \{\emptyset\}$. For $(\eta, x) \in \Psi^*$, $(\eta, x) \otimes^* e^* = e^* \otimes^* (\eta, x) = (\eta, x)$ and $e^* \otimes^* e^* = e^*$. Hence, e^* is an identity element of $\mathfrak{A}^*$.

(A8) For $s \in D$, the elements $e_s^* = (e, s)$, are neutral elements since for $(\phi, s) \in \Psi^*$,

$$e_s^* \otimes^* (\phi, s) = (\phi, s).$$

The neutrality axiom holds since for $s, t \in D$

$$e_s^* \otimes^* e_t^* = (e, s \cup t) = e_{s \cup t}^*.$$

(A9) Furthermore, stability holds since for $s \subseteq t \in D$ it holds that $s \in 2^t \cap D = 2^t \cap \mathcal{M}(e) = \mathcal{M}(e_t^*)$ and

$$e_t^{*\downarrow s} = (e^{\downarrow s \cap \emptyset}, s) = e_s^*.$$ □

The following theorem shows that going from a stable labelled valuation (and closed under vacuous reduction) to its domain-free version and deriving a labelled valuation algebra from the domain-free yields an isomorphic valuation algebra. The situation is depicted in Figure 2.4.

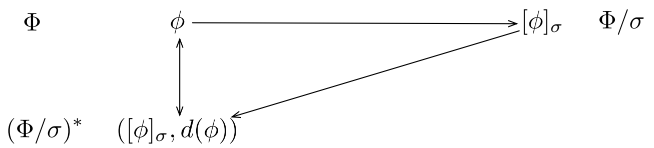


FIGURE 2.4: From a stable labelled valuation algebra Φ to a domain-free Φ/σ and back yields the labelled valuation algebra $(\Phi/\sigma)^*$ isomorphic to Φ.

THEOREM **2.36**. *Given a stable labelled valuation algebra* $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ *such that* σ *is complete under vacuous reduction. Let* $\mathfrak{A}_\sigma = (\Phi/\sigma, D, \otimes_\sigma, \mathcal{M}_\sigma, \Rightarrow_\sigma)$ *be the domain-free valulation algebra derived from* $\mathfrak{A}$. *Finally,* $\mathfrak{A}_\sigma$ *induces the labelled valuation algebra* $\mathfrak{A}^* = ((\Phi/\sigma)^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*)$. *Then, the mapping* $i : \Phi \to (\Phi/\sigma)^*$, $\phi \mapsto ([\phi]_\sigma, d(\phi))$ *is an isomorphism.* ⊘

PROOF. It has to be verified that i is homomorphism:

- $([\phi]_\sigma, d(\phi)) = i(\phi) = i(\psi) = ([\psi]_\sigma, d(\psi))$ implies $d(\phi) = d(\psi)$;
- $i(\phi \otimes \psi) = ([\phi \otimes \psi]_\sigma, d(\phi) \cup d(\psi) = i(\phi) \otimes^* i(\psi)$;
- on the one hand, $\mathcal{M}(\phi) \subseteq \mathcal{M}^*([\phi]_\sigma, d(\phi)) = \mathcal{M}^*(i(\phi))$; on the other hand, $s \in \mathcal{M}^*([\phi]_\sigma, d(\phi))$ implies $s \in \mathcal{M}_\sigma([\phi]_\sigma)$ and $s \in \mathcal{M}(\phi)$ by Lemma 2.26 (2), thus $s \in \mathcal{M}^*([\phi]_\sigma, d(\phi)) = \mathcal{M}^*(i(\phi))$; then, $i(\phi^{\downarrow s}) = ([\phi^{\downarrow s}]_\sigma, s) = (\phi, d(\phi))^{\downarrow^* s} = i(\phi)^{\downarrow^* s}$.

- $i(e) = (e, \emptyset) = e^*$.

It remains to be verified that i is a bijection. Assume $([\phi]_\sigma, d(\phi)) = ([\psi]_\sigma, d(\psi))$. This implies that $\phi = \phi \otimes e_{d(\phi)} = \psi \otimes e_{d(\psi)} = \psi$. Hence, i is injective. Let $([\phi]_\sigma, s) \in (\Phi/\sigma)^*$. Then, $\phi' = \Delta\phi \otimes e_s \in [\phi]_\sigma$ and $i(\phi') = ([\phi]_\sigma, s)$. Hence, i is surjective. □

2.8 Valuation Algebras with Division

Definition **2.37**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a valuation algebra and let γ be an* idempotent *congruence in it such that the equivalence classes $\gamma(\phi) = [\phi]_\gamma$ are groups with identity element $f_{\gamma(\phi)}$, i.e.*

$$\phi \otimes f_{\gamma(\phi)} = \phi; \tag{2.86}$$

the inverse of an element ϕ in its group $\gamma(\phi)$ is denoted $\phi^{-1} \in \gamma(\phi)$, i.e.

$$\phi \otimes \phi^{-1} = \phi^{-1} \otimes \phi = f_{\gamma(\phi)}. \tag{2.87}$$

Then, $\mathfrak{A}$ is called a valuation algebra with division *induced by γ.* ⊘

On the one hand, the whole semigroup (Φ, D) cannot be a group (if $D \neq \{\emptyset\}$) since e is a multiplicative identity and since $\phi \otimes \chi = e$ implies $d(\phi) = d(\chi) = \emptyset$ (in light of the labelling axiom). Therefore, Φ is decomposed into disjoint equivalence classes,

$$\Phi = \bigcup_{\phi \in \Phi} \{[\phi]_\gamma\}. \tag{2.88}$$

On the other hand, as seen in Section 2.5, the idempotency of γ induces a partial order,

$$\gamma(\phi) \leq \gamma(\psi) \iff \gamma(\psi) = \gamma(\phi \otimes \psi) \tag{2.89}$$

and in particular

$$\gamma(\phi^{\downarrow s}) \leq \gamma(\phi), \tag{2.90}$$

see Lemma 2.19. This property will be exploited in the Lauritzen-Spiegelhalter architecture for local computation, see Section 4.5. The following lemma gives some basic properties of valuation algebras with division.

Lemma **2.38**. (1) $f_{\gamma(\phi)} \otimes f_{\gamma(\psi)} = f_{\gamma(\phi \otimes \psi)}$

(2) $(\phi \otimes \psi)^{-1} = \phi^{-1} \otimes \psi^{-1}$.

(3) *If $\gamma(\psi) \leq \gamma(\phi)$, then*

$$\phi \otimes f_{\gamma(\psi)} = \phi \tag{2.91}$$

(4) $\phi \otimes \chi \otimes \phi = \phi$ *and* $\chi \otimes \phi \otimes \chi = \chi$ *imply* $\chi = \phi^{-1}$. ⊘

Proof. (1) Since γ is a congruence, $f_{\gamma(\phi)} \otimes f_{\gamma(\psi)} \in \gamma(\phi \otimes \psi)$. Using the commutativity of $\otimes$, $(\phi \otimes \psi) \otimes (f_{\gamma(\phi)} \otimes f_{\gamma(\psi)}) = (\phi \otimes f_{\gamma(\phi)}) \otimes (\psi \otimes f_{\gamma(\psi)}) = \phi \otimes \psi$, hence $f_{\gamma(\phi)} \otimes f_{\gamma(\psi)}$ is an identity element in $\gamma(\phi \otimes \psi)$. Since identity elements in groups are unique, it follows that $f_{\gamma(\phi)} \otimes f_{\gamma(\psi)} = f_{\gamma(\phi\otimes\psi)}$.

(2) Observe that $\phi^{-1} \otimes \psi^{-1} \in \gamma(\phi \otimes \psi)$. Hence, $(\phi \otimes \psi) \otimes (\phi^{-1} \otimes \psi^{-1}) = (\phi \otimes \phi^{-1}) \otimes (\psi \otimes \psi^{-1}) = f_{\gamma(\phi)} \otimes f_{\gamma(\psi)} = f_{\gamma(\phi\otimes\psi)}$ in light of (1). Since inverses are unique in a group, it follows that $(\phi \otimes \psi)^{-1} = \phi^{-1} \otimes \psi^{-1}$.

(3) Using (1) and $\gamma(\phi \otimes \psi) = \gamma(\phi)$,

$$\phi \otimes f_{\gamma(\psi)} = \phi \otimes f_{\gamma(\phi)} \otimes f_{\gamma(\psi)} = \phi \otimes f_{\gamma(\phi\otimes\psi)} = \phi \otimes f_{\gamma(\phi)} = \phi.$$

(4) Since $\phi \otimes \phi \equiv_\gamma \phi$,

$$\gamma(\phi) = \gamma(\phi \otimes \chi \otimes \phi) = \gamma((\phi \otimes \phi) \otimes \chi) = \gamma(\phi \otimes \chi);$$

similarly, $\gamma(\chi \otimes \phi) = \gamma(\chi)$. Hence, $\gamma(\chi) \leq \gamma(\phi)$ and $\gamma(\phi) \leq \gamma(\chi)$. Since $\leq$ is a partial order, $\gamma(\phi) = \gamma(\chi)$. Therefore, $\phi \otimes \chi = \phi \otimes \chi \otimes f_{\gamma(\chi)} = \phi \otimes \chi \otimes \phi \otimes \phi^{-1} = \phi \otimes \phi^{-1} = f_{\gamma(\phi)}$. Since inverses are unique in the group $\gamma(\phi) = \gamma(\chi)$, it follows that $\chi = \phi^{-1}$. □

There are several sufficient conditions which allow to introduce division in a valuation algebra with *full marginalisation*, see (Kohlas, 2003; Pouly, 2008). For instance, if it holds that $\phi \otimes \phi^{\downarrow t} = \phi$ for all $t \subseteq d(\phi)$, then the valuation algebra is called idempotent or an information algebra (Kohlas, 2003). In this case, $[\phi]_\gamma = \{\phi\}$ is a trivial idempotent congruence. Furthermore, since $\phi \otimes \phi = \phi$, ϕ is at the same time its own inverse and also the identity element of the group $[\phi]_\gamma$. There are many information algebras (Kohlas, 2003), for instance propositional logic, relational databases, and systems of linear equations.

2.9 Examples

Example **2.39** (Relational Algebra). (Kohlas and Stärk, 1996) A *tuple* is a function f which associates a value $f(X) \in \Omega_X$ to each variable X of a finite set x; the set x is called the *domain* of f, denoted $d(f) = x$. Let E_x be the set of all tuples of domain x. For a tuple f of domain x, the restriction of f to a subset $y \subseteq x$ is denoted $f[y] \in E_y$.
A *relation* R is a set of tuples of the same domain x, which is denoted $d(R) = x$. The projection of a relation R to $y \subseteq x = d(R)$ is the relation

$$\pi_y(R) = \{f[y] : f \in R\}$$

and the join of two relations R_1 and R_2 of domains x and y is the relation

$$R_1 \bowtie R_2 = \{f \in E_{x\cup y} : f[x] \in R_1, f[y] \in R_2\}.$$

Relations with join and projection form a labelled valuation algebra. Furthermore, joining is idempotent, i.e. $R \bowtie R = R$, and the valuation algebra is stable since it has neutral elements E_x and $\pi_y(E_x) = E_y$ for $y \subseteq x$. ⊘

EXAMPLE **2.40** (PROBABILITY DENSITIES). (Kohlas, 2003) Let $x \subseteq r$ be a finite set of real-valued variables with configurations $\mathbf{x} : x \to \mathbb{R}$, i.e. mappings which associate a real value to each variable $X \in x$. Configurations will often be denoted by bold-faced letters $\mathbf{x}$ corresponding to a set x of variables. Alternatively, a configuration $\mathbf{x}$ will be regarded as an $|x|$-dimensional vector whose components are indexed by the variables in x, and the space of all configurations of x will be denoted $\mathbb{R}^x = x \to \mathbb{R}$. The only configuration of the empty set of variables is denoted $\diamond$; it denotes an "empty vector." Define $\mathbb{R}^{\emptyset} = \{\diamond\}$. A continuous, non-negative function f on $\mathbb{R}^x$ is called a *probability density* if its (Riemann) integral is finite,

$$\int_{\mathbf{x}\in\mathbb{R}^x} f(\mathbf{x})d\mathbf{x} < \infty.$$

If the integral equals 1, this definition corresponds to ordinary normalised probability densities. The benefit of this more general definition is that one needs not be bothered about renormalisation on combination as defined below.
For any finite $x \in D$ consider the set Φ_x of probability densities $f : \mathbb{R}^x \to \mathbb{R}$ with domain $d(f) = x$. If f is a probability density with domain x and $s \subseteq x$, then the marginal $f^{\downarrow s} : \mathbb{R}^s \to \mathbb{R}$ of f with respect to s is defined for $\mathbf{s} \in \mathbb{R}^s$ by

$$f^{\downarrow s}(\mathbf{s}) = \int_{\mathbf{t}\in\mathbb{R}^t} f(\mathbf{s},\mathbf{t})d\mathbf{t}, \qquad t = x - s.$$

If f and g are two probability densities on x and y, respectively, then the combination of the two probability densities is defined for configurations $\mathbf{z}$ of $s \cup t$ by

$$(f \otimes g)(\mathbf{z}) = f(\mathbf{z}^{\downarrow z}) \cdot g(\mathbf{z}^{\downarrow t}).$$

It can be shown that probability densities on finite sets of variables form a valuation algebra (Kohlas, 2003). The identity element of this valuation algebra is the constant $e(\diamond) = 1$. It has no neutral elements since $f(\mathbf{x}) = 1$ for all $\mathbf{x}$ is not a probability density (for $x \neq \emptyset$). However, it has a null elements $z_x(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^x$, $x \in D$, which are probability densities (at least according to this definition). $\oslash$

Chapter Synopsis & Discussion

The axioms of valuation algebras were initially introduced by (Shenoy and Shafer, 1990) motivated by probability networks and belief function propagation and shown to be applicable to other domains in (Shafer, 1991). Valuation algebras with partial variable elimination were axiomatised in (Kohlas, 2003) and with partial marginalisation in (Schneuwly et al., 2004). The transitivity axiom of the marginalisation variant had the following form:

(A4)″ *Transitivity:* If $\phi \in \Phi$ and $x \subseteq y \subseteq d(\phi)$, then $x \in \mathcal{M}(\phi)$ implies $x \in \mathcal{M}(\phi^{\downarrow y}) \wedge y \in \mathcal{M}(\phi)$ and

$$(\phi^{\downarrow y})^{\downarrow x} = \phi^{\downarrow x}.$$

However, the following objection can be raised against this definition:

- If $t, t' \in \mathcal{M}(\phi)$ and $s \in \mathcal{M}(\phi^{\downarrow t}), \mathcal{M}(\phi^{\downarrow t'})$, it cannot be derived from the axioms whether $\phi^{\downarrow t \downarrow s} = \phi^{\downarrow t' \downarrow s}$ since the linking intermediate "$\phi^{\downarrow s}$" may not be defined (i.e. s may not be in $\mathcal{M}(\phi)$).

This modification of the transitivity axiom is irrelevant if marginalisation is fully defined for all subsets.

If neutral elements are presupposed in the definition of a labelled valuation algebra as in (Kohlas, 2003), the combination can be weakened as follows:

(A5)$''$ If $\phi, \psi \in \Phi$ with $d(\phi) = x$, $d(\psi) = y$, then $x \cap y \in \mathcal{M}(\psi)$ implies $x \in \mathcal{M}(\phi \otimes \psi)$ and

$$(\phi \otimes \psi)^{\downarrow x} = \phi \otimes \psi^{\downarrow x \cap y}.$$

The stronger (A5) then follows from (A5)$''$ in the same way as (U5)$'$ follows from (U5) in the domain-free case (see Lemma 2.32).

The definition of a valuation algebra can be generalised by admitting a general lattice D of domains, which needs not be Boolean as a set lattice; see (Kohlas, 2003). Then, some of the proofs using the differences of domains (for instance Lemma 2.23) are not feasible. However, local computation in join trees as discussed in Chapter 4 can be generalised to local computation in so-called *Markov Trees* (Mellouli, 1988).

Most definitions and results of this chapter are generalisations to partial marginalisation and the stronger transitivity axiom from those in (Kohlas, 2003; Schneuwly et al., 2004). In contrast, Lemma 2.22 is a new result, stating that any valuation algebra can be extended to a stable one.

3

Valuation Algebra of Gaussian Potentials

In modelling, Gaussian densities are often used to describe unknown errors. This has two main reasons:

- On the one hand, the *central limit theorem* states that the distribution of the (relocated, rescaled and averaged) sum of a sufficiently large number of independent and identically-distributed random variables is approximately normal. This may often be a reasonable assumption for measurement errors.

- On the other hand, multivariate Gaussian densities have properties which make them a very *tractable* choice: Marginal distributions have a Gaussian density (obtained through integration), and the joint distribution of two independent random variables has a Gaussian density (obtained by multiplying and renormalising two Gaussian densities). Furthermore, Gaussian densities can be easily represented by their *mean vector* and their *concentration matrix*. Such pairs of a mean vector and a concentration matrix representing a Gaussian density are called *Gaussian potentials* (Kohlas, 2003; Kohlas and Monney, 2008). The operations of *marginalisation* and *combination* of Gaussian densities correspond to simple matrix operations on the associated Gaussian potentials. Furthermore, the algebraic structure of Gaussian potentials is a *valuation algebra* (Kohlas, 2003).

Alternatively, Gaussian densities can be represented by their mean vector and their *variance-covariance matrix* (i.e. the inverse of the concentration matrix). These pairs are called *extended matrices* (Dempster, 1990a; Liu, 1996a; 1999). The operations of marginalisation and combination can be carried over to extended matrices. Here, the matrix operations for combination are called *sweepings*. Swept matrices represent Gaussian densities conditioned on a fixed value for some variables.

Models such as Bayesian networks usually consist of conditional distributions. In the case of (unnormalised) discrete probability potentials, the valuation algebra already contains conditional distributions. However, the algebra of continuous probability densities does not contain conditional distributions since the integral over all variables including the conditioning ones does not exist. Therefore, this chapter serves as a preparation for the discussion of conditional Gaussian densities in Part II and as an illustration of valuation algebras.

Chapter Outline

In Section 3.1, the terminology and notation of real-valued variables, vectors and matrices are given. In contrast to the standard definition, columns and rows may be indexed by variables. In Section 3.2, Gaussian potentials and the operations of combination and marginalisation are formally defined and shown to correspond to integration and normalised multiplication of Gaussian densities. A more efficient formulas of marginalisation in terms of the concentration matrix is presented. In Section 3.3, it is shown that the algebraic structure of Gaussian potentials is a valuation algebra. A geometric interpretation of vacuously extended Gaussian potentials is given in Section 3.4. Finally, moment matrices are introduced in Section 3.5. In contrast to Gaussian potentials, they represent Gaussian densities by their variance-covariance matrix. The combination rule of moment matrices can then be used to derive a more efficient way of computing the combined mean of Gaussian potentials.

3.1 Terminology and Notation

Variables and Lattice of Domains

Gaussian potentials always refer to a finite set x of variables. Since finite sets are closed under set union and intersection, they form a lattice D, whose elements are called domains. Such a domain $x \in D$ of variables of cardinality $|x| = n > 0$ will often be indexed by a bijective *indexation* $\mathcal{I} : \{1, \ldots, n\} \to x$. Posing

$$X_i = \mathcal{I}(i),$$

then

$$x = \{X_1, \ldots, X_n\}$$

without reference to the particular indexation. Variables will be designated by (possibly indexed) capital letters, sets of variables by lower-case letters.

Frames and Configurations

Variables may be relative to a frame. Intuitively, a variable can take a value in its frame. The frame associated with a variable X is denoted $\Theta(X)$. A configuration $\mathbf{x}$ associates to every variable X in a set x of variables the value $\mathbf{x}(X) \in \Theta(X)$ out of its frame. The only configuration of the empty set is denoted $\diamond$, which is not specified further. Configurations will be designated by the bold-faced lower-case letter corresponding to its set.

Real Vectors

Gaussian potentials refer to real-valued variables. A configuration $\mathbf{x}$ of a set x of real-valued variables is called real vector,

$$\mathbf{x} : x \to \mathbb{R}$$

or

$$\mathbf{x} \in \mathbb{R}^x,$$

and x is called the *domain* of $\mathbf{x}$. Such a configuration $\mathbf{x} \in \mathbb{R}^x$ associates to every variable X in its domain x a real value $\mathbf{x}(X) \in \mathbb{R}$ which can be thought of as the "component" of the real vector $\mathbf{x}$ corresponding to the "index" X. Notice that real vectors according to this definition are "unordered", i.e. there is no a priori ordering of the variables of the domain x in a vector $\mathbf{x} \in \mathbb{R}^x$. The real vectors of a domain x of cardinality $n = |x|$ form an n-dimensional *vector space* over the field of real numbers. If $x \neq \emptyset$, then *vector addition* is defined variable-wise by

$$(\mathbf{x}_1 + \mathbf{x}_2)(X) = \mathbf{x}_1(X) + \mathbf{x}_2(X), \qquad \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^x, \quad X \in x$$

and *scalar multiplication* by

$$(\lambda \cdot \mathbf{x})(X) = \lambda \cdot (\mathbf{x}(X)), \qquad \lambda \in \mathbb{R}, \quad \mathbf{x} : \mathbb{R}^x, \quad X \in x;$$

the *zero* or *null vector* (or *additive identity*) is

$$0_x(X) = 0, \qquad X \in x.$$

If $x = \emptyset$, then $\mathbb{R}^\emptyset = \{\diamond\}$ is a trivial 0-dimensional vector space over the field $\mathbb{R}$ only consisting of $\diamond$. Here,

$$\diamond + \diamond = \diamond$$

and

$$\lambda \cdot \diamond = \diamond, \qquad \lambda \in \mathbb{R}.$$

Under an indexation $\mathcal{I} : \{1, \ldots, |x| = n\} \to x$, the value corresponding to the ith variable, $i \in \{1, \ldots, n\}$, is denoted

$$\mathbf{x}_i = \mathbf{x}(X_i) = \mathbf{x}(\mathcal{I}(i)),$$

which leads to the standard matrix notation

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} \in \mathbb{R}^n.$$

Real Matrices

Analagously, the sets

$$\mathbb{R}(x,y) = x \times y \to \mathbb{R}, \quad \mathbb{R}(m,n) = \{1,\ldots,m\} \times \{1,\ldots,n\} \to \mathbb{R}$$

and

$$\mathbb{R}(m,y) = \{1,\ldots,m\} \times y \to \mathbb{R}, \quad \mathbb{R}(x,n) = x \times \{1,\ldots,n\} \to \mathbb{R}$$

where $x, y \in D$, $m, n \in \{0, 1, \ldots\}$ are vector spaces whose elements are called real matrices over the field $\mathbb{R}$. Addition is defined component-wise,

$$(A + B)(X,Y) = A(X,Y) + B(X,Y), \qquad X \in x,\, Y \in y,$$

as well as scalar multiplication,

$$(\lambda \cdot A)(X,Y) = \lambda \cdot A(X,Y), \qquad X \in x,\, Y \in y.$$

The corresponding null elements (which are constant 0) are denoted $0_{x,y}$, $0_{m,n}$, $0_{m,y}$, $0_{x,n}$, respectively. If either x or y is empty or m or n equals 0, then

$$\mathbb{R}(x,y) = \mathbb{R}(m,n) = \mathbb{R}(m,y) = \mathbb{R}(x,n) = \{\diamond\}.$$

Let now x and y be either finite sets of variables or index sets $x = \{1,\ldots,m\}$, $y = \{1,\ldots,n\}$ for non-negative integers m, n. Then, the matrix-matrix product $A \cdot B$ or just AB and the matrix-vector product $A \cdot \mathbf{y}$ or just $A\mathbf{y}$ are defined for compatible domains, i.e. for matrices

$$A : x \times y \to \mathbb{R} \quad \text{and} \quad B : y \times z \to \mathbb{R},$$

by $AB \in \mathbb{R}(x,z)$,

$$(AB)(X,Z) = \begin{cases} \sum_{Y \in y} A(X,Y) \cdot B(Y,Z) & \text{if } y \neq \emptyset, \\ 0 & \text{if } y = \emptyset, \end{cases} \tag{3.1}$$

and for matrices and vectors

$$A : x \times y \to \mathbb{R} \quad \text{and} \quad \mathbf{y} : y \to \mathbb{R}$$

define $A\mathbf{y} \in \mathbb{R}^x$ in the same way by identifying a vector $\mathbf{z} \in \mathbb{R}^z$ and the matrix $\tilde{\mathbf{z}} \in \mathbb{R}(z,1)$ such that $\tilde{\mathbf{z}}(Z,1) = \mathbf{z}(Z)$ for all $Z \in z$.

The following definitions are directly carried over from standard definitions on integer-indexed matrices and vectors. The transpose of a real matrix $A \in \mathbb{R}(x,y)$ is the function $A' \in \mathbb{R}(y,x)$ defined by

$$A'(Y,X) = A(X,Y). \tag{3.2}$$

A matrix $A \in \mathbb{R}(x,x)$, $x \in D$, is symmetric if for all $X, Y \in x$, it holds that

$$A(X,Y) = A(Y,X),$$

or, equivalently, if $A = A'$. If $A \in \mathbb{R}(x, x)$ is a symmetric matrix and if x is a set of variables, then x is called the *domain* of A and denoted $d(A) = x$. Let $\langle \cdot, \cdot \rangle$ be the *standard scalar product* on the Euclidean space $\mathbb{R}^x$,

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1' \mathbf{x}_2 = \begin{cases} \sum_{X \in x} \mathbf{x}_1(X) \cdot \mathbf{x}_2(X) & \text{if } x \neq \emptyset, \\ 0 & \text{if } x = \emptyset, \end{cases}$$

for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^x$. A symmetric matrix $A \in \mathbb{R}(x, x)$ is positive definite if $\langle \mathbf{x}, A\mathbf{x} \rangle > 0$ for all $\mathbf{x} \in \mathbb{R}^x$, $\mathbf{x} \neq 0_x$. The configuration of the empty domain, $\diamond$, is trivially a symmetric positive definite real matrix since the only element of $\mathbb{R}^\emptyset$ is the null element $\diamond = 0_\emptyset$.

The identity matrix $I_x \in \mathbb{R}(x, x)$ for a finite domain x or $I_m \in \mathbb{R}(m, m)$ for a non-negative integer m is defined by

$$I_x(X, Y) = \begin{cases} 1 & \text{if } X = Y, \\ 0 & \text{else} \end{cases}, \qquad X, Y \in x$$

and

$$I_m(i, j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else} \end{cases}, \qquad i, j \in \{1, \ldots, m\}.$$

They are the multiplicative identities of $\mathbb{R}(x, x)$ and $\mathbb{R}(m, m)$, respectively. A similar notation $I_{m,x} \in \mathbb{R}(m, x)$ will be used for

$$I_{m,x}(i, X) = \begin{cases} 1 & \text{if } i = \mathcal{I}(X), \\ 0 & \text{else,} \end{cases} \qquad i \in \{1, \ldots, m\},\ X \in x,$$

if the reference to the particular indexation $\mathcal{I}$ is either clear from the context or irrelevant.

The determinant $\det(A)$ of a real square matrix $A \in \mathbb{R}(x, y)$ is the standard determinant of the square matrix $\tilde{A} \in \mathbb{R}(m, m)$ for $m = |x| = |y|$, defined by

$$\tilde{A}(i, j) = A(\mathcal{I}(i), \mathcal{I}(j)).$$

This is well defined since interchanging rows or columns in $\tilde{A}$ leaves the determinant unaltered. Therefore, $\det(A)$ does not depend on the particular indexation $\mathcal{I}$. If $\det(A) > 0$, then A is called regular or non-singular. If $A \in \mathbb{R}(x, y)$ is regular, then there is a unique function $A^{-1} \in \mathbb{R}(y, x)$, called its inverse, such that $AA^{-1} = I_x$ and $A^{-1}A = I_y$.

Projection, Vacuous Extension and Transport

Consider a finite domain $x \in D$. For a real vector $\mathbf{x} : x \to \mathbb{R}$, $x \in D$, the projection of $\mathbf{x}$ onto $s \subseteq x, s \in D$, denoted

$$\mathbf{x}^{\downarrow s} : s \to \mathbb{R},$$

is the restriction of $\mathbf{x}$ to $s \subseteq d(\mathbf{x}) = x$, i.e.

$$\mathbf{x}^{\downarrow s}(X) = \mathbf{x}(X)$$

for $X \in s$. The vacuous extension of $\mathbf{x}$ to $u \supseteq d(\mathbf{x}) = x$, $u \in D$, denoted $\mathbf{x}^{\uparrow u} : u \to \mathbb{R}$, is

$$\mathbf{x}^{\uparrow u}(X) = \begin{cases} \mathbf{x}(X) & \text{if } X \in x, \\ 0 & \text{if } X \in u - x. \end{cases}$$

For some (arbitrary) $z \in D$, the transport operator $\Rightarrow$ will be used,

$$\mathbf{x}^{\Rightarrow z} = \left(\mathbf{x}^{\downarrow x \cap z}\right)^{\uparrow z}.$$

If $z \subseteq x$ or $x \subseteq z$, this corresponds to projection and vacuous extension, respectively. Similarly, for a matrix $A \in \mathbb{R}(t_1, t_2)$, $t_1, t_2 \in D$, the projection of A to $s_1 \times s_2$, $s_1 \subseteq t_1$, $s_2 \subseteq t_2$, $s_1, s_2 \in D$, denoted

$$A^{\downarrow s_1, s_2} : s_1 \times s_2 \to \mathbb{R},$$

is the restriction of A to $s_1 \times s_2$, and the vacuous extension of A to $u_1 \times u_2, u_1 \supseteq t_1$, $u_2 \supseteq t_2$, $u_1, u_2 \in D$, denoted $A^{\uparrow u_1, u_2} : u_1 \times u_2 \to \mathbb{R}$,

$$A^{\uparrow u}(X_i, X_j) = \begin{cases} A(X_i, X_j) & \text{if } X_i \in t_1, X_j \in t_2, \\ 0 & \text{if } X_i \in u_1 - t_1 \text{ or } X_j \in u_2 - t_2. \end{cases}$$

If $s = s_1 = s_2$, resp. $u = u_1 = u_2$, then the alternative notation

$$A^{\downarrow s} = A^{\downarrow s,s},$$

resp.

$$A^{\uparrow u} = A^{\uparrow u,u}$$

will be used. Further, for a matrix $A \in \mathbb{R}(m, x)$ ($m \in \mathbb{N}$ and $x \in D$), the restriction to a subset $y \subseteq x$ will often be denoted $A^{\downarrow y} \in \mathbb{R}(m, x)$ without reference to the set indexing the rows. For some (arbitrary) $z_1, z_2 \in D$, define the transport operator $\Rightarrow$ by

$$A^{\Rightarrow z_1, z_2} = \left(A^{\downarrow t_1 \cap z_1, t_2 \cap z_2}\right)^{\uparrow z_1, z_2}.$$

Again, if , $z_1 \subseteq t_1$ and $z_2 \subseteq t_2$ or $t_1 \subseteq z_1$ and $t_2 \subseteq z_2$, this corresponds to projection and vacuous extension, respectively, and, if $z_1 = z_2$, then the shortcut notation

$$A^{\Rightarrow z_1} = A^{\Rightarrow z_2} = A^{\Rightarrow z_1, z_2}$$

will be used. The following lemma summarises some important properties of vacuous extension of real vectors and matrices. In particular, property (2) is called the transitivity of vacuous extension.

LEMMA **3.1**. *For $\mathbf{x} \in \mathbb{R}^x$ and $A \in \mathbb{R}(x, x)$ and $x \subseteq s \subseteq t$,*

(1) $\mathbf{x}^{\uparrow x} = \mathbf{x}$ *and* $A^{\uparrow x} = A$,

(2) $\mathbf{x}^{\uparrow s \uparrow t} = \mathbf{x}^{\uparrow t}$ *and* $A^{\uparrow s \uparrow t} = A^{\uparrow t}$,

(3) $\mathbf{x}^{\uparrow t \downarrow s} = \mathbf{x}^{\uparrow s}$ *and* $A^{\uparrow t \downarrow s} = A^{\uparrow s}$. ⊘

PROOF. The claim follows from the definition of vacuous extension. □

Partitioned Matrices and Vectors

Let $K \in \mathrm{IR}(x, y)$, $K_{11} \in \mathrm{IR}(x, x)$, $K_{12} \in \mathrm{IR}(x_2, x_1)$, $K_{22} \in \mathrm{IR}(x_2, x_2)$ be real matrices. Further, let $x_1 \cup x_2 = x$ and $y_1 \cup y_2 = y$ be disjoint subsets of domains x and y, $x_1 \cap x_2 = \emptyset = y_1 \cap y_2$. Then, K is said to be partitioned according to x_1 and x_2, denoted

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

if and only if

$$K(X, Y) = \begin{cases} K_{11}(X, Y) & \text{if } X \in x_1, Y \in x_1, \\ K_{12}(X, Y) & \text{if } X \in x_1, Y \in x_2, \\ K_{21}(X, Y) & \text{if } X \in x_2, Y \in x_1, \\ K_{22}(X, Y) & \text{if } X \in x_2, Y \in x_2. \end{cases}$$

The function K is unambiguously defined by the elements $K_{11}, K_{12}, K_{21}, K_{22}$ as long as the block rows and the block columns have disjoint labels. Of course, different arrangement of the block rows and block columns are possible:

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = K = \begin{pmatrix} K_{22} & K_{21} \\ K_{12} & K_{11} \end{pmatrix}.$$

The partitioning subsets may be empty; for instance, if x_2 or y_2 is empty, then $K = K_{11}$. If the rows or the columns of K are indexed by numbers, then the order is of course relevant. For instance, a matrix $K \in \mathrm{IR}(m, x)$ may be partitioned according to m_1 and m_2 rows and according to x_1 and x_2 into matrices $K_{ij} \in \mathrm{IR}(m_i, x_j)$ for $i, j \in \{1, 2\}$ such that $m_1 + m_2 = m$. This is denoted

$$\begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} = K = \begin{pmatrix} K_{12} & K_{11} \\ K_{22} & K_{21} \end{pmatrix}.$$

If one block is not specified, then it is assumed zero; for instance,

$$\begin{pmatrix} K_{11} & K_{12} \\ & K_{22} \end{pmatrix} = \begin{pmatrix} K_{11} & K_{12} \\ 0_{x_2, y_1} & K_{22} \end{pmatrix}.$$

In order to save space, columns of a real matrix may also be separated by a comma, for instance

$$\begin{pmatrix} K_{11} & K_{12} \end{pmatrix} = (K_{11}, K_{12}).$$

(Column) vectors $\mathbf{x} \in \mathrm{IR}^x$ and one-column matrices $\tilde{\mathbf{x}} \in \mathrm{IR}(x, 1)$ are often identified if $\mathbf{x}(X) = \tilde{\mathbf{x}}(X)$ for all $X \in x$. This is denoted $\mathrm{IR}^x = \mathrm{IR}(x, 1)$. Furthermore, if $\mathbf{x} = \tilde{\mathbf{x}}$ is partitioned according to x_1 and x_2 into $\mathbf{x}_1$ and $\mathbf{x}_2$, either the matrix notation

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}$$

or the shorter notation $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ is used. Since $\mathbf{x}_1$ and $\mathbf{x}_2$ have different row labels, there is no clash with the notation used for matrices partitioned into two column blocks. Also, all these definitions are easily generalised to more than two blocks of rows or columns.

LEMMA **3.2**. *For real matrices $A \in \mathbb{R}(s,t)$, $B \in \mathbb{R}(t,u)$ and $C \in \mathbb{R}(u,v)$ and any index sets $\tilde{s}, \tilde{v}$ it holds*

$$(ABC)^{\Rightarrow \tilde{s},\tilde{v}} = A^{\Rightarrow \tilde{s},t} B C^{\Rightarrow u,\tilde{v}}. \tag{3.3}$$

⊘

PROOF.

$$\begin{aligned}(ABC)^{\Rightarrow \tilde{s},\tilde{v}}(S,V) &= \begin{cases} (ABC)(S,V) & \text{if } S \in s, V \in v \\ 0 & \text{else} \end{cases} \\ &= \begin{cases} \sum_{T \in t, U \in u} A(S,T)B(T,U)C(U,V) & \text{if } S \in s, V \in v \\ 0 & \text{else} \end{cases} \\ &= A^{\Rightarrow \tilde{s},t} B C^{\Rightarrow u,\tilde{v}}(S,V).\end{aligned}$$

□

3.2 Definition of Gaussian Potentials

A Gaussian potential on a finite domain $x \in D$ is a pair $\phi = (\mu, K)$ where $\mu \in \mathbb{R}^x, K \in \mathbb{R}(x,x)$, K symmetric and positive definite. If $n = |x| > 0$, it represents the Gaussian density function

$$\phi_{\mu,K}(\mathbf{x}) = \sqrt{\frac{|\det(K)|}{(2\pi)^n}} e^{-\frac{1}{2}(\mathbf{x}-\mu)' K (\mathbf{x}-\mu)}, \quad \mathbf{x} \in \mathbb{R}^n, \tag{3.4}$$

with respect to Lebesgue measure λ^n. The vector μ is called mean vector, K its concentration matrix, and $\Sigma = K^{-1}$(variance-)covariance matrix. The unique Gaussian potential with domain $\emptyset$ is

$$e = \phi_{\diamond,\diamond} = (\diamond, \diamond),$$

representing the probability mass function

$$\phi_{\diamond,\diamond}(\diamond) = 1, \tag{3.5}$$

or, equivalently, the Dirac measure $\delta_\diamond$ at $\diamond$ on the σ-algebra $2^{\{\diamond\}} = \{\emptyset, \{\diamond\}\}$, i.e.

$$\delta_\diamond(\{\diamond\}) = 1, \quad \delta_\diamond(\emptyset) = 0. \tag{3.6}$$

Let $\mathcal{G}$ denote the set of all such Gaussian potentials $\phi = (\mu, K)$. The operation $d : \mathcal{G} \to D$ is called *labelling* and $d(\phi)$ is called the *domain* of ϕ.

Combination of Gaussian Potentials

First, the combination of Gaussian potentials is formally defined and then shown to correspond to the normalised product of the two corresponding Gaussian densities. Let $\phi = (\mu_1, K_1)$ and $\psi = (\mu_2, K_2)$ be Gaussian potentials Define their combination $\phi \otimes \psi = (\mu, K)$ by

$$K = {K_1}^{\uparrow u} + {K_2}^{\uparrow u}, \qquad \text{and} \tag{3.7}$$

$$\mu = K^{-1}({K_1}^{\uparrow u} \cdot {\mu_1}^{\uparrow u} + {K_2}^{\uparrow u} \cdot {\mu_2}^{\uparrow u}) \tag{3.8}$$

for $u = d(\phi) \cup d(\psi)$. The combination $\otimes$ is well defined since K is symmetric and positive definite in light of Lemma A.5. Further, ${K_1}^{\uparrow u} \cdot {\mu_1}^{\uparrow u} = (K_1 \cdot \mu_1)^{\uparrow u}$. Define

$$\begin{aligned} z &= d(\phi) \cap d(\psi), \\ x &= d(\phi) - d(\psi), \qquad \text{and} \\ y &= d(\psi) - d(\phi). \end{aligned}$$

Then, the concentration matrix and the mean vector of the combination can also be written as

$$K = \begin{pmatrix} {K_1}^{\downarrow z} + {K_2}^{\downarrow z} & {K_1}^{\downarrow z,x} & {K_2}^{\downarrow z,y} \\ {K_1}^{\downarrow x,z} & {K_1}^{\downarrow x} & 0_{x,y} \\ {K_2}^{\downarrow y,z} & 0_{y,x} & {K_2}^{\downarrow y} \end{pmatrix} \tag{3.9}$$

and

$$\mu = K^{-1} \begin{pmatrix} {K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + {K_2}^{\downarrow z}{\mu_2}^{\downarrow z} + {K_1}^{\downarrow z,x}{\mu_1}^{\downarrow x} + {K_2}^{\downarrow z,y}{\mu_2}^{\downarrow y} \\ {K_1}^{\downarrow x}{\mu_1}^{\downarrow x} + {K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} \\ {K_2}^{\downarrow y}{\mu_2}^{\downarrow y} + {K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z} \end{pmatrix}. \tag{3.10}$$

THEOREM **3.3**. *The combination $\phi \otimes \psi = (\mu, K)$ represents the density function*

$$(\phi \otimes \psi)(\mathbf{u}) = k \cdot \phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}), \tag{3.11}$$

where the normalisation constant

$$k = \frac{1}{\int_{\mathbf{u} \in \mathbb{R}^u} \phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}) d\mathbf{u}} \tag{3.12}$$

does not depend on $\mathbf{u} \in \mathbb{R}^u$. ⊘

PROOF. On the one hand,

$$(\phi \otimes \psi)(\mathbf{u}) = \sqrt{\frac{|\det(K)|}{(2\pi)^n}} \cdot e^{-\frac{1}{2}\left[\sum_{s,t \in \{x,y,z\}} (\mathbf{u}^{\downarrow s} - \mu^{\downarrow s})' K^{\downarrow s,t} (\mathbf{u}^{\downarrow t} - \mu^{\downarrow t})\right]} \tag{3.13}$$

for $n = |d(\phi) \cup d(\psi)|$. On the other hand,

$$\begin{aligned} &\phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}) \\ =&\sqrt{\frac{|\det(K_1)|}{(2\pi)^{n_1}}} e^{-\frac{1}{2}(\mathbf{u}^{\downarrow z \cup x} - \mu_1)' K_1 (\mathbf{u}^{\downarrow z \cup x} - \mu_1)} \\ &\cdot \sqrt{\frac{|\det(K_2)|}{(2\pi)^{n_2}}} e^{-\frac{1}{2}(\mathbf{u}^{\downarrow z \cup y} - \mu_2)' K_2 (\mathbf{u}^{\downarrow z \cup y} - \mu_2)} \\ =&k' \cdot e^{-\frac{1}{2}\left[\left\{\begin{array}{c} \sum_{s,t \in \{x,z\}} (\mathbf{u}^{\downarrow s} - {\mu_1}^{\downarrow s})' {K_1}^{\downarrow s,t} (\mathbf{u}^{\downarrow t} - {\mu_1}^{\downarrow t}) \\ + \sum_{s,t \in \{y,z\}} (\mathbf{u}^{\downarrow s} - {\mu_2}^{\downarrow s})' {K_2}^{\downarrow s,t} (\mathbf{u}^{\downarrow t} - {\mu_2}^{\downarrow t}) \end{array}\right\}\right]} \end{aligned} \tag{3.14}$$

for

$$k' = \sqrt{\frac{|\det(K_1)||\det(K_2)|}{(2\pi)^{n_1+n_2}}},$$

$n_1 = |d(\phi)|$, $n_2 = |d(\psi)|$. Then, the sum in the exponent of the right-hand side of equations (3.13) can be developed as follows:

$$\begin{aligned}
&\sum_{s,t\in\{x,y,z\}} (\mathbf{u}^{\downarrow s} - \mu^{\downarrow s})' K^{\downarrow s,t}(\mathbf{u}^{\downarrow t} - \mu^{\downarrow t}) \\
=& \sum_{s,t\in\{x,y,z\}} \mathbf{u}^{\downarrow s'} K^{\downarrow s,t} \mathbf{u}^{\downarrow t} \\
&- \sum_{s,t\in\{x,y,z\}} \mathbf{u}^{\downarrow s'} K^{\downarrow s,t} \mu^{\downarrow t} - \sum_{s,t\in\{x,y,z\}} \mu^{\downarrow s'} K^{\downarrow s,t} \mathbf{u}^{\downarrow t} \\
&+ k'' \\
=&\mathbf{u}' K \mathbf{u} \\
&- \mathbf{u}' K (K^{-1}(K_1{}^{\uparrow u} \mu_1{}^{\uparrow u} + K_2{}^{\uparrow u} \mu_2{}^{\uparrow u})) \\
&- (K^{-1}(K_1{}^{\uparrow u} \mu_1{}^{\uparrow u} + K_2{}^{\uparrow u} \mu_2{}^{\uparrow u}))' K \mathbf{u} \\
&+ k'' \\
=&\mathbf{u}' K_1{}^{\uparrow u} \mathbf{u} + \mathbf{u}' K_2{}^{\uparrow u} \mathbf{u} \\
&- \mathbf{u}' K_1{}^{\uparrow u} \mu_1{}^{\uparrow u} - \mathbf{u}' K_2{}^{\uparrow u} \mu_2{}^{\uparrow u} \\
&- \mu_1{}^{\uparrow u'} K_1{}^{\uparrow u} \mathbf{u} - \mu_2{}^{\uparrow u'} K_2{}^{\uparrow u} \mathbf{u} \\
&+ k'' \\
=&\mathbf{u}^{\downarrow z\cup x'} K_1 \mathbf{u}^{\downarrow z\cup x} - \mu_1' K_1 \mathbf{u}^{\downarrow z\cup x} - \mathbf{u}^{\downarrow z\cup x'} K_1 \mu_1 \\
&+ \mathbf{u}^{\downarrow z\cup y'} K_2 \mathbf{u}^{\downarrow z\cup y} - \mu_2' K_2 \mathbf{u}^{\downarrow z\cup y} - \mathbf{u}^{\downarrow z\cup y'} K_2 \mu_2 \\
&+ k'' \\
=& \sum_{s,t\in\{x,z\}} (\mathbf{u}^{\downarrow s} - \mu_1{}^{\downarrow s})' K_1{}^{\downarrow s,t}(\mathbf{u}^{\downarrow t} - \mu_1{}^{\downarrow t}) \\
&+ \sum_{s,t\in\{y,z\}} (\mathbf{u}^{\downarrow s} - \mu_2{}^{\downarrow s})' K_2{}^{\downarrow s,t}(\mathbf{u}^{\downarrow t} - \mu_2{}^{\downarrow t}) \\
&+ k'''
\end{aligned}$$

where

$$k'' = \sum_{s,t\in\{x,y,z\}} \mu^{\downarrow s'} K^{\downarrow s,t} \mu^{\downarrow t}$$

and

$$k''' = \mu^{\downarrow x'} K^{\downarrow x,y} \mu^{\downarrow y} + \mu^{\downarrow y'} K^{\downarrow y,x} \mu^{\downarrow x} - \mu^{\downarrow z'} K^{\downarrow z} \mu^{\downarrow z}$$

do not depend on $\mathbf{u}$. Hence,

$$(\phi \otimes \psi)(\mathbf{u}) = k \cdot \phi(\mathbf{u}^{\downarrow z\cup x}) \cdot \psi(\mathbf{u}^{\downarrow z\cup y})$$

for some $k \in \mathbb{R}$ not depending on $\mathbf{u}$. Since

$$\begin{aligned} 1 = \int_{\mathbb{R}^x} (\phi \otimes \psi)(\mathbf{u}) &= \int_{\mathbb{R}^x} k \cdot \phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}) \\ = k \cdot \int_{\mathbb{R}^x} \phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}), \end{aligned}$$

indeed

$$k = \frac{1}{\int_{\mathbb{R}^x} \phi(\mathbf{u}^{\downarrow z \cup x}) \cdot \psi(\mathbf{u}^{\downarrow z \cup y}) d\mathbf{u}}.$$

□

Hence, the combination of Gaussian potentials represents the joint probability density of two *independent* random variables with Gaussian distribution on x and y. It will be shown below in Section 6.7 that this rule is a special case of the more general Dempster Rule of combination.

EXAMPLE **3.4**. The wholesale price W of a car is estimated by two independent experts (Pearl, 1988; Lehmann et al., 2005; Kohlas and Monney, 2008) by

- the difference μ_i $(i \in \{1, 2\})$ of the estimated asking price and the estimated mean profit and
- a standard deviation σ_i $(i \in \{1, 2\})$ which expresses the expert's confidence or reliability.

This yields two Gaussian potentials (μ_i, σ_i^{-2}) on the domain $\{W\}$ (for $i \in \{1, 2\}$). Let $\mu_1 = 7000$, $\sigma_1 = 300$, $\mu_2 = 9000$, and $\sigma_2 = 1000$. Combining the two experts' estimates then yields $\mu = (300^{-2} + 1000^{-2})^{-1}(300^{-2} \cdot 7000 + 1000^{-2} \cdot 9000) \approx 7165$ and standard deviation $\sigma = \sqrt{(300^{-2} + 1000^{-2})^{-1}} \approx 287$. So the combined estimate has a narrower range of uncertainty, but it is closer to the more confident (or more reliable) first expert's estimate. ⊘

Marginalisation of Gaussian Potentials

The marginal $\phi^{\downarrow s}$ of a Gaussian potential $\phi = (\mu, K)$ with domain $d(\phi) = s \cup t$, $s \cap t = \emptyset$, is supposed to represent the marginal probability density function

$$\phi^{\downarrow s}(\mathbf{s}) = \int_{\mathbf{t} \in \mathbb{R}^t} \phi_{\mu,K}(\mathbf{s}, \mathbf{t}) \mathrm{d}\mathbf{t} \tag{3.15}$$

for $\mathbf{s} \in \mathbb{R}^s$. According to Appendix B, define marginalisation $\downarrow : \mathcal{G} \times D \to \mathcal{G}, (\phi, s) \mapsto \phi^{\downarrow s}$, by

$$\phi^{\downarrow s} = (\mu^{\downarrow s}, \left((K^{-1})^{\downarrow s}\right)^{-1}) \tag{3.16}$$

This is well defined since, according to Corollaries 14.2.11 and 14.2.12 of (Harville, 1997; p.214), the inverse of a symmetric positive definite matrix exists and is symmetric positive definite and every principal submatrix of a symmetric positive definite matrix is symmetric and positive definite.

According to equation (3.15), in order to compute the marginal of a Gaussian potential, the whole concentration matrix K has to be inverted. An upper bound of the computational complexity of matrix inversion is given by $\mathcal{O}(n^3)$ flops (floating point operations)[1] in the schoolbook approach of Gaussian elimination. On the other hand, the theoretical lower bound time complexity of the inversion of a matrix of dimension n is of order $\mathcal{O}(n^2)$ flops since, informally, at least one operation is required for the n^2 entries. Notice that symmetry does not reduce this lower bound essentially since the upper (or lower) triangular submatrix has $\frac{n(n+1)}{2}$ entries. Hence, the complexity is of order $\mathcal{O}(n^\kappa)$ flops for some constant $\kappa \geq 2$ depending on the algorithm used.[2] The most widely implemented algorithms used for matrix inversion have complexity $\mathcal{O}(c{\cdot}n^3)$ ($c = \frac{2}{3}$ for the LU factorisation, $c = \frac{2}{3}$ the QR factorisation, $c = \frac{1}{3}$ for the Cholesky factorsation, $c = 4$ for the singular-value decomposition; see (Golub and Van Loan, 1989)).

However, in light of Lemma A.6,

$$\left((K^{-1})^{\downarrow s}\right)^{-1} = K^{\downarrow s} - K^{\downarrow s,t}(K^{\downarrow t})^{-1}K^{\downarrow t,s}. \tag{3.17}$$

This shows that, in order to compute the marginal of a Gaussian potential, only a submatrix of size $|t|$, the number of variables to be eliminated, needs to be inverted and not the whole matrix K of size $|d(\phi)| \geq |t|$. Let $n_1 = |t|$, $n_2 = |s|$ and $n = n_1 + n_2 = |d(\phi)|$. Therefore, a computation according to the formula on the left-hand side of (3.17) requires $\mathcal{O}(c \cdot (n^3 + {n_1}^3))$ flops. On the other hand, a computation according to the formula on the right-hand side of (3.17) requires $\mathcal{O}({n_1}^3 + {n_1}^2 n_2 + n_1 {n_2}^2 + {n_2}^2)$ flops: The inversion of $K^{\downarrow t}$ has complexity of order $\mathcal{O}({n_1}^3)$, the matrix products require $\mathcal{O}(n_2 {n_1}^2 + {n_2}^2 n_1)$ flops, and the final difference is of order $\mathcal{O}(\frac{n_2(n_2+1)}{2}) = \mathcal{O}({n_2}^2)$. However, the marginal can also be computed by successive variable eliminations (since it correspond to extracting the corresponding elements in the mean vector μ and the variance-covariance matrix K^{-1}). If a single variable $X_0 \in t$ is eliminated at a time, only multiplication, division and subtraction of real numbers are required:

$$\left((K^{-1})^{\downarrow d(\phi)-\{X_0\}}\right)^{-1}(Y,Z) = K(Y,Z) - K(Y,X_0)K(X_0,X_0)^{-1}K(X_0,Z) \tag{3.18}$$

for $Y, Z \in s_0 = d(\phi) - \{X_0\}$. Since $n_1 = 1$ in each step, the iteration needs

$$\sum_{j=n}^{n_2+1} 1 + j + j^2 + j^2 \approx 2(\frac{n(n+1)(2n+1)}{6} - \frac{n_2(n_2+1)(2n_2+1)}{6})$$

flops for n large enough. In an ad-hoc empirical test, the iterative method turns out to be faster.[3]

---

[1] See (Golub and Van Loan, 1989). The question of whether flops are a good measure of complexity and whether memory accesses should be taken into account as well will not be discussed here.

[2] The best known algorithm has $\kappa \approx 2.376$, see `http://en.wikipedia.org/wiki/Computational_complexity_of_mathematical_operations`, accessed 2009/02/27.

[3] In the marginalisation of a matrix with 500 and 1000 variables, the iterative method is about 5 times faster than using the LU factorisation for the inversion. For details of the iterative implementation, see Chapter 12.

3.3 Valuation Algebra of Gaussian Potentials

Theorem **3.5** (Valuation Algebra of Gaussian Potentials). *The algebraic structure $(\mathcal{G}, D, d, \otimes, \downarrow)$ of Gaussian potentials is a valuation algebra.* ⊘

Proof. It has to be verified that the operations satisfy the axioms (A1)-(A7) imposed on a valuation algebra.

(A1) Let

$$\phi_1 = (\mu_1, K_1), \phi_2 = (\mu_2, K_2), \phi_3 = (\mu_3, K_3) \in \mathcal{G},$$

be Gaussian potentials on domains

$$x = d(\phi_1), \quad y = d(\phi_2), \quad z = d(\phi_3),$$

and let

$$u = d(\phi_1) \cup d(\phi_2), \quad v = d(\phi_2) \cup d(\phi_3), \quad s = d(\phi_1) \cup d(\phi_2) \cup d(\phi_3).$$

Then, by the commutativity of vector and matrix addition,

$$K_1{}^{\uparrow u} + K_2{}^{\uparrow u} = K_2{}^{\uparrow u} + K_1{}^{\uparrow u} = K$$

and

$$\begin{aligned}\phi_1 \otimes \phi_2 &= \Big(K^{-1}(K_1{}^{\uparrow u} \cdot \mu_1{}^{\uparrow u} + K_2{}^{\uparrow u} \cdot \mu_2{}^{\uparrow u}), K\Big)\\ &= \Big(K^{-1}(K_2{}^{\uparrow u} \cdot \mu_2{}^{\uparrow u} + K_1{}^{\uparrow u} \cdot \mu_1{}^{\uparrow u}), K\Big)\\ &= \phi_2 \otimes \phi_1,\end{aligned}$$

hence combination of Gaussian potentials is indeed commutative. By the transitivity of vacuous extension

$$\begin{aligned}(K_1{}^{\uparrow u} + K_2{}^{\uparrow u})^{\uparrow s} + K_3{}^{\uparrow s} &= K_1{}^{\uparrow s} + K_2{}^{\uparrow s} + K_3{}^{\uparrow s}\\ &= K_1{}^{\uparrow s} + (K_2{}^{\uparrow v} + K_3{}^{\uparrow v})^{\uparrow s}\end{aligned}$$

and, using Lemma 3.2,

$$\begin{aligned}&\Big(K_1{}^{\uparrow u} \cdot \mu_1{}^{\uparrow u} + K_2{}^{\uparrow u} \cdot \mu_2{}^{\uparrow u}\Big)^{\uparrow s} + K_3{}^{\uparrow s} \cdot \mu_3{}^{\uparrow s}\\ =&K_1{}^{\uparrow s} \cdot \mu_1{}^{\uparrow s} + K_2{}^{\uparrow s} \cdot \mu_2{}^{\uparrow s} + K_3{}^{\uparrow s} \cdot \mu_3{}^{\uparrow s}\\ =&K_1{}^{\uparrow s} \cdot \mu_1{}^{\uparrow s} + \Big(K_2{}^{\uparrow v} \cdot \mu_2{}^{\uparrow v} + K_3{}^{\uparrow v} \cdot \mu_3{}^{\uparrow v}\Big)^{\uparrow s}.\end{aligned}$$

Therefore,

$$\begin{aligned}(\phi_1 \otimes \phi_2) \otimes \phi_3 &= \Big(K^{-1}\Big(K_1{}^{\uparrow s} \cdot \mu_1{}^{\uparrow s} + K_2{}^{\uparrow s} \cdot \mu_2{}^{\uparrow s} + K_3{}^{\uparrow s} \cdot \mu_3{}^{\uparrow s}\Big), K\Big)\\ &= \phi_1 \otimes (\phi_2 \otimes \phi_3)\end{aligned}$$

where

$$K = K_1{}^{\uparrow s} + K_2{}^{\uparrow s} + K_3{}^{\uparrow s},$$

hence combination of Gaussian potentials is also associative.

(A2) Let $\phi_1 = (\mu_1, K_1), \phi_2 = (\mu_2, K_2)$ with $\mu_1 \in \mathbb{R}^x$, $K_1 \in \mathbb{R}(x,x)$, $\mu_2 \in \mathbb{R}^y$, $K_2 \in \mathbb{R}(y,y)$. Let $\phi = (\mu, K) = \phi_1 \otimes \phi_2$. Then, $\mu \in \mathbb{R}^{x \cup y}$ and $K \in \mathbb{R}(x \cup y, x \cup y)$ in light of equations (3.8) and (3.7), hence $d(\phi_1 \otimes \phi_2) = d(\phi) = x \cup y = d(\phi_1) \cup d(\phi_2)$. This shows that Gaussian potentials verify the labelling axiom.

(A3) Let $\phi = (\mu, K)$ with $\mu \in \mathbb{R}^x$ and $K \in \mathbb{R}(x,x)$, i.e. $x = d(\phi)$. Then, for $s \subseteq x$,

$$\phi^{\downarrow s} = (\mu^{\downarrow s}, ((K^{-1})^{\downarrow s})^{-1})$$

where $\mu^{\downarrow s} \in \mathbb{R}^s$ and $((K^{-1})^{\downarrow s})^{-1} \in \mathbb{R}(s,s)$, hence $d(\phi^{\downarrow s}) = s$. This shows that Gaussian potentials verify the marginalisation axiom.

(A4) In order to prove the transitivity axiom, let $\phi = (\mu, K)$ be a Gaussian potential and let $t \subseteq s \subseteq d(\phi)$. Then, using the transitivity of projection of vectors and matrices,

$$\phi^{\downarrow t} = (\mu^{\downarrow t}, ((K^{-1})^{\downarrow t})^{-1}) = ((\mu^{\downarrow s})^{\downarrow t}, ((((K^{-1\downarrow s})^{-1})^{-1\downarrow t})^{-1}) = (\phi^{\downarrow s})^{\downarrow t},$$

which shows the transitivity of marginalisation.

(A5) Let $\phi = (\mu_1, K_1), \psi = (\mu_2, K_2) \in \mathcal{G}$ with $x = d(\phi)$, $y = d(\psi)$, $u = x \cup y$ and $z \in D$ such that

$$x \subseteq z \subseteq x \cup y.$$

Let $(\mu, K) = \phi \otimes \psi$. Partition K and μ according to z and $(x \cup y) - z$,

$$K = \begin{pmatrix} {K_1}^{\uparrow z} + {K_2}^{\Rightarrow z} & {K_2}^{\Rightarrow z, y-z} \\ {K_2}^{\Rightarrow y-z, z} & {K_2}^{\downarrow y-z} \end{pmatrix}$$

and

$$\mu = K^{-1}({K_1}^{\uparrow u}{\mu_1}^{\uparrow u} + {K_2}^{\uparrow u}{\mu_2}^{\uparrow u}).$$

On the one hand, let

$$(\mu^{\downarrow z}, \tilde{K}) = (\phi \otimes \psi)^{\downarrow z},$$

where, in light of Lemma 3.2,

$$\mu^{\downarrow z} = (K^{-1})^{\downarrow z, u}({K_1}^{\uparrow u}{\mu_1}^{\uparrow u} + {K_2}^{\uparrow u}{\mu_2}^{\uparrow u}),$$

and, according to equation (3.17),

$$\tilde{K} = {K_1}^{\uparrow z} + {K_2}^{\Rightarrow z} - {K_2}^{\Rightarrow z, y-z}({K_2}^{\downarrow y-z})^{-1}{K_2}^{\Rightarrow y-z, z}.$$

On the other hand, according to equation (3.17),

$$\psi^{\downarrow y \cap z} = ({\mu_2}^{\downarrow y \cap z}, \tilde{K}_2)$$

where

$$\tilde{K}_2 = \qquad K_2{}^{\downarrow y\cap z} - K_2{}^{\downarrow y\cap z, y-z}\left(K_2{}^{\downarrow y-z}\right)^{-1} K_2{}^{\downarrow y-z, y\cap z}.$$

Then,

$$\phi \otimes \psi^{\downarrow y\cap z} = ((K_1{}^{\uparrow z} + \tilde{K}_2{}^{\uparrow z})^{-1}(K_1{}^{\uparrow z}\mu_1{}^{\uparrow z} + \tilde{K}_2^{\uparrow z}\mu_2{}^{\Rightarrow z}), K_1{}^{\uparrow z} + \tilde{K}_2{}^{\uparrow z}).$$

In order to prove $(\phi \otimes \psi)^{\downarrow z} = \phi \otimes \psi^{\downarrow y\cap z}$, it has therefore to be verified that

$$\tilde{K} = K_1{}^{\uparrow z} + \tilde{K}_2{}^{\uparrow z} \tag{3.19}$$

and that

$$\mu^{\downarrow z} = (K_1{}^{\uparrow z} + \tilde{K}_2{}^{\uparrow z})^{-1}(K_1{}^{\uparrow z}\mu_1{}^{\uparrow z} + \tilde{K}_2^{\uparrow z}\mu_2{}^{\Rightarrow z}). \tag{3.20}$$

Observe that

$$\tilde{K}_2^{\uparrow z} = K_2{}^{\Rightarrow z} - K_2{}^{\Rightarrow z, y-z}\left(K_2{}^{\downarrow y-z}\right)^{-1} K_2{}^{\Rightarrow y-z, z} \tag{3.21}$$

in light of Lemma 3.2. Hence, equation (3.19) holds indeed. Second, recalling that $\tilde{K} = ((K^{-1})^{\downarrow z})^{-1}$ and posing $c_{11} = \tilde{K}^{-1}$ in Lemma A.6,

$$(K^{-1})^{\downarrow z,u} = \begin{pmatrix} \tilde{K}^{-1} & -\tilde{K}^{-1} K_2{}^{\Rightarrow z, y-z}\left(K_2{}^{\downarrow y-z}\right)^{-1} \end{pmatrix},$$

Since $x \subseteq z \subseteq u$,

$$(K^{-1})^{\downarrow z,u} K_1{}^{\uparrow u}\mu_1{}^{\uparrow u} = \tilde{K}^{-1} K_1{}^{\uparrow z}\mu_1{}^{\uparrow z}.$$

Using (3.21)

$$\begin{aligned}
&(K^{-1})^{\downarrow z,u} K_2{}^{\uparrow u} \\
&= \begin{pmatrix} \tilde{K}^{-1} & -\tilde{K}^{-1} K_2{}^{\Rightarrow z, y-z}\left(K_2{}^{\downarrow y-z}\right)^{-1} \end{pmatrix} \begin{pmatrix} K_2{}^{\Rightarrow z} & K_2{}^{\Rightarrow z, y-z} \\ K_2{}^{\Rightarrow y-z, z} & K_2{}^{\Rightarrow y-z} \end{pmatrix} \\
&= \begin{pmatrix} \tilde{K}^{-1}\tilde{K}_2{}^{\uparrow z} & 0_{z, y-z} \end{pmatrix}
\end{aligned}$$

and thus

$$(K^{-1})^{\downarrow z,u} K_2{}^{\uparrow u}\mu^{\uparrow u} = \tilde{K}^{-1}\tilde{K}_2{}^{\uparrow z}\mu_2{}^{\Rightarrow z}.$$

Hence, equation (3.20) holds indeed, too.

(A6) The domain axiom follows by the definition of marginalisation.

(A7) Finally, $e = (0_\emptyset, 0_{\emptyset,\emptyset})$ is an identity element. □

In light of the transitivity of marginalisation of Gaussian potentials, marginals can be calculated step-wise, using (3.17), or even variable-wise, using (3.18).

3.4 Vacuous Extension of Gaussian Potentials

It has been shown in Lemma 2.22 how a valuation without neutral elements can be extended to a stable valuation algebra. The elements in the extension have the form (ϕ, x) for ϕ and $d(\phi) \subseteq x \in D$. A geometric interpretation will now be given for the extension of Gaussian potentials. A Gaussian potential can be viewed as a random variable whose outcomes are sets of configurations of its domain, i.e. a Gaussian potential (ϕ, x) with domain $x = d(\phi)$ induces the surjective mapping $\Gamma_x(\mathbf{x}) = \{\mathbf{x}\}$, called *focal* mapping. Formally, $\Gamma_x : \mathbb{R}^x \to 2^{\mathbb{R}^x}$ is a random variable whose outcomes are the singleton sets $\{\mathbf{x}\}$, which form a partition of $\mathbb{R}^x$. In the spirit of this approach, a neutral element $e'_x = (\diamond, x)$, expressing complete ignorance on the domain x, is represented by the mapping $\Gamma_\emptyset{}^{\uparrow x}(\omega) = \mathbb{R}^x$. This corresponds to a distribution over the trivial partition consisting of $\mathbb{R}^x$ only.

Gaussian potentials in the extension and neutral elements are two extreme cases of the following general situation: An element (ϕ, y) with $x = d(\phi) \subseteq y$ induces a focal mapping

$$\Gamma_x{}^{\uparrow y}(\mathbf{x}) = \{\mathbf{x}\} \times \mathbb{R}^{y-x}. \tag{3.22}$$

The image of such a mapping $\Gamma_x{}^{\uparrow y}$ is always a partition of $\mathbb{R}^x$, which consists of parallel linear manifolds which stand orthogonal on the linear hyperplane spanned by the singleton domain x. Hence, the extension of Gaussian potentials leads to distributions over parallel linear manifolds, including distributions over singeltons as a special case.

EXAMPLE **3.6**. An example of the vacuous extension is shown in Figure 3.1: Here, the points of the x-axis are mapped to planes orthogonal to the x-axis in Figure 3.1(a) and points of the xy-plane to straight lines perpendicular to that plane in Figure 3.1(b). ⊘

The marginal

$$(\phi, y)^{\downarrow z} = (\phi^{\downarrow x \cap z}, z)$$

induces the focal mapping $\Gamma_{x \cap z}{}^{\uparrow z}$. Marginalisation in the extension corresponds to the *projection* of the configurations in the image of $\Gamma_x{}^{\uparrow y}$ since

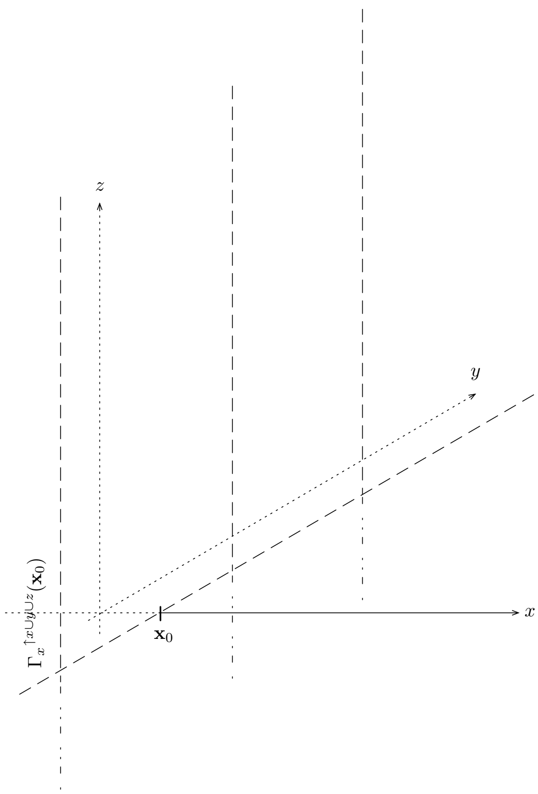
$$\begin{aligned}
\Gamma_x{}^{\uparrow y}(\mathbf{x})^{\downarrow z} &= \{\mathbf{z} \in \mathbb{R}^z : \exists \mathbf{y} \in \Gamma_x{}^{\uparrow y}(\mathbf{x}), \, \mathbf{z} = \mathbf{y}^{\downarrow z}\} \\
&= \{\mathbf{z} \in \mathbb{R}^z : \exists \mathbf{y} \in \mathbb{R}^y, \, \mathbf{z} = \mathbf{y}^{\downarrow z}, \, \mathbf{y}^{\downarrow x} = \mathbf{x}\} \\
&= \mathbf{x}^{\downarrow x \cap z} \times \mathbb{R}^{z-x} \\
&= \Gamma_{x \cap z}{}^{\uparrow z}(\mathbf{x}^{\downarrow x \cap z})
\end{aligned}$$

for $\mathbf{x} \in \mathbb{R}^x$. Furthermore, the combination

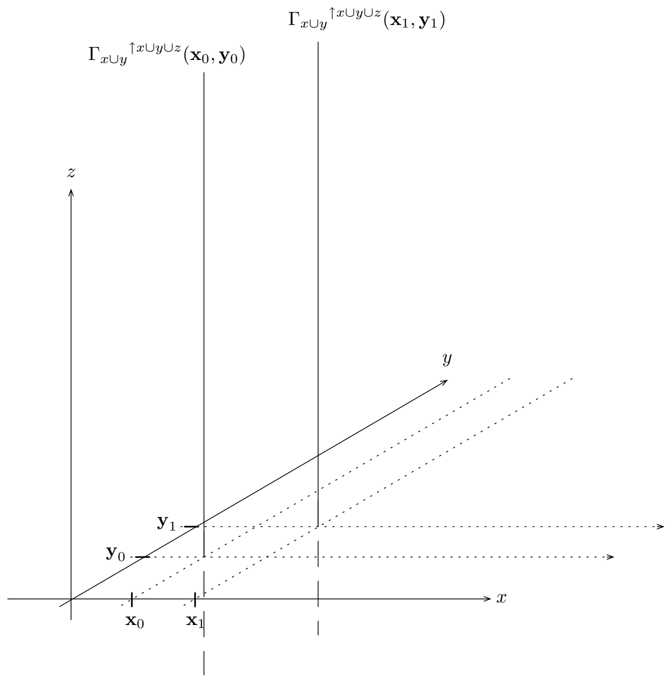
$$(\phi_1, y_1) \otimes' (\phi_2, y_2) = (\phi_1 \otimes \phi_2, y_1 \cup y_1)$$

induces to the *intersection of the sets* since

$$\begin{aligned}
\Gamma_{x_1 \cup x_2}{}^{\uparrow y_1 \cup y_2}(\mathbf{u}) &= \{\mathbf{u}\} \times \mathbb{R}^{y_1 \cup y_2 - (x_1 \cup x_2)} \\
&= \left(\{\mathbf{u}^{\downarrow x_1}\} \times \mathbb{R}^{y_1 \cup y_2 - (x_1 \cup x_2)}\right) \cap \left(\{\mathbf{u}^{\downarrow x_2}\} \times \mathbb{R}^{y_1 \cup y_2 - (x_1 \cup x_2)}\right) \\
&= \Gamma_{x_1}{}^{\uparrow y_1 \cup y_2}(\mathbf{u}^{\downarrow x_1}) \cap \Gamma_{x_2}{}^{\uparrow y_1 \cup y_2}(\mathbf{u}^{\downarrow x_2})
\end{aligned}$$



(a) Vacuous extension from one dimension to three dimensions



(b) Vacuous extension from two dimensions to three

Figure 3.1: Geometric interpretation of the vacuous extension of Gaussian potentials

for $x_1 = d(\phi_1)$ and $x_2 = d(\phi_2)$ and $\mathbf{u} \in \mathbb{R}^{x_1 \cup x_2}$. This product-intersection gives a geometric interpretation to marginalisation and combination as well.

These mappings will be generalised in the Chapter 6. Furthermore, they will be derived within a precise theory of (statistical) inference will be given to explain how such focal mappings arise.

3.5 Moment Matrices and Sweeping

In this section, *moment matrices* (Dempster, 1990a) or *extended matrices* (Liu, 1999) are introduced as an equivalent representation of Gaussian potentials. Combination is expressed in terms of matrix operations called *sweepings* on the common variables. From the sweeping-based combination of moment matrices, a more efficient way of computing the combined mean of Gaussian potentials will be deduced. Extended matrices are used for the probabilistic variables in Gaussian belief functions (Dempster, 1990a; Liu, 1996a; 1999); see also Section 10.7.

Instead of representing a Gaussian density by a Gaussian potential (which is the pair of its mean vector and its concentration matrix), such a distribution may be represented by an extended matrix, which is the pair of its mean vector and its variance-covariance matrix: A corresponding extended matrix can be associated to every Gaussian potential (μ, K),

$$\sigma(\mu, K) = (\mu, K^{-1}). \tag{3.23}$$

Marginalisation of Gaussian potentials can then be carried over to such extended matrices as $\downarrow_M$ by

$$(\mu, K^{-1})^{\downarrow_M s} = \left(\mu^{\downarrow s}, (K^{-1})^{\downarrow s}\right). \tag{3.24}$$

The marginalisation operations of Gaussian potentials and of moment matrices are compatible since

$$(\mu, K^{-1})^{\downarrow_M s} = \left(\mu^{\downarrow s}, (K^{-1})^{\downarrow s}\right) = \sigma\left(\mu^{\downarrow s}, \left((K^{-1})^{\downarrow s}\right)^{-1}\right) = \sigma\left((\mu, K)^{\downarrow s}\right).$$

Marginalisation in terms of extended matrices is an easy operation since only the relevant elements of the mean vector and the concentration matrix have to be extracted. Furthermore, combination $\otimes$ can be carried over as $\otimes_M$ by

$$\begin{aligned}(\mu_1, {K_1}^{-1}) \otimes_M (\mu_2, {K_2}^{-1}) &= (K^{-1}((K_1\mu_1)^{\uparrow x \cup y} + (K_2\mu_2)^{\uparrow x \cup y}), K^{-1}) \\ &= \sigma((\mu_1, K_1) \otimes (\mu_2, K_2))\end{aligned} \tag{3.25}$$

where

$$K = {K_1}^{\uparrow x \cup y} + {K_2}^{\uparrow x \cup y}, \tag{3.26}$$

$\phi_1 = (\mu_1, K_1), \phi_2 = (\mu_2, K_2) \in \mathcal{G}$, $x = d(\phi_1)$, $y = d(\phi_2)$. According to formula (3.25), matrices of the dimension of the number of all variables $x \cup y$ have to be inverted in order to compute the combination of extended matrices, which is expensive. However, it turns out that there is a less expensive way of computing the combination of extended matrices: It suffices to invert matrices whose dimension is at most the number of common variables.

In light of equations (B.10) and (B.9), define the Gaussian potential obtained by conditioning on a subset of variables.

DEFINITION **3.7**. *Let $\phi = (\mu, K)$ be a Gaussian potential with domain $d(\phi) = x \cup z$ such that $x \cap z = \emptyset$. Then, $\phi_{x|\mathbf{z}} \in \mathcal{G}$ with $d(\phi_{x|\mathbf{z}}) = x$, given by*

$$\phi_{x|\mathbf{z}} = (\mu^{\downarrow x} - K^{\downarrow x^{-1}} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z}), K^{\downarrow x}), \tag{3.27}$$

is called the conditioned Gaussian potential of ϕ on $\mathbf{z} \in \mathbb{R}^z$. ⊘

Let $\phi = (\mu, K) \in \mathcal{G}$, $(\mu, \Sigma) = \sigma(\mu, K)$, and $x \cup z = d(\phi)$, $x \cap z = \emptyset$. In light of Lemma B.1, the extended matrix corresponding to a conditioned Gaussian potential is given by

$$\sigma(\phi_{x|\mathbf{z}}) = (\mu_{x|\mathbf{z}}, \Sigma_{x|\mathbf{z}}) \tag{3.28}$$

where

$$\mu_{x|\mathbf{z}} = \mu^{\downarrow x} + \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} (\mathbf{z} - \mu^{\downarrow z}) \tag{3.29}$$

and

$$\Sigma_{x|\mathbf{z}} = \Sigma^{\downarrow x} - \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} \Sigma^{\downarrow z,x}. \tag{3.30}$$

From this extended matrix, the original extended matrix (μ, Σ) can be reconstructed if $\mathbf{z}$, $\mu^{\downarrow z}$, $\Sigma^{\downarrow z}$ and $\Sigma^{\downarrow x,z}$ are retained since then

$$\begin{pmatrix} \mu_{x|\mathbf{z}} - \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} (\mathbf{z} - \mu^{\downarrow z}) \\ \mu^{\downarrow z} \end{pmatrix} = \begin{pmatrix} \mu^{\downarrow x} \\ \mu^{\downarrow z} \end{pmatrix} = \mu$$

and

$$\begin{pmatrix} \Sigma_{x|\mathbf{z}} + \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} (\Sigma^{\downarrow x,z})' & \Sigma^{\downarrow x,z} \\ (\Sigma^{\downarrow x,z})' & \Sigma^{\downarrow z} \end{pmatrix} = \begin{pmatrix} \Sigma^{\downarrow x} & \Sigma^{\downarrow x,z} \\ \Sigma^{\downarrow z,x} & \Sigma^{\downarrow z} \end{pmatrix} = \Sigma.$$

In order to represent the distribution conditioned on $\mathbf{z} \in \mathbb{R}^z$, retaining the information necessary to reverse the conditioning, (Liu, 1999) defines the forward sweep of the extended matrix $M = (\mu, \Sigma)$ on $\mathbf{z}$ to be the pair $\triangleright(M, \mathbf{z}) = (\tilde{\mu}, \tilde{M})$,

$$\tilde{\mu} = \begin{pmatrix} \mu^{\downarrow x} + \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} (\mathbf{z} - \mu^{\downarrow z}) \\ \Sigma^{\downarrow z^{-1}} \mu^{\downarrow z} \end{pmatrix}, \tag{3.31}$$

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma^{\downarrow x} - \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} \Sigma^{\downarrow z,x} & \Sigma^{\downarrow x,z} \Sigma^{\downarrow z^{-1}} \\ \Sigma^{\downarrow z^{-1}} \Sigma^{\downarrow z,x} & -\Sigma^{\downarrow z^{-1}} \end{pmatrix} \tag{3.32}$$

and the reverse sweep of such a pair $\tilde{M} = (\tilde{\mu}, \tilde{\Sigma})$ from $\tilde{\mathbf{z}} \in \mathbb{R}^z$ by $\triangleleft(\tilde{\phi}, \tilde{\mathbf{z}}) = (\mu, \Sigma)$,

$$\mu = \begin{pmatrix} \tilde{\mu}^{\downarrow x} - \tilde{\Sigma}^{\downarrow x,z} (\tilde{\mathbf{z}} + \tilde{\Sigma}^{\downarrow z^{-1}} \tilde{\mu}^{\downarrow z}) \\ -\tilde{\Sigma}^{\downarrow z^{-1}} \tilde{\mu}^{\downarrow z} \end{pmatrix}, \tag{3.33}$$

$$\Sigma = \begin{pmatrix} \tilde{\Sigma}^{\downarrow x} - \tilde{\Sigma}^{\downarrow x,z} \tilde{\Sigma}^{\downarrow z^{-1}} \tilde{\Sigma}^{\downarrow z,x} & -\tilde{\Sigma}^{\downarrow x,z} \tilde{\Sigma}^{\downarrow z^{-1}} \\ -\tilde{\Sigma}^{\downarrow z^{-1}} \tilde{\Sigma}^{\downarrow z,x} & -\tilde{\Sigma}^{\downarrow z^{-1}} \end{pmatrix}. \tag{3.34}$$

Then, it can be verified that

$$\triangleleft(\triangleright(M, \mathbf{z}), \mathbf{z}) = M. \tag{3.35}$$

Forward and reverse sweepings can be used to express combination in terms of the associated extended matrices as shown by the following lemma. (Liu, 1996a; 1999), derived the same combination rule for moment matrices from *Dempster's Rule of Combination*. It will be shown below in Section 6.7 in a more general setting that the combination of Gaussian potentials complies with Dempster's Rule. Here, a direct a proof is given instead.

Lemma **3.8**. *Let $\phi, \psi \in \mathcal{G}$ with $z = d(\phi) \cap d(\psi)$, $x = d(\phi) - d(\psi)$, $y = d(\psi) - d(\phi)$. Then*

$$\sigma(\phi \otimes \psi) = \triangleleft([\triangleright(\sigma(\phi), \mathbf{z}) \oplus \triangleright(\sigma(\psi), \mathbf{z})], \mathbf{z}) \tag{3.36}$$

for any $\mathbf{z} \in \mathbb{R}^z$ where $\oplus$ stands for addition of the two vectors and matrices vacuously extended to $x \cup z \cup y$. ⊘

Proof. Let $(\mu_1, K_1) = \phi$, $(\mu_2, K_2) = \psi$, $\Sigma_1 = {K_1}^{-1}$, and $\Sigma_2 = {K_2}^{-1}$. Then, $\sigma(\phi) = (\mu_1, \Sigma_1)$ and $\sigma(\psi) = (\mu_2, \Sigma_2)$.
On the one hand,

$$\triangleright(\sigma(\phi), \mathbf{z}) \oplus \triangleright(\sigma(\psi), \mathbf{z})$$

is given by the vector

$$\begin{pmatrix} {\mu_1}^{\downarrow x} + {\Sigma_1}^{\downarrow x,z} {{\Sigma_1}^{\downarrow z}}^{-1} (\mathbf{z} - {\mu_1}^{\downarrow z}) \\ {\mu_2}^{\downarrow y} + {\Sigma_2}^{\downarrow y,z} {{\Sigma_2}^{\downarrow z}}^{-1} (\mathbf{z} - {\mu_2}^{\downarrow z}) \\ ({\Sigma_1}^{\downarrow z})^{-1} {\mu_1}^{\downarrow z} + ({\Sigma_2}^{\downarrow z})^{-1} {\mu_2}^{\downarrow z} \end{pmatrix}$$

and the matrix

$$\begin{pmatrix} {\Sigma_1}^{\downarrow x} - {\Sigma_1}^{\downarrow x,z} {{\Sigma_1}^{\downarrow z}}^{-1} {\Sigma_1}^{\downarrow z,x} & 0_{x,y} & {\Sigma_1}^{\downarrow x,z} ({\Sigma_1}^{\downarrow z})^{-1} \\ 0_{y,x} & {\Sigma_2}^{\downarrow y} - {\Sigma_2}^{\downarrow y,z} {{\Sigma_2}^{\downarrow z}}^{-1} {\Sigma_2}^{\downarrow z,y} & {\Sigma_2}^{\downarrow y,z} ({\Sigma_2}^{\downarrow z})^{-1} \\ ({\Sigma_1}^{\downarrow z})^{-1} {\Sigma_1}^{\downarrow z,x} & ({\Sigma_2}^{\downarrow z})^{-1} {\Sigma_2}^{\downarrow z,y} & -(({\Sigma_1}^{\downarrow z})^{-1} + ({\Sigma_2}^{\downarrow z})^{-1}) \end{pmatrix}.$$

Let

$$(\mu, \Sigma) = \triangleleft([\triangleright(\sigma(\phi), \mathbf{z}) \oplus \triangleright(\sigma(\psi), \mathbf{z})], \mathbf{z})$$

where

$$\mu = \begin{pmatrix} {\mu_1}^{\downarrow x} - {\Sigma_1}^{\downarrow x,z} {{\Sigma_1}^{\downarrow z}}^{-1} ({\mu_1}^{\downarrow z} - \mu_z) \\ {\mu_2}^{\downarrow y} - {\Sigma_2}^{\downarrow y,z} {{\Sigma_2}^{\downarrow z}}^{-1} ({\mu_2}^{\downarrow z} - \mu_z) \\ \mu_z \end{pmatrix} \tag{3.37}$$

for

$$\mu_z = (({\Sigma_1}^{\downarrow z})^{-1} + ({\Sigma_2}^{\downarrow z})^{-1})^{-1} (({\Sigma_1}^{\downarrow z})^{-1} {\mu_1}^{\downarrow z} + ({\Sigma_2}^{\downarrow z})^{-1} {\mu_2}^{\downarrow z})$$

and

$$\Sigma = \begin{pmatrix} \Sigma_x & \Sigma_{x,y} & {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z \\ \Sigma_{y,x} & \Sigma_y & {\Sigma_2}^{\downarrow y,z}({\Sigma_2}^{\downarrow z})^{-1}\Sigma_z \\ \Sigma_z({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x} & \Sigma_z({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y} & \Sigma_z \end{pmatrix} \tag{3.38}$$

for

$$\begin{aligned} \Sigma_x &= {\Sigma_1}^{\downarrow x} - {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x} + {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}, \\ \Sigma_{x,y} &= {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}, \\ \Sigma_{y,x} &= {\Sigma_2}^{\downarrow y,z}({\Sigma_2}^{\downarrow z})^{-1}\Sigma_z({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}, \\ \Sigma_y &= {\Sigma_2}^{\downarrow y} - {\Sigma_2}^{\downarrow y,z}({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y} + {\Sigma_2}^{\downarrow y,z}({\Sigma_2}^{\downarrow z})^{-1}\Sigma_z({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}, \end{aligned}$$

and

$$\Sigma_z = \left(({\Sigma_1}^{\downarrow z})^{-1} + ({\Sigma_2}^{\downarrow z})^{-1}\right)^{-1}.$$

On the other hand, let

$$(\tilde{\mu}, K) = \phi \otimes \psi$$

where, in light of equation (3.10),

$$\begin{aligned} \tilde{\mu} &= K^{-1}\begin{pmatrix} {K_1}^{\downarrow x}{\mu_1}^{\downarrow x} + {K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} \\ {K_2}^{\downarrow y}{\mu_2}^{\downarrow y} + {K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z} \\ {K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + {K_2}^{\downarrow z}{\mu_2}^{\downarrow z} + {K_1}^{\downarrow z,x}{\mu_1}^{\downarrow x} + {K_2}^{\downarrow z,y}{\mu_2}^{\downarrow y} \end{pmatrix} \\ &= K^{-1}\left[\begin{pmatrix} {K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} \\ {K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z} \\ {K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + {K_2}^{\downarrow z}{\mu_2}^{\downarrow z} \end{pmatrix} + \begin{pmatrix} {K_1}^{\downarrow x}{\mu_1}^{\downarrow x} \\ 0_y \\ {K_1}^{\downarrow z,x}{\mu_1}^{\downarrow x} \end{pmatrix} + \begin{pmatrix} 0_x \\ {K_2}^{\downarrow y}{\mu_2}^{\downarrow y} \\ {K_2}^{\downarrow z,y}{\mu_2}^{\downarrow y} \end{pmatrix}\right] \\ &= K^{-1}\left[\begin{pmatrix} {K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} \\ {K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z} \\ {K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + {K_2}^{\downarrow z}{\mu_2}^{\downarrow z} \end{pmatrix} + K^{\downarrow z\cup x\cup y,x}{\mu_1}^{\downarrow x} + K^{\downarrow z\cup x\cup y,y}{\mu_2}^{\downarrow y}\right], \end{aligned}$$

and, applying Lemma A.6 to both Σ_1 and Σ_2,

$$\begin{aligned} K &= {K_1}^{\uparrow x\cup y\cup z} + {K_2}^{\uparrow x\cup y\cup z} = \left({\Sigma_1}^{-1}\right)^{\uparrow x\cup y\cup z} + \left({\Sigma_2}^{-1}\right)^{\uparrow x\cup y\cup z} \\ &= \begin{pmatrix} s_{12} & 0_{x,y} & -({\Sigma_1}^{\downarrow x})^{-1}{\Sigma_1}^{\downarrow x,z}s_{11} \\ 0_{y,x} & s_{22} & -({\Sigma_2}^{\downarrow y})^{-1}{\Sigma_2}^{\downarrow y,z}s_{21} \\ -({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}s_{12} & -({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}s_{22} & s_{11} + s_{21} \end{pmatrix} \end{aligned} \tag{3.39}$$

for

$$\begin{aligned} s_{12} &= \left({\Sigma_1}^{\downarrow x} - {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}\right)^{-1}, \\ s_{11} &= \left({\Sigma_1}^{\downarrow z} - {\Sigma_1}^{\downarrow z,x}({\Sigma_1}^{\downarrow x})^{-1}{\Sigma_1}^{\downarrow x,z}\right)^{-1}, \\ s_{22} &= \left({\Sigma_2}^{\downarrow y} - {\Sigma_2}^{\downarrow y,z}({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}\right)^{-1}, \end{aligned}$$

and

$$s_{21} = \left(\Sigma_2{}^{\downarrow z} - \Sigma_2{}^{\downarrow z,y}(\Sigma_2{}^{\downarrow y})^{-1}\Sigma_2{}^{\downarrow y,z}\right)^{-1}.$$

With these definitions, the claim can be written as

$$\sigma(\mu, K) = (\mu, \Sigma),$$

so it has to be proved that $\Sigma K = I_{x\cup y\cup z}$ and $\mu = \tilde{\mu}$. Firstly, using equations (3.39) and (3.38) and the definitions of their submatrices,

$$\begin{aligned}(\Sigma K)^{\downarrow z} =& \Sigma_z(s_{11} + s_{21}) + \Sigma_z(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_1{}^{\downarrow z,x}(-(\Sigma_1{}^{\downarrow x})^{-1}\Sigma_1{}^{\downarrow x,z}s_{11})\\ &+ \Sigma_z(\Sigma_2{}^{\downarrow z})^{-1}\Sigma_2{}^{\downarrow z,y}(-(\Sigma_2{}^{\downarrow y})^{-1}\Sigma_2{}^{\downarrow y,z}s_{21})\\ =& \Sigma_z(s_{11} + s_{21}) + \Sigma_z(\Sigma_1{}^{\downarrow z})^{-1}(s_{11}{}^{-1} - \Sigma_1{}^{\downarrow z})s_{11}\\ &+ \Sigma_z(\Sigma_2{}^{\downarrow z})^{-1}(s_{21}{}^{-1} - \Sigma_2{}^{\downarrow z})s_{21}\\ =& \Sigma_z(s_{11} + s_{21}) + \Sigma_z((\Sigma_1{}^{\downarrow z})^{-1} - s_{11}) + \Sigma_z((\Sigma_2{}^{\downarrow z})^{-1} - s_{21})\\ =& \Sigma_z((\Sigma_1{}^{\downarrow z})^{-1} + (\Sigma_2{}^{\downarrow z})^{-1}) = I_z,\end{aligned}$$

$$\begin{aligned}(\Sigma K)^{\downarrow x} =& \Sigma_1{}^{\downarrow x,z}(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_z(-(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_1{}^{\downarrow z,x}s_{12})\\ &+ (s_{12}{}^{-1} + \Sigma_1{}^{\downarrow x,z}(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_z(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_1{}^{\downarrow z,x})s_{12}\\ =& s_{12}{}^{-1}s_{12} = I_x,\end{aligned}$$

$$(\Sigma K)^{\downarrow z,x} = \Sigma_z(-\Sigma_1{}^{\downarrow z^{-1}}\Sigma_1{}^{\downarrow z,x}s_{12}) + \Sigma_z\Sigma_1{}^{\downarrow z^{-1}}\Sigma_1{}^{\downarrow z,x}s_{12} \;=\; 0_{z,x},$$

$$\begin{aligned}(\Sigma K)^{\downarrow x,y} =& \Sigma_1{}^{\downarrow x,z}(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_z(-(\Sigma_2{}^{\downarrow z})^{-1}\Sigma_2{}^{\downarrow z,y})s_{22}\\ &+ \Sigma_1{}^{\downarrow x,z}(\Sigma_1{}^{\downarrow z})^{-1}\Sigma_z(\Sigma_2{}^{\downarrow z})^{-1}\Sigma_2{}^{\downarrow z,y}s_{22} = \; 0_{x,y}.\end{aligned}$$

By similar arguments, it can proved that $(\Sigma K)^{\downarrow y} = I_y$ and $(\Sigma K)^{\downarrow z,y} = 0_{z,y}$, and it then follows by the symmetry of ΣK that $\Sigma K = I_{x\cup z}$.
Secondly, using $K^{-1} = \Sigma$ and applying equation (B.13) [taking $K_{11} = \Sigma_1{}^{\downarrow z}$ and $\Sigma_{22} = K_1{}^{\downarrow x}$],

$$\begin{aligned}\tilde{\mu}^{\downarrow z} =& \Sigma_z(K_1{}^{\downarrow z}\mu_1{}^{\downarrow z} + K_2{}^{\downarrow z}\mu_2{}^{\downarrow z})\\ &+ \Sigma_z(\Sigma_1)^{\downarrow z^{-1}}\Sigma_1{}^{\downarrow z,x}K_1{}^{\downarrow x,z}\mu_1{}^{\downarrow z} + \Sigma_z(\Sigma_2)^{\downarrow z^{-1}}\Sigma_2{}^{\downarrow z,y}K_2{}^{\downarrow y,z}\mu_2{}^{\downarrow z}\\ =& \Sigma_z((K_1{}^{\downarrow z} - K_1{}^{\downarrow z,x}(K_1{}^{\downarrow x})^{-1}K_1{}^{\downarrow x,z})\mu_1{}^{\downarrow z} + (K_2{}^{\downarrow z} - K_2{}^{\downarrow z,y}(K_2{}^{\downarrow y})^{-1}K_2{}^{\downarrow y,z})\mu_2{}^{\downarrow z})\\ =& ((\Sigma_1{}^{\downarrow z})^{-1} + (\Sigma_2{}^{\downarrow z})^{-1})^{-1}((\Sigma_1{}^{\downarrow z})^{-1}\mu_1{}^{\downarrow z} + (\Sigma_2{}^{\downarrow z})^{-1}\mu_2{}^{\downarrow z})\\ =& \mu^{\downarrow z},\end{aligned}$$

and

$$\begin{aligned}
\tilde{\mu}^{\downarrow x} =& {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + {K_2}^{\downarrow z}{\mu_2}^{\downarrow z}) \\
&+ (({K_1}^{\downarrow x})^{-1} + {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}){K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} \\
&+ {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}{K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z} + I_x{\mu_1}^{\downarrow x} \\
=& {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({K_1}^{\downarrow z}{\mu_1}^{\downarrow z} + ({\Sigma_1}^{\downarrow z})^{-1}{\Sigma_1}^{\downarrow z,x}{K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z}) \\
&+ {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({K_2}^{\downarrow z}{\mu_2}^{\downarrow z} + ({\Sigma_2}^{\downarrow z})^{-1}{\Sigma_2}^{\downarrow z,y}{K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z}) \\
&+ ({K_1}^{\downarrow x})^{-1}{K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z} + {\mu_1}^{\downarrow x} \\
=& {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({K_1}^{\downarrow z}{\mu_1}^{\downarrow z} - {K_1}^{\downarrow z,x}({K_1}^{\downarrow x})^{-1}{K_1}^{\downarrow x,z}{\mu_1}^{\downarrow z}) \\
&+ {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z({K_2}^{\downarrow z}{\mu_2}^{\downarrow z} - {K_2}^{\downarrow z,y}({K_2}^{\downarrow y})^{-1}{K_2}^{\downarrow y,z}{\mu_2}^{\downarrow z}) \\
&- {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}{\mu_1}^{\downarrow z} + {\mu_1}^{\downarrow x} \\
=& {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}\Sigma_z(({\Sigma_1}^{\downarrow z})^{-1}{\mu_1}^{\downarrow z} + ({\Sigma_2}^{\downarrow z})^{-1}{\mu_2}^{\downarrow z}) \\
&- {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}{\mu_1}^{\downarrow z} + {\mu_1}^{\downarrow x} \\
=& {\mu_1}^{\downarrow x} + {\Sigma_1}^{\downarrow x,z}({\Sigma_1}^{\downarrow z})^{-1}(\mu_z - {\mu_1}^{\downarrow z}) \\
=& \mu^{\downarrow x}.
\end{aligned}$$

In the same way it can be proved that $\tilde{\mu}^{\downarrow y} = \mu^{\downarrow y}$, hence $\tilde{\mu} = \mu$. □

The combination of extended matrices associated with Gaussian potentials can be summarised in the following way:

- sweep the extended matrices forward to the same but arbitrary value of the shared variables,
- add the mean vectors and matrices, and
- sweep backwards on the same value.

This is a remarkable property of the Gaussian distribution.

In order to compute the combination of extended matrices, only the submatrices corresponding to the common variables z have to be inverted according to equation (3.37). This can be carried over to Gaussian potentials: By using equation (B.10), instead of equation (3.10), the mean of the combination of two Gaussian potentials can be computed more efficiently

$$\mu = \begin{pmatrix} {\mu_1}^{\downarrow x} + {K_1}^{\downarrow x^{-1}}{K_1}^{\downarrow x,z}({\mu_1}^{\downarrow z} - \mu_z) \\ {\mu_2}^{\downarrow y} + {K_2}^{\downarrow y^{-1}}{K_2}^{\downarrow y,z}({\mu_2}^{\downarrow z} - \mu_z) \\ \mu_z \end{pmatrix} \tag{3.40}$$

Chapter Synopsis & Discussion

In Section 3.3, the proof that Gaussian potentials form a valuation algebra was direct. However, there are at least two indirect proofs.

- In (Kohlas, 2003), it is proved that probability densities form a valuation algebra. Therefore, since Gaussian potentials correspond to Gaussian densities and since they are closed under combination and marginalisation, this results an indirect proof of the theorem.

- In (Kohlas and Monney, 2008), it is proved that Gaussian potentials correspond to precise Gaussian hints, which belong to the valuation algebra of Gaussian hints. Again, since Gaussian potentials are closed under combination and marginalisation, this yields a second indirect proof of the theorem.

The extension of the valuation algebra of Gaussian potentials by neutral elements corresponds to the subalgebra of only those Gaussian hints whose focal sets are parallel to the axes of (a part of) the variables.

The combination of moment matrices has been derived from the product-intersection rule (Liu, 1996a; 1999). This gives an alternative proof of the compatibility of the combination of Gaussian potentials and moment matrices (Lemma 3.8).

Furthermore, alternative formulas for marginalisation and combination have been derived for the "direct" formulas of (Kohlas, 2003; Kohlas and Monney, 2008):

- computing the marginal concentration matrix according to (3.17) requires the inversion of a matrix corresponding to the variables to be eliminated;

- computing the combined mean according to (3.40) requires the inversion of two matrices corresponding to the variables occurring in one factor only.

Further, the marginalisation of moment matrices is easy, whereas the combination requires the inversion of two matrices corresponding to the common variables.

4

Join Trees and Local Computation

Information may come in pieces, since it may either become available only little by little in an interactive system, or since its modular structure allows to decompose it. In order to capture the overall knowledge, the different pieces have to be combined. One may then be interested in what can be said about a certain subdomain of this bundled piece of knowledge. This subdomain will be called a *query*. However, this process of first combining and then marginalising information pieces may be intractable or at least computationally inefficient: In the case of relational databases, for instance, this intuitive approach would result in a huge table with lots of redundancy (i.e. functional dependencies). Therefore, by an astute interplay of marginalisations and combinations, the construction of such a huge table can be avoided, and computations are done locally on smaller domains which do not exceed those of the factors. The combination axiom and its derivative Lemma 2.4 are the key to these local computation schemes. In a first step, valuations are assigned to the nodes of a *join tree*, which has the *running intersection property*, i.e. a variable occurring in two nodes resides in every node on the path between these two nodes. Then, messages (i.e. marginals to the intersection with the child node) are sent towards a designated *root node*, whose label contains the query. When all messages have arrived, the root node, marginalised to the query, will contain the same information as obtained by first combining all pieces and then marginalising them to the query. This scheme is called *collect algorithm* (Kohlas, 2003). If there are several queries, a join tree has to be constructed which covers all queries; then, the same messages as in the collect algorithm can be computed for every root node. However, the messages cannot be added to the nodes' content, but they are stored in separate mailboxes associated with the edges, one for each direction. This scheme is called *Shenoy-Shafer* architecture (Shenoy and Shafer, 1990). However, if there are inverses in a valuation algebra, the message does not need to be stored separately since the node's content can be divided out before sending a message. This scheme is called *Lauritzen-Spiegelhalter* architecture (Lauritzen and Spiegelhalter, 1988).

Applying these techniques to Gaussian potentials, one can reduce the space requirements from order $\mathcal{O}(m^2)$ to $\mathcal{O}(\max\{{m_i}^2\})$ (m being the number of all variables, and m_i being the number of variables of the factor i).

Chapter Outline

In Section 4.1, the *query problem* is formally defined, and *covering join trees* are introduced as a graphical structure bearing the different valuations of a *knowledge base*. How a join tree can be constructed for such a factorisation by use of the *fusion* algorithm is set out in Section 4.2. The *collect* algorithm for answering a single query is discussed in Section 4.3, the *Shenoy-Shafer* architecture for several queries in Section 4.4. If valuations can be divided out in the valuation algebra, this property can be used for more efficient computations in the *Lauritzen-Spiegelhalter* architecture of Section 4.5. Finally, if marginalisation is only partially defined, not every required marginal for local computation may be defined. If the factorisation forms a *construction sequence*, there is always such a scheduling of the local computation algorithms that all required marginals are defined. This is the topic of Section 4.6.

4.1 Covering Join Trees

Given a *knowledge base* consisting of several pieces of information, one may be interested in several *questions* or *queries*.

DEFINITION **4.1**. *A set $\{\phi_1, \ldots, \phi_n\}$ is called a* factorisation *of $\phi \in \Phi$ if $\phi = \phi_1 \otimes \cdots \otimes \phi_n$. A factorisation $\{\phi_1, \ldots, \phi_n\}$ of ϕ together with a set of* queries *$q_1, \ldots, q_k \in \mathcal{M}(\phi)$ is called a* projection problem. *If there is only one query q, then the projection problem is called* simple. *The factors $\phi_1, \ldots, \phi_n$ form the* knowledge base *of the projection problem.* ⊘

The direct approach to query answering is to combine all factors to compute ϕ and then to marginalise ϕ to the queries. In many applications, the *complexity* increases *exponentially* with the number of variables involved. For instance, computing the marginal of a probability potential involves exponentially many terms in the summation. Furthermore, the *storage space* required to represent the valuations may increase exponentially, too; for instance, a probability table for n binary variables has 2^n entries. In the case of Gaussian potentials, an upper bound of space and time requirements is $\mathcal{O}(m^2)$ where m is the number of variables involved.

However, there are efficient generic methods (such as the Shenoy-Shafer (Shenoy and Shafer, 1990), the Lauritzen-Spiegelhalter (Lauritzen and Spiegelhalter, 1988) and the HUGIN (Jensen et al., 1990) architectures), which are all based on join trees, where computations can be carried out locally on smaller domains. Join trees express how the different information pieces are related, for instance *functional dependencies* in relational databases and *conditional independences* in joint probability distributions. Furthermore, if there are several queries, a join tree allows to avoid redundant computations.

In order to formalise join trees, some basic graph-theoretic notions are introduced first.

DEFINITION **4.2**. *An* (undirected) graph *is a pair $G = (V, E)$, where*

- *the elements of V are called* vertices *or* nodes *and*
- *the elements of E are two-element sets $\{i, j\}$, $i, j \in V$, and are called* edges.

If two vertices $i, j \in V$ are connected by an edge $e = \{i, j\} \in E$, then they are called neighbours *or* adjacent. *The set of all neighbours of a vertex i is denoted $ne(i)$. A* path *from a node i_1 to a node i_n is a sequence of different edges $e_1, e_2, \ldots, e_{n-1}$ such that $e_k = \{i_k, i_{k+1}\} \in E$ for all $k \in \{1, \ldots, n\}$. Notice that, if there is path $e_1, \ldots, e_{n-1}$ from i_1 to i_n, then $e_{n-1}, \ldots, e_1$ is a path from i_n to i_1. Therefore, without loss of information, a path from i to j is also said to be a path* between *i and j. A graph $G = (V, E)$ is* connected *whenever there is a path between two different nodes, i.e. if $i \neq j$ implies that there is a path $e_1, \ldots, e_n$ such that $i \in e_1$ and $j \in e_n$. A* tree *is a connected graph with the property that removing an edge leads to an unconnected graph. A* labelled tree *is a quartuple (V, E, λ, D) where*

- *(V, E) is a tree,*
- *D is a lattice of domains,*
- *$\lambda : V \to D$ assigns the* label *$\lambda(v)$ to every node $v \in V$.* ⊘

Remark **4.3**. In a tree $T = (V, E)$, there is a unique path between distinct nodes $i, j \in V$. This can be proved as follows: Since a tree is connected, there is a path between any distinct nodes $i, j \in V$. Assume the contraposition, i.e. that there are two paths between two nodes $i, j \in V$. Then, there is at least one edge $e = \{k, l\}$ not common to both paths. When that edge is removed, the graph remains connected since the gap can be bridged by the path from k to i, over the alternative path to l and from there to j. Therefore, T cannot be a tree. ⊘

Example **4.4** (Tree). Consider the graph $G = (V, E)$ of Figure 4.1: The vertices $V = \{1, \ldots, 5\}$ are represented by dots labelled by the vertex, and the edges $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{2, 5\}\}$ are represented by straight lines. For any pair of vertices, there is a path between them in the graph G: for instance, the path $\{1, 2\}, \{2, 5\}$ connects the nodes 2 and 5. Furthermore, if any edge is removed, G becomes disconnected: for instance, after removing the edge $\{3, 4\}$, there is no path from the isolated node 4 to any other node of the modified graph. Therefore, G is a tree. Furthermore, the graph obtained by adding the edge $\{4, 5\}$ is not a tree since this adds a second path from 4 to 5. Finally, the graph $(V \cup \{6\}, E)$ is not connected since there is no path from the isolated node 6 to any other node of the graph. ⊘

Definition **4.5**. *A labelled tree (V, E, λ, D) satisfies the* running intersection property *(or* Markov property *or* join tree property*) if, for every variable X common to the label of distinct nodes $i, j \in V$, i.e. for $X \in \lambda(i) \cap \lambda(j)$, and for every node k on the path from i to j, it holds that X is also in the node k's label, i.e. that $X \in \lambda(k)$. A* join tree *(or* Markov tree*) is a labelled tree satisfying the running intersection property.* ⊘

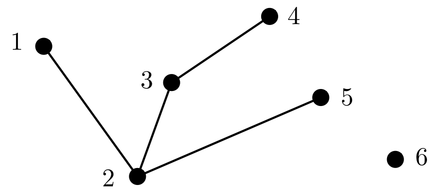


FIGURE 4.1: A graph a part of which forms a tree

EXAMPLE **4.6** (JOIN TREE). The vertices of the tree in Example 4.4 can be labelled by the variables $r = \{A, B, C, D\}$ forming the lattice $D = 2^r$ as shown in Figure 4.2, where each node i is depicted by an oval bearing the label $\lambda(i)$. The quartuple (V, E, λ, D) forms a labelled tree, satisfying the running intersection property since the variable A, for instance, is contained in the labels of vertices 1, 2, and 5, and it is contained in all possible paths between these nodes. However, adding the variable C to the label of node 3, then 3 and 5 would both contain C in their label, but not 2, which is on the path between 3 and 5, hence, the modified tree would not satisfy the running intersection property. ⊘

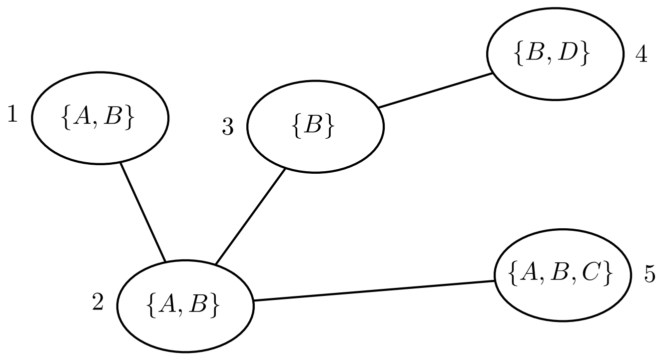


FIGURE 4.2: A tree satisfying the running intersection property

DEFINITION **4.7**. *A join tree $JT = (V, E, \lambda, D)$ is called a* covering join tree *for a factorisation $\{\phi_1, \ldots, \phi_n\}$ if, for each factor ϕ_i, $i \in \{1, \ldots, n\}$, there is a node $j \in V$ such that $d(\phi_i) \subseteq \lambda(j)$; j is then said to* cover *ϕ_i. A function $a : \{\phi_1, \ldots, \phi_m\} \to V$ is called an* assignment mapping *if $d(\phi_i) \subseteq \lambda(a(\phi_i))$ for all $i \in \{1, \ldots, m\}$.* ⊘

EXAMPLE **4.8** (ASSIGNMENT MAPPING). Figure 4.3 shows an assignment mapping from the knowledge base $\phi_1, \ldots, \phi_6$ to the join tree of Example 4.6: The valuations are depicted by diamonds bearing their domain. The assignment of a valuation to a node is shown by an arrow from the valuation to the node. ⊘

The following lemma shows that every assignment mapping a induces a surjective assignment mapping a' and a new factorisation as shown by the following lemma.

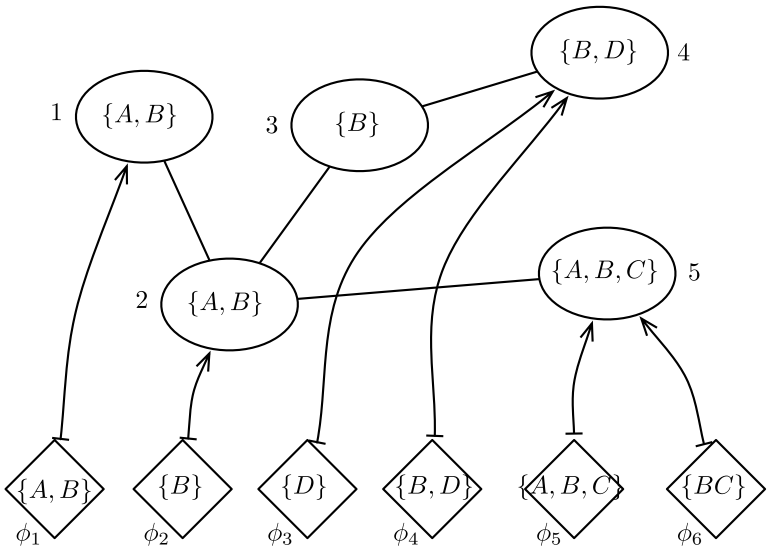


Figure 4.3: An assignment of a knowledge base to a join tree

Lemma **4.9**. *Let $a : \{\phi_1, \ldots, \phi_n\} \rightarrow V$ be an assignment mapping to the covering join tree (V, E, λ, D). For $j \in \{1, \ldots, m = |V|\}$, define*

$$\psi_j = \begin{cases} \bigotimes_{\phi_i : a(\phi_i) = j} \phi_i & \text{if } \exists \phi_i : a(\phi_i) = j, \\ e & \text{else} \end{cases}$$

and

$$a'(\psi_j) = j.$$

Then, a' is a bijective assignment mapping for the new factorisation $\{\psi_1, \ldots, \psi_m\}$ of $\phi = \phi_1 \otimes \cdots \otimes \phi_n$. ⊘

Proof. In light of the identity and the commutativity axioms and since a assigns every ϕ_i to exactly one node j,

$$\begin{aligned} \bigotimes_{j \in \{1,\ldots,m\}} \psi_j &= \bigotimes_{j \in \{1,\ldots,m\}, \phi_i : a(\phi_i) = j} \phi_i \\ &= \bigotimes_{i \in \{1,\ldots,n\}} \phi_i \quad = \quad \phi. \end{aligned}$$

□

In light of this lemma, it can be assumed without loss of generality that an assignment mapping is bijective.

Definition **4.10**. *Let a be a bijective assignment mapping $a : \{\psi_1, \ldots, \psi_m\} \rightarrow V$. Then,*

- *$\lambda(j)$ is called the* label of the node *$j \in V$ and*

- $\omega_j = d(a^{-1}(j))$ *is called the* domain *of the valuation assigned to node* $j \in V$*.* ⊘

Notice the difference between the *label* $\lambda(j)$ and the *domain* ω_j of a node j: By the definition of an assignment mapping,

$$\omega_j \subseteq \lambda(j),$$

i.e. the valuation $\psi_j = a^{-1}(j)$ assigned to the node j is not required to fill the node completely. The label may be seen as the "capacity" of the node, as the largest domain it may cover.

EXAMPLE **4.11** (BIJECTIVE ASSIGNMENT MAPPING). The assignment mapping of Example 4.8 induces a bijective assignment mapping for a new factorisation $\psi_1, \ldots, \psi_5$ as shown in Figure 4.4: Several valuations assigned to the same node are combined ($\psi_4 = \phi_3 \otimes \phi_4$ and $\psi_5 = \phi_5 \otimes \phi_6$); if there is no factor assigned to a node, then the identity element is assigned to it instead ($\psi_3 = e$); all other assignments are not changed ($\psi_1 = \phi_1$ and $\psi_2 = \phi_2$). ⊘

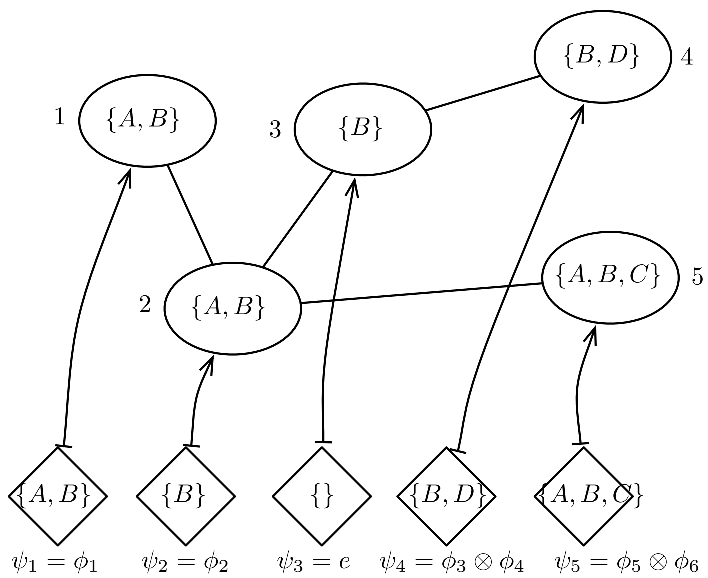


FIGURE 4.4: Induced bijective assignment mapping

DEFINITION **4.12**. *A join tree* (V, E, λ, D) *is a* covering join tree *for the projection problem consisting of the knowledge base* $\{\phi_1, \ldots, \phi_n\}$ *and the queries* $\{q_1, \ldots, q_k\}$ *if*

1. *it is a covering join tree for the factorisation* $\{\phi_1, \ldots, \phi_n\}$,
2. *it covers the domains* q_j, $j \in \{1, \ldots, k\}$, *and*
3. $\bigcup_{i=1}^{n} d(\phi_i) = \bigcup_{j \in V} \lambda(j)$. ⊘

The third point in this definition requires that every variable occurring in a node of the join tree also occurs in one of the factors. However, the nodes' labels need not be filled up by the factors (initially) assigned to them.

4.2 Fusion Algorithm for Join Tree Construction

In order to solve a projection problem, the first step is to find a covering join tree for it. Notice that it is always possible to assign all factors to one node, which results in a trivial join tree. Since the complexity of the operations or the required storage space may increase exponentially in the number of variables involved, one might want to find a join tree whose biggest label has minimal cardinality. This problem is known to be NP-hard (Arnborg et al., 1987). However, the fusion algorithm (Shenoy, 1992) can be used to compute join trees: Given a factorisation $\Phi = \{\phi_1, \ldots, \phi_n\}$ of ϕ and a sequence $X_1, \ldots, X_k$ of all variables $d(\phi) - t = \{X_1, \ldots, X_k\}$, variables are successively eliminated, i.e.

$$\phi^{\downarrow t} = \left(\left(\phi^{-X_1} \right)^{\cdots} \right)^{-X_k}, \tag{4.1}$$

Such a sequence is therefore called an *elimination sequence*.

Definition **4.13**. *Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ be a factorisation of ϕ and let $x = d(\phi)$. Then, a sequence $X_1, \ldots, X_k$ such that $x = \{X_1, \ldots, X_k\}$ is called an eliminiation sequence.* ⊘

The fusion algorithm eliminates variables *locally*: Consider eliminating Y in $\phi = \phi_1 \otimes \cdots \otimes \phi_n$; here, the factors not containing Y remain unchanged in light of the combination axiom, so the factors that do contain Y in their domain may first be combined and Y may then be eliminated from their combination, i.e.

$$\phi^{-Y} = \left(\bigotimes_{i:Y \in d(\phi_i)} \phi_i \right)^{-Y} \otimes \bigotimes_{i:Y \notin d(\phi_i)} \phi_i. \tag{4.2}$$

Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ and

$$\psi = \left(\bigotimes_{i:Y \in d(\phi_i)} \phi_i \right)^{-Y}.$$

Then, in terms of

$$\mathrm{Fus}_Y(\Phi) = \{\psi\} \cup \{\phi_i : Y \notin d(\phi_i)\}, \tag{4.3}$$

it thus holds that

$$\phi^{\downarrow t} = \bigotimes \mathrm{Fus}_{X_k}(\cdots (\mathrm{Fus}_{X_1}(\Phi))).$$

The fusion algorithm can be used to construct a join tree from a factorisation (Shafer, 1996; Kohlas, 2003) as given in the pseudo-code of Algorithm 1: At each step i, the variable X_i is eliminated and the labels of remaining factors containing X_i in their label are removed from the list l, their union is added to the list l, and the other labels remain untouched in the list l. The following lemma shows that the fusion algorithm produces a covering join tree for the initial factorisation Φ; see (Kohlas, 2003).

Lemma **4.14**. *The graph $G = (V, E)$ constructed by the fusion Algorithm 1 is a covering join tree for Φ.* ⊘

Algorithm 1: *Fusion*

input: a set of domains $\{x_1, \ldots, x_n\}$ and a sequence $X_1, \ldots, X_k$ of all variables occurring in a domain x_j $(j \in \{1, \ldots, n\})$

output: a covering join tree (V, E, λ, D)

$V := \emptyset, E := \emptyset, l := \{x_1, \ldots, x_n\}$
loop for i **from** 1 **to** k
do

1. $v_i := \bigcup\{x \in l : X_i \in x\}$
2. $v'_i := v_i - \{X_i\}$
3. $V := V \cup \{v_i, v'_i\}$
4. $E := E \cup \{\{v_i, v'_i\}\} \cup \{\{v, v_i\} : \ v \in l \cap V, X_i \in v\}$
5. $l := l \cup \{v'_i\} - \{x \in l : X_i \in x\}$

done
$\lambda(v) := v$

PROOF. 1. Notice that, at the start of step i, the domains in the list l do not contain the eliminated variables $X_1, \ldots, X_{i-1}$. This trivially holds for $i = 1$. Assume the claim holds at the beginning of step $i-1$; then all domains containing X_i are removed from the list and the only domain v'_i added does not contain X_i by definition, hence the claim holds indeed at the start of step i.

2. Observe that, at the end of the algorithm, the list l contains only the empty set, $l = \{\emptyset\}$. On the one hand, by the previous observation, at the end of step k, the list does not contain any non-empty domain, i.e. $l \subseteq \{\emptyset\}$; on the other hand, $v'_k = \emptyset$ and $v_k \in l$. Thence, $l = \{\emptyset\}$.

3. It will now be proved that every factor x_j in Φ is covered by a node of G. Initially $x_j \in l$; after each step i, either $x_j \in l$ or $x_j \subseteq v_i$, not both. As soon as $X_i \in x_j$, then $x_j \subseteq v_j$ and then x_j is covered by G; since the iteration goes over all variables, finally all nodes will be covered.

4. It will now be proved that G is a tree. Since $l = \{v'_k\}$ at the end and since every vertex being removed from the list is connected to some vertex remaining in the list, there is a path from any vertex to the node v_k; since also $(v_k, v'_k) \in E$, the graph G is connected. On the other hand, notice that there are only edges (v_i, v'_i) and (v'_i, v_j) with $i < j$, $i, j \in \{1, \ldots, k\}$; therefore, an edge of the first type (v_i, v'_i) lies in every path from the vertices $v_1, v'_1, \ldots, v_{i-1}, v'_{i-1}$ to v_k and an edge of the second type (v'_i, v_j) lies in every path from the ver-

tices $v_1, v'_1, \ldots, v_{j-1}, v'_{j-1}$ to v_k, hence removing any edge makes the graph disconnected. Thence, G is indeed a tree.

5. It remains to be proved that G satisfies the running intersection property. Observe that a variable X_i may only occur in nodes v_j where $j \leq i$ or v'_j where $j < i$. By construction, for $j < i$, if it is in v_j, then it is in v'_j, and it is in every node on the unique path from v_j to v_i, where it is eliminated in the edge to v'_j. Therefore, whenever X_i is in v_j, then it is in v_i; hence, the unique path between two nodes v_{j1}, v_{j2} containing X_i goes through v_i, whence X_i is contained in every node in the unique path between v_{j1} and v_{j2}. This shows that G indeed satisfies the running intersection property. □

Several heuristics for finding an elimination sequence resulting in a "good" join tree have been proposed in the literature, attempting to minimise the size of the largest label of the join tree obtained by the fusion algorithm; see (Lehmann, 2001) for an overview of such heuristics.

Example **4.15** (Fusion Algorithm). Figure 4.5 shows the join tree constructed by the fusion algorithm for the elimination sequence C, D, A, B; each one of the six diamonds stands either for a valuation or a query; the nodes of the tree are depicted by ovals and the edges by straight lines; to the right of the dashed vertical lines, there are only nodes and valuations which do not contain the variable in their label; an arrow shows a possible assignment of the valuation to a covering node, and corresponds to eliminating the valuation from the list l in Algorithm 1. ⊘

4.3 Collect Algorithm

The collect algorithm (Kohlas, 2003; Schneuwly et al., 2004) allows to solve a *simple projection problem* on a corresponding join tree covering all factors $\phi_1, \ldots, \phi_m$ of the knowledge base and also the single query q. The factors are supposed to belong to a valuation algebra with *full marginalisation*. The collect algorithm can be viewed as a message-passing scheme between distributed processors, the nodes of the covering join tree, to a selected node covering q, called the root node; messages are marginals of the node content, which are then combined with the node content of the receiving node. At the end of the collect algorithm, the root node will contain the full information of the knowledge base with respect to the query.

As seen above, it can be assumed without loss of generality that the assignment mapping $a : \phi_1, \ldots, \phi_m \to V$ is bijective from the factors of the knowledge base to the vertices of the join tree (V, E, λ, D), i.e. $m = |V|$. The vertices will be denoted by integers $1, \ldots, m$. A tree (V, E) with nodes $V = \{1, \ldots, m\}$ is said to be directed towards the node m if $j > i$ for every node $j \in V$ lying on the path between i and m. By a suitable permutation π of the nodes V, a join tree can be assumed directed towards the node $r = \pi^{-1}(m)$ covering the query q, i.e. $q \subseteq \lambda(r)$; the permutation can be constructed using Algorithm 2. The correctness of the algorithm is straightforward.

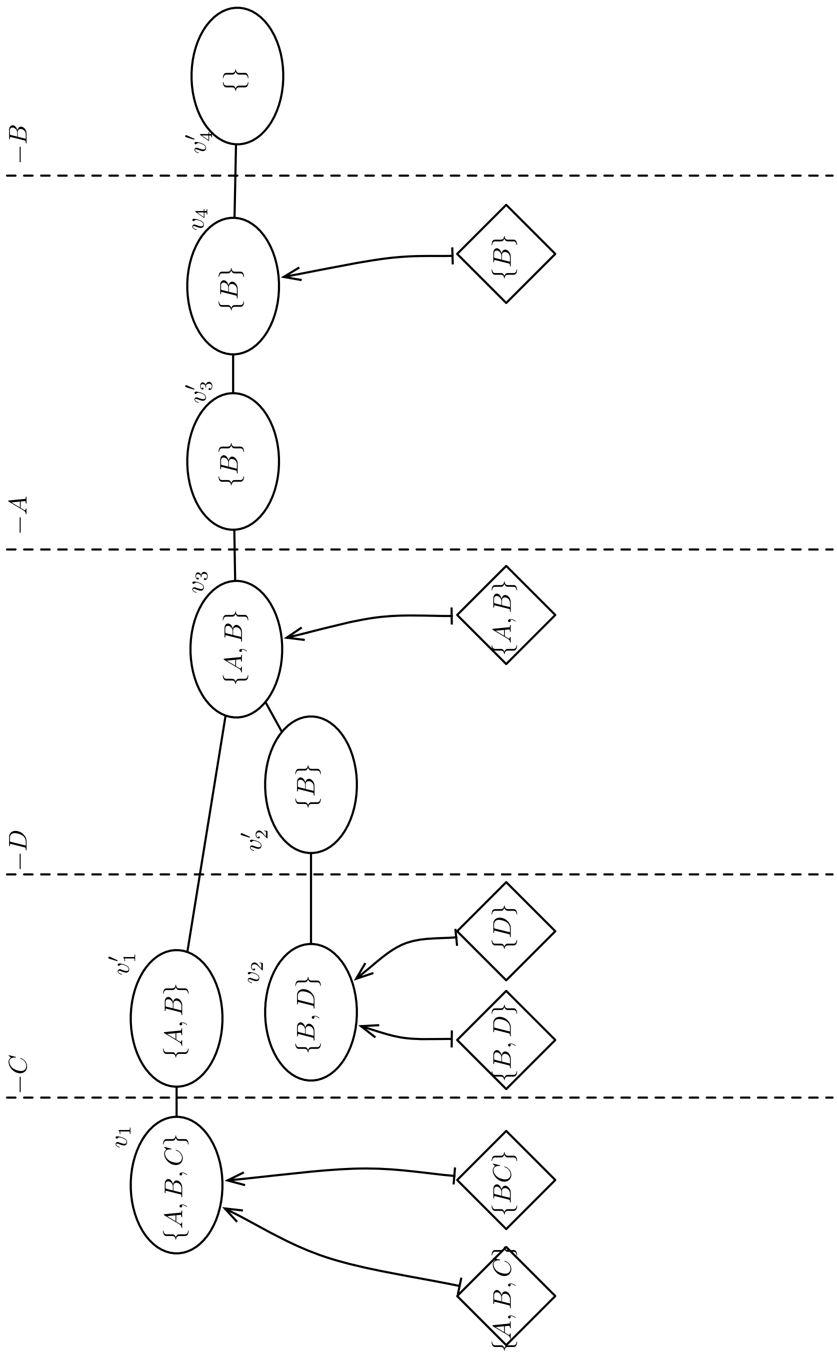


FIGURE 4.5: The covering join the constructed by the fusion algorithm

Algorithm 2: *Direct Join Tree*

→ input: a covering join tree (V, E, λ, D) for a simple projection problem $\phi_1, \ldots, \phi_m$ with query q, a root node $r \in V$ covering q (i.e. $q \subseteq \lambda(r)$).

← output: a permutation π such that $\pi^{-1}(r) = m$ and $\pi^{-1}(j) > \pi^{-1}(i)$ for every node $j \in V$ lying on the path between i and m

---

Done := $\emptyset$; *Next* := $\{r\}$; i:=m
loop until *Done* == V
do

1. *Pre* := *Next*; *Next* := $\emptyset$
2. **for** j **in** *Pre*
   do
   $\pi(j) = i$; $i := i - 1$; *Next* := *Next* $\cup$ $ne(j)$ − *Done*
   done
3. *Done* := *Done* $\cup$ *Pre*

done

In a directed tree (V, E) with the root node $m = |V|$, terms of human ancestry and the tree metaphor are used to denominate particular nodes and relations between nodes:

DEFINITION **4.16**. *The* **parents** *of a node i are the nodes*

$$pa(i) = \{j : \ j < i, \{i, j\} \in E\}.$$

A node without parents is called **leaf (node)**. *The* **child** *$ch(i)$ of a node $i \in V$ is the unique node j with $j > i$ and $\{i, j\} \in E$. The nodes on the path from a node i to the root node are called* **descendants** *of i, the nodes on a path from a node i to a leaf its* **ancestors**. ⊘

The reader should not be misled by the facts that in this terminology a child may have none, only one or even several parents and that a directed tree is rooted at the youngest member of the family.

EXAMPLE **4.17**. Consider the tree in Figure 4.1 (see Example 4.4). The tree is directed since the nodes on the paths from the leaves 1 and 4 to the root node 5 are increasing. Node 2 has parents $pa(2) = \{1, 3\}$, ancestors $\{1, 3, 4\}$ and its child (and unique descendant) is $ch(2) = 5$. ⊘

With this terminology, the collect algorithm for solving the simple projection can now be specified. Let (V, E, λ, D) be a covering join tree directed towards the root node $m = |V|$ for a simple projection problem consisting of the knowledge base $\phi_1, \ldots, \phi_m$ and the query $q \subseteq \lambda(m)$, and let a be a bijective assignment mapping, $a : \{\phi_1, \ldots, \phi_m\} \rightarrow V$. In the collect algorithm, each node acts according to the following two rules:

R1 If a message from a parent arrives, combine it with your content.

R2 If every parent has sent a message and it is combined with your own content, marginalise your content to the intersection of its domain and the domain of your child, and set that message to your child.

According to these rules, every node except the root node sends exactly one message to its unique child; leaves can send immediately. This leads to $m-1$ messages in a run of the collect algorithm. Since combination is commutative, the messages from the parent nodes can be combined in any order. Thus, different schedulings of the collect algorithm lead to the same result. Since the covering join tree is directed, the parents of a node i all have a number $j < i$, so node i becomes ready to send after all nodes $j < i$ have sent their message, i.e. the nodes may send their messages in ascending order of their node number. This particular scheduling leads to the following description of the collect algorithm in m steps. Let

$$\psi_j^{(i)}$$

be the content of the node $j \in V = \{1, \ldots, m\}$ before step $i \in \{1, \ldots, m\}$ of the collect algorithm; in particular, the initial contents are given by

$$\psi_j^{(1)} = a^{-1}(j).$$

Since a node's content and its domain may change during a run of the collect algorithm, the domain of a node j's content before step i is referred to as

$$\omega_j^{(i)} = d(\psi_j^{(i)}).$$

Then, at each step i, the node i computes the message

$$\mu_{i \to ch(i)} = \psi_i^{(i)\downarrow \omega_i^{(i)} \cap \lambda(ch(i))}, \tag{4.4}$$

which is sent to the child $ch(i)$ with node label $\lambda(ch(i))$. In node $ch(i)$, upon reception, the message is combined with the old content and the result is stored as the new content,

$$\psi_{ch(i)}^{(i+1)} = \psi_{ch(i)}^{(i)} \otimes \mu_{i \to ch(i)}; \tag{4.5}$$

all other nodes do not change their content,

$$\psi_k^{(i+1)} = \psi_k^{(i)}, \qquad k \neq ch(i). \tag{4.6}$$

At the end of the collect algorithm, the root node m is filled up (i.e. $d(\psi_m^{(m)} = \lambda(m))$ since all variables of $\lambda(m)$ occur in some factor in light of the third condition of Definition 4.12. Furthermore, the root node's content reflects the overall information of the knowledge base with respect to $\lambda(m)$.

Theorem **4.18**. *A the end of the collect algorithm, the root node m contains the marginal of $\phi = \phi_1 \otimes \cdots \otimes \phi_m$ relative to $\lambda(m)$, i.e.*

$$\psi_m^{(m)} = \phi^{\downarrow \lambda(m)}.$$

In particular,

$$\psi_m^{(m)\downarrow q} = \phi^{\downarrow q}.$$

⊘

PROOF. The first assertion is proved in (Schneuwly et al., 2004; Schneuwly, 2007). The second claim then follows by the transitivity axiom. □

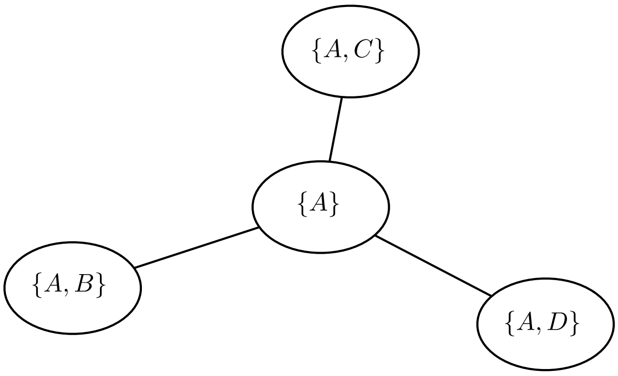


FIGURE 4.6: A join tree for the four variables A, B, C, D

EXAMPLE **4.19** (COLLECTING GAUSSIAN POTENTIALS). Consider the covering join tree of Figure 4.6 and the corresponding *adjacency matrix* for the variables A, B, C, D in Figure 4.7, where the grey-shaded areas show the pairs of variables which are in the same node in the join tree. It is a covering join tree for three Gaussian potentials with domains $d(\phi_1) = \{A, D\}$, $d(\phi_2) = \{A, C\}$ and $d(\phi_3) = \{A, B\}$ and the query $q = \{A, B\}$. The assignments are shown in Figure 4.8. By this factorisation, the storage cost is reduced from order $\mathcal{O}(4^2) = \mathcal{O}(16)$ to $\mathcal{O}(3 \cdot 2^2 + 1^2) = \mathcal{O}(12)$. For simplicity's sake, assume that the mean vectors are all zero. Let the concentration matrices be K_1, K_2 and K_3, respectively. Then, using the alternative formula (3.17) for marginalisation of Gaussian potentials, $(\phi_1 \otimes \phi_2 \otimes \phi_3)^{\downarrow\{A,B\}}$ is given by

$$K^{\downarrow\{A,B\}} - K^{\downarrow\{A,B\},\{C,D\}} {K^{\downarrow\{C,D\}}}^{-1} K^{\downarrow\{C,D\},\{A,B\}}$$

for

$$K = {K_1}^{\uparrow\{A,B,C,D\}} + {K_2}^{\uparrow\{A,B,C,D\}} + {K_3}^{\uparrow\{A,B,C,D\}}.$$

Notice that the non-zero entries of K correspond exactly to the grey-shaded areas in the adjacency matrix of Figure 4.6.

However, it is not necessary to build the *sparse matrix* K since, using the combination axiom,

$$(\phi_1 \otimes \phi_2 \otimes \phi_3)^{\downarrow\{A,B\}} = {\phi_1}^{\downarrow\{A\}} \otimes {\phi_2}^{\downarrow\{A\}} \otimes \phi_3,$$

which is given by

$$\tilde{K}_1^{\uparrow\{A,B\}} + \tilde{K}_2^{\uparrow\{A,B\}} + K_3$$

where

$$\tilde{K}_1 = {K_1}^{\downarrow\{A\}} - {K_1}^{\downarrow\{A\},\{D\}} {{K_1}^{\downarrow\{D\}}}^{-1} {K_1}^{\downarrow\{D\},\{A\}}$$

and

$$\tilde{K}_2 = {K_2}^{\downarrow\{A\}} - {K_2}^{\downarrow\{A\},\{C\}} {{K_2}^{\downarrow\{C\}}}^{-1} {K_2}^{\downarrow\{C\},\{A\}}.$$

Thereby, the time complexity is reduced from $\mathcal{O}(2^2)$ to $\mathcal{O}(\max\{1^2, 1^2\}) = \mathcal{O}(1)$. This can be directly translated into an execution of the collect algorithm as shown in Figure 4.8: first, the messages ${\phi_1}^{\downarrow\{A\}}$ and ${\phi_2}^{\downarrow\{A\}}$ are sent to the node with label $\{A\}$, where they are combined with the node's content, the identity element e; then, the message ${\phi_1}^{\downarrow\{A\}} \otimes {\phi_2}^{\downarrow\{A\}}$ is sent to the root node with label $\{A, B\}$ where it is finally combined with ϕ_3. ⊘

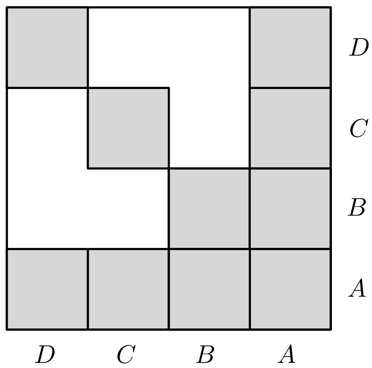


Figure 4.7: The adjacency matrix for the join tree of Figure 4.6

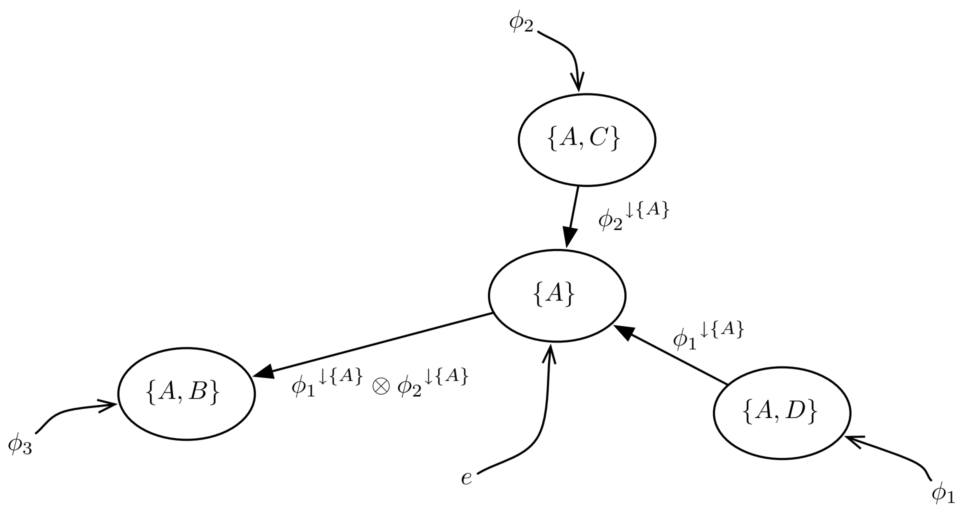


Figure 4.8: An execution of the collect algorithm towards the node with label $\{A, B\}$

4.4 Shenoy-Shafer Architecture

In order to solve a general projection problem with several queries $q_1, \ldots, q_k$, the collect algorithm could be run for each query one after another. However, the join tree covers all queries and some of the messages of the collect algorithm for the first root node q_1 covering the first query, $r_1 \subseteq \lambda(r_1)$, some messages for a second root node r_2 covering q_2 remain the same. Therefore, those messages could be cached and then re-used. In fact, if a message $\mu_{i \to ch(i)}$ has already been computed, then it can be used for propagating messages towards any other root node. Therefore, the Shenoy-Shafer architecture (Shenoy and Shafer, 1990; Schneuwly et al., 2004) introduces *mailboxes* on every edge to store the messages. The vertices compute the messages for every root node according to the following rule:

R A node i sends a message to its neighbour j as soon as a message has arrived from all other neighbours; the message is the combination of the initial content ψ_i with the messages from all other nodes, which is then marginalised to the intersection of its domain with the receiving neighbour's label.

The message sent from a node i to a neighbour j in the Shenoy-Shafer architecture is therefore

$$\mu_{i \to j} = \left(\psi_i \otimes \bigotimes_{k \in ne(i), k \neq j} \mu_{k \to i} \right)^{\downarrow \omega_{i \to j} \cap \lambda(j)}$$

where

$$\omega_{i \to j} = d(\psi_i) \cup \bigcup_{k \in ne(i), k \neq j} d(\mu_{k \to i}).$$

Theorem **4.20**. *At the end of the message passing in the Shenoy-Shafer architecture,*

$$\phi^{\downarrow \lambda(i)} = \psi_i \otimes \bigotimes_{j \in ne(i)} \mu_{j \to i}. \tag{4.7}$$

⊘

Proof. Since the messages $\mu_{j \to i}$ are the same as for the collect algorithm with the root node i, the assertion follows from Theorem 4.18. □

After the execution of the Shenoy-Shafer algorithm, every query q_k can be answered in a node i covering $q_k \subseteq \lambda(i)$ since

$$\phi^{\downarrow q_k} = \left(\phi^{\downarrow \lambda(i)} \right)^{\downarrow q_k} = (\psi_i \otimes \bigotimes_{j \in ne(i)} \mu_{j \to i})^{\downarrow q_k}.$$

in light of the transitivity of axiom.

Example **4.21** (Redirecting a Join Tree). Figure 4.9 shows the same tree directed to two different root nodes (after permutation π); an arrow from node i pointing to j is to be read "$i < j$." Only one message has to be computed for the new root nodes, the other three are the same as for the first root node. ⊘

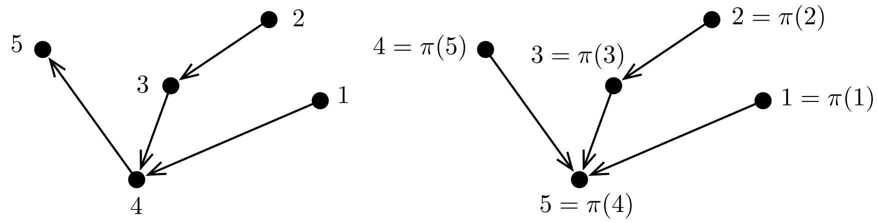


FIGURE 4.9: Most of the messages can be re-used for a different root node.

Although the Shenoy-Shafer algorithm has been described in terms of rules that the nodes have to apply independently, it can also be described in terms of a particular scheduling induced by choosing an arbitrary node as root node and (re)directing the join tree towards that node $m = |V|$. Then, as in the collect algorithm, the nodes can send their message in ascending order of their node number i; in this first phase, called collect phase or inward propagation, messages are sent towards the particular root node. After this first phase, nodes may send in decreasing order of their node number. This second phase, called distribute phase or outward propagation, messages are sent from the root node to the leaves.

EXAMPLE **4.22** (SHENOY-SHAFER ALGORITHM WITH GAUSSIAN POTENTIALS). The join tree of Example 4.19 may be redirected towards the other nodes. The corresponding messages are shown in Figure 4.10. ⊘

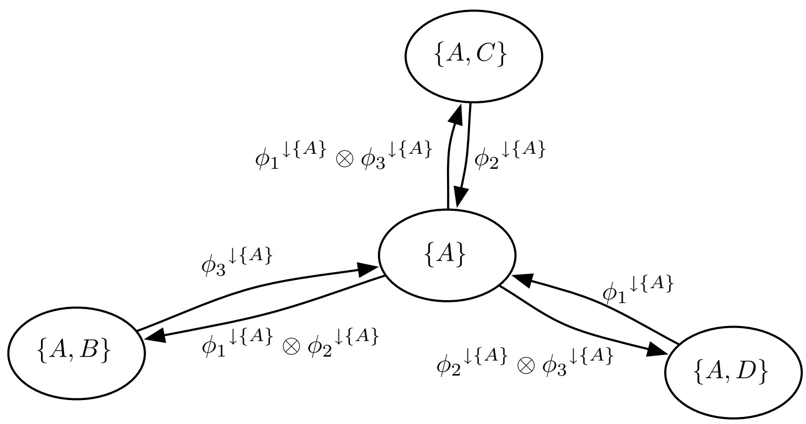


FIGURE 4.10: The messages in the Shenoy-Shafer architecture

4.5 Lauritzen-Spiegelhalter Architecture

In the collect algorithm, the messages are immediately combined with the receiving node's content at arrival. In contrast, in the Shenoy-Shafer architecture, all the messages are stored in mailboxes on the edges and only combined when a node becomes ready to send. The reason why the messages cannot be directly combined is that the messages from the inward propagation phase must not be sent in the outward propagation phase. However, if the messages from the inward propagation phase can be divided out before sending the message of the outward propagation phase, then the messages of the inward propagation phase could be directly combined with the receiving node's content as in the collect algorithm and then divided out before the outward propagation phase of the Shenoy-Shafer architecture. This scheme is called Lauritzen-Spiegelhalter architecture (Lauritzen and Spiegelhalter, 1988; Schneuwly et al., 2004) and can be used to solve a projection problem in a valuation algebra with *division* (see Section 2.8).

The Lauritzen-Spiegelhalter architecture will now be described in terms of an inward/collect and an outward/distribute phase as the Shenoy-Shafer architecture. In the first phase of the Lauritzen-Spiegelnhalter architecture, the same messages as in the Shenoy-Shafer architecture are used: Every node i sends the message

$$\mu_{i\to ch(i)} = \psi_i^{(i)\downarrow\omega_i^{(i)}\cap\lambda(ch(i))} \tag{4.8}$$

to its unique child $ch(i)$, where

$$\psi_i^{(i)} = \psi_i \otimes \bigotimes_{j\in pa(i)} \mu_{j\to i}.$$

As in the collect algorithm, the message is combined with receiver node's content on reception,

$$\psi_{ch(i)}^{(i+1)} = \psi_{ch(i)}^{(i)} \otimes \mu_{i\to ch(i)}; \tag{4.9}$$

however, the sender divides the message out after sending its message,

$$\psi_i^{(i+1)} = \psi_i^{(i)} \otimes {\mu_{i\to ch(i)}}^{-1}, \tag{4.10}$$

all other nodes do not change their content,

$$\psi_k^{(i+1)} = \psi_k^{(i)}, \qquad k \neq i, ch(i). \tag{4.11}$$

The inward propagation phase is terminated when the root node m has received all messages from its parents. In the outward propagation phase, when a node has received the message from its child, it combines the message with its content and sends the message

$$\mu_{j\to i} = (\psi_j^{(m)} \otimes \mu_{ch(i)\to i})^{\downarrow\lambda(j)\cap\lambda(i)} \tag{4.12}$$

to every parent $j \in pa(i)$.

Theorem **4.23**. *At the end of the Lauritzen-Spiegelhalter architecture, every node i contains $\phi^{\downarrow\lambda(i)}$.* ⊘

Proof. The scheduling of the outward propagation is irrelevant. The proof goes by induction over the particular scheduling $m, \ldots, 1$, where step i means that the node i has received the message from its child.

Since the messages are the same as in the Shenoy-Shafer architecture, it follows by the correctness of the Shenoy-Shafer architecture, Theorem 31, that the root node m contains $\phi^{\downarrow\lambda(m)}$ at the end of the inward propagation phase. That is the basis of the induction.

Assume the induction hypothesis holds, i.e., at step i, the node i has received the message

$$\mu_{ch(i)\to i} = \phi^{\downarrow\lambda(ch(i))\cap\lambda(i)}.$$

Then, by Lemma 2.19 (6),

$$\gamma(\mu_{i\to ch(i)}) \leq \gamma(\mu_{ch(i)\to i} \otimes \mu_{i\to ch(i)}) \leq \gamma(\phi^{\downarrow\lambda(ch(i))\cap\lambda(i)}) \leq \gamma(\phi^{\downarrow\lambda(i)})$$

hence, by the correctness of the Shenoy-Shafer architecture and Lemma 2.38 (3),

$$\begin{aligned}
\psi_i^{(m)} \otimes \mu_{ch(i)\to i} =& \psi_i \otimes \bigotimes_{k\in pa(i)} \mu_{k\to i} \otimes \mu_{i\to ch(i)}{}^{-1} \otimes \phi^{\downarrow\lambda(ch(i))\cap\lambda(i)} \\
=& \psi_i \otimes \bigotimes_{k\in pa(i)} \mu_{k\to i} \otimes \mu_{i\to ch(i)}{}^{-1} \otimes \bigotimes_{k\in ne(ch(i))} \mu_{k\to ch(i)} \\
=& \psi_i \otimes \bigotimes_{k\in pa(i)} \mu_{k\to i} \otimes \bigotimes_{k\in ne(ch(i)),k\neq i} \mu_{k\to ch(i)} \otimes f_{\gamma(\mu_{i\to ch(i)})} \\
=& \psi_i \otimes \bigotimes_{k\in pa(i)} \mu_{k\to i} \otimes \mu_{ch(i)\to i} \otimes f_{\gamma(\mu_{i\to ch(i)})} \\
=& \psi_i \otimes \bigotimes_{k\in ne(i)} \mu_{k\to i} \otimes f_{\gamma(\mu_{i\to ch(i)})} \\
=& \phi^{\downarrow\lambda(i)} \otimes f_{\gamma(\mu_{i\to ch(i)})} \\
=& \phi^{\downarrow\lambda(i)}.
\end{aligned}$$

□

4.6 Local Computation in Valuation Algebras with Partial Marginalisation

So far, it has been assumed that local computation is performed in valuation algebras with full marginalisation, i.e. that marginals exist for all subsets of a valuation's domain. If marginalisation is only partially defined, it has to be verified that the marginals required as messages exist in the collect algorithm and in the Shenoy-Shafer architecture. In an even slightly more general context, it has been shown in (Schneuwly et al., 2004; Lemma 12) that all the messages for the Shenoy-Shafer algorithm exist whenever the collect algorithm can be executed. Furthermore, if the Shenoy-Shafer algorithm can be executed, then Lauritzen-Spiegelhalter can also be executed (Schneuwly et al., 2004; Theorem 4). Hence, the problem of whether the messages for local computation exist is reduced to finding a scheduling for the collect algorithm.

A particular case where there is such a scheduling are *construction sequences* (Shafer, 1996; Kohlas, 2003), which generalise the chain rule of probability calculus:

a sequence $\phi_1, \ldots, \phi_n$ of conditionals such that the combination $\phi_1 \otimes \cdots \otimes \phi_i$ ($1 \leq i \leq n$) of every initial subsequence is a density with full marginalisation again. Such factorisations arise in particular from Bayesian networks. The following definitions are a further generalisation of such construction sequences (Kohlas, 2003).

Definition **4.24**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a valuation algebra* with division. *Then, a valuation $\phi \in \Phi$ with domain $d(\phi) = h \cup t$, $h \cap t = \emptyset$ is a* kernel *for h given t if*

- $t \in \mathcal{M}(\phi)$ *and*
- $\phi^{\downarrow t} = f_{\gamma(\phi^{\downarrow t})}$.

A kernel for h given $\emptyset$ is called density. *A* construction sequence *is a sequence $\phi_1, \phi_2, \ldots, \phi_n$ of kernels with heads h_i and tails t_i, such that*

- $t_1 = \emptyset$, $t_i \subseteq d(\phi_1) \cup \cdots \cup d(\phi_{i-1})$,
- h_i *disjoint from* $d(\phi_1) \cup \cdots \cup d(\phi_{i-1})$, *and*
- $\gamma({\phi_i}^{\downarrow t_i}) \leq \gamma(\phi_1 \otimes \cdots \otimes \phi_{i-1})$. ⊘

Lemma **4.25**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \mathcal{M}, \downarrow)$ be a valuation algebra with division.*

(1) *Every marginal of a density is well defined and is a density as well.*

(2) *Let $\phi_1, \ldots, \phi_n$ be a construction sequence. Then, for $i = 1, \ldots, n$,*

$$\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_i$$

is a density and

$$\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_i = (\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_n)^{\downarrow d(\phi_1) \cup \cdots \cup d(\phi_i)}.$$ ⊘

Proof. (1) Let $\phi \in \Phi$ be a density. Since $\emptyset \in \mathcal{M}(\phi)$, the transitivity axiom implies that the marginal for every $s \subseteq d(\phi)$ exists, i.e. $s \in \mathcal{M}(\phi)$. Since $\phi^{\downarrow \emptyset} = f_{\gamma(\phi^{\downarrow \emptyset})}$, the transitivity axiom implies that $\emptyset \in \mathcal{M}(\phi^{\downarrow s})$ and that $(\phi^{\downarrow s})^{\downarrow \emptyset} = \phi^{\downarrow \emptyset} = f_{\gamma(\phi^{\downarrow \emptyset})}$, hence $\phi^{\downarrow s}$ is a density as well.

(2) The first claim holds by induction over i. It clearly holds for $i = 1$. Assume that $\phi_1 \otimes \cdots \otimes \phi_{i-1}$ is a density for some $i \in \{2, \ldots, n\}$. Since $t_i = d(\phi_1 \otimes \cdots \otimes \phi_{i-1}) \cap d(\phi_i) \in \mathcal{M}(\phi_i)$, the combination axiom implies that $d(\phi_1) \cup \cdots \cup d(\phi_{i-1}) \in \mathcal{M}(\phi_1 \otimes \cdots \otimes \phi_i)$ and

$$\begin{aligned}(\phi_1 \otimes \cdots \otimes \phi_i)^{\downarrow d(\phi_1) \cup \cdots \cup d(\phi_{i-1})} &= \phi_1 \otimes \cdots \otimes \phi_{i-1} \otimes {\phi_i}^{\downarrow t_i} \\ &= \phi_1 \otimes \cdots \otimes \phi_{i-1} \otimes f_{\gamma({\phi_i}^{\downarrow t_i})} \\ &= \phi_1 \otimes \cdots \otimes \phi_{i-1},\end{aligned}$$

using the combination axiom and Lemma 2.38 (3) with the third condition of a construction sequence. Hence, using the induction hypothesis and the transitivity axiom, $\emptyset \in \mathcal{M}(\phi_1 \otimes \cdots \otimes \phi_i)$ and

$$(\phi_1 \otimes \cdots \otimes \phi_i)^{\downarrow\emptyset} = ((\phi_1 \otimes \cdots \otimes \phi_i)^{\downarrow d(\phi_1)\cup\cdots\cup d(\phi_{i-1})})^{\downarrow\emptyset},$$

which by the induction hypothesis is the identity element of its group. Hence, $\phi_1 \otimes \phi_2 \otimes \cdots \otimes \phi_i$ is indeed a density.
Therefore, $\phi_1 \otimes \cdots \otimes \phi_n$ is a density, hence the second claim holds for $i = n$. Assume that the claim holds for some $i \in \{2, \ldots, n\}$. Then, using the transitivity, the induction hypotesis, the combination axioms and Lemma 2.38 (3) with the third condition of a construction sequence,

$$\begin{aligned}(\phi_1 \otimes \cdots \otimes \phi_n)^{\downarrow d(\phi_1)\cup\cdots\cup d(\phi_{i-1})} =& ((\phi_1 \otimes \cdots \otimes \phi_n)^{\downarrow d(\phi_1)\cup\cdots\cup d(\phi_i)})^{\downarrow d(\phi_1)\cup\cdots\cup d(\phi_{i-1})} \\ =& (\phi_1 \otimes \cdots \phi_i)^{\downarrow d(\phi_1)\cup\cdots\cup d(\phi_{i-1})} \\ =& \phi_1 \otimes \cdots \otimes \phi_{i-1} \otimes {\phi_i}^{\downarrow t_i} \\ =& \phi_1 \otimes \cdots \otimes \phi_{i-1} \otimes f_{\gamma({\phi_i}^{\downarrow t_i})} \\ =& \phi_1 \otimes \cdots \otimes \phi_{i-1}.\end{aligned}$$

By induction, this shows that the second claim also holds. □

If the factors $\psi_1, \ldots, \psi_m$ assigned to a join tree form a construction sequence, then the collect algorithm can be executed in that order. Since the messages in the subtree T_i correspond to the collect algorithm with root node i in that tree, it follows from Theorem 4.18 that

$$\psi_i^{(i)} = \bigotimes_{\psi_j \in T_i} \psi_j.$$

Since the initial subsequence $\psi_1, \ldots, \psi_i$ is a construction sequence, it follows from Lemma 4.25 that $\psi_1 \otimes \cdots \otimes \psi_i$ is a density, hence, using the combination axiom,

$$(\psi_1 \otimes \cdots \otimes \psi_i)^{\downarrow d(T_i)} = \bigotimes_{\psi_j \in T_i} \psi_j \otimes (\bigotimes_{\psi_j \notin T_i} \psi_j)^{\downarrow\emptyset} = \psi_i^{(i)} \otimes (\bigotimes_{\psi_j \notin T_i} \psi_j)^{\downarrow\emptyset}$$

for $d(T_i) = \bigcup_{\psi_j \in T_i} d(\psi_j)$. Therefore, using the transitivity and the combination axioms, every marginal of $\psi_i^{(i)}$ is well defined and hence node i is ready to send at step i.

Chapter Synopsis & Discussion

The presentation and notation used here closely follow (Schneuwly et al., 2004; Schneuwly, 2007).

The term join tree is borrowed from database theory (Maier, 1983). In other domains, join trees have different names: *qualitative Markov trees* (Shenoy and

Shafer, 1986) and *hypertrees* (Shenoy and Shafer, 1990) in the context of belief-function propagation, *junction trees* (Jensen et al., 1990) and *clique trees* (Lauritzen and Spiegelhalter, 1988) in the field of probabilistic inference and expert systems.

In (Shenoy, 1997), *binary* join trees are introduced, i.e. join trees which have at most three neighbours. It is always possible to convert a join tree into a binary one by introducing additional nodes, see for instance (Lehmann, 2001). When a join tree is made binary, there are more nodes in the tree, hence more marginalisations and additional storage space may be needed for the new nodes (Kohlas and Shenoy, 2000). However, these additional requirements may be outweighed by several improvements in efficiency.

1. *Redundant combinations in the Shenoy-Shafer architecture*: In the Shenoy-Shafer architecture, an incoming message $\mu_{k\to i}$ to a node i appears in every outgoing message $\mu_{i\to j}$ from node i with node valuation ψ_i to any other neighbour $j \neq k$, i.e. the factors ψ and $\mu_{k\to i}$ have to be combined several times; in a binary join tree, this number is cut down to 2.

2. *Locality*: The size of the domains of the nodes may be reduced; in particular, even the tree width (i.e. the maximum size of the node domains) may be reduced as observed in (Kohlas and Shenoy, 2000). Roughly speaking, the operations take place on smaller domains and also require less storage space. More precisely,

   - the domain of the combinations are smaller. This is desirable when valuations are enumerations since the combination may then cut down on the number of elements, for instance in relational databases.
   - Furthermore, when the messages are computed, some variables may not be propagated and marginalised out earlier.

   As an example, take Figure 4.11 (Pouly, 2008): The four valuations $\alpha, \beta, \gamma, \delta$ with domains $d(\alpha) = \{S, T\}$, ... are assigned to a non-binary join tree. Figure 4.12 shows a binary join tree for the same knowledge base. When messages are propagated to the root node $\{U, V\}$, variable T is already eliminated in the message coming to the node $\{S, V, U\}$, while this is not the case in the first join tree.

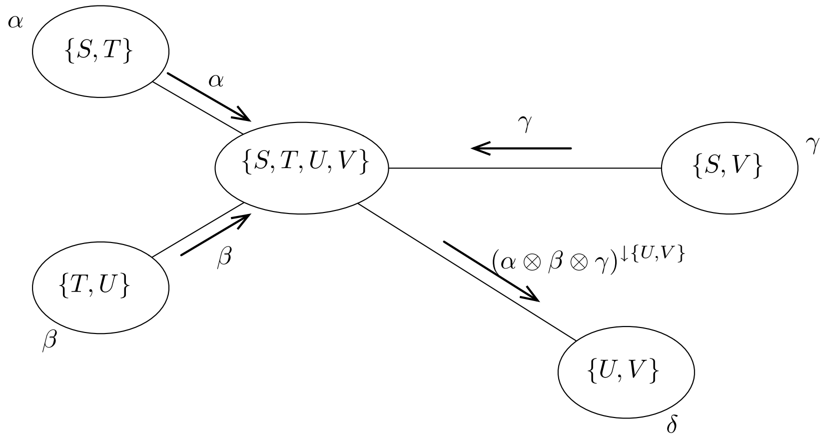


FIGURE 4.11: A non-binary join tree with four valuations

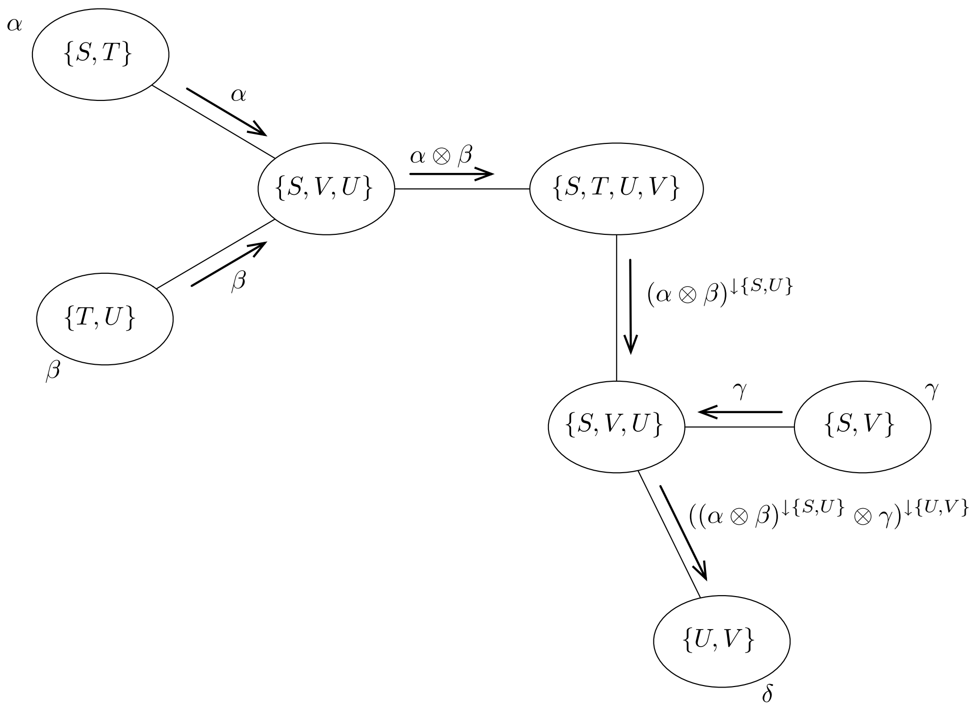


FIGURE 4.12: A binary join tree for the same four valuations

Part II

Conditional Gaussian Densities

5

Conditional Gaussian Densities

Probabilistic statements often make no sense without a specification of the conditions under which they hold. For instance, weather forecasts depend on the current weather and the expert's experience; the success of medical treatment depends on the patient's age, health etc. Therefore, (Rényi, 1970) concludes that *every probability is a conditional probability.* Consequently, the primary objects of his treatise are *conditional probability spaces.* Similarly, in Bayesian networks, a family of conditional distributions form a factorisation of a joint distribution (due to conditional independences expressed in the network). This may be called the *engineer's* point of view. In contrast, mainstream probability theory usually takes probability spaces as their primary objects, from which conditional distributions are derived. This may be called the *mathematician's* point of view.

The starting point of statistical reasoning is often a *statistical specification*: a parameterised family of probability distributions on the observation space. Given an observation, statistical inference seeks to make statements about the unknown parameter. In this chapter, a particular type of statistical specification is introduced: *conditional Gaussian densities.* Conditioning a Gaussian density $\phi = \phi_{\mu,K}$ with domain $d(\phi) = x \cup z$, $x \cap z = \emptyset$, on an event "$z = \mathbf{z}$" results in the Gaussian density $\phi_{x|\mathbf{z}}$ on x,

$$\phi_{x|\mathbf{z}}(\mathbf{x}) = \phi_{\mu(\mathbf{z}),K^{\downarrow x}}(\mathbf{x}) \tag{5.1}$$

for

$$\mu(\mathbf{z}) = \mu^{\downarrow x} - {K^{\downarrow x}}^{-1} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z}), \tag{5.2}$$

see Appendix B.4. Different observations of the variables z lead to the family

$$\{\phi_{x|\mathbf{z}} : \mathbf{z} \in \mathbb{R}^z\} \tag{5.3}$$

of densities on x (one for each value $\mathbf{z} \in \mathbb{R}^z$). This family constitutes the conditional Gaussian density (CGD) function

$$\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \phi_{x|\mathbf{z}}(\mathbf{x}) = \phi_{\mu(\mathbf{z}),K^{\downarrow x}}(\mathbf{x}). \tag{5.4}$$

This function is a density on the *head* variables x given a fixed value for the *tail* variables z, but not on both arguments.

A related notion is the likelihood function

$$l_{\mathbf{x}}(\mathbf{z}) = \phi_{x|\mathbf{z}}(\mathbf{x}|\mathbf{z}) \tag{5.5}$$

of the observations $\mathbf{x}$ under a fixed parameter $\mathbf{z}$. However, the likelihood function is not a probability density on z, and it makes no sense to normalise it. So the interpretation of the likelihood function is delicate, and one must be careful not to interpret likelihoods as probabilities. The Bayesian answer is to require a prior distribution on z and to derive the posterior distribution on z given an observation of x. An alternative approach is provided by the theory of hints (Kohlas and Monney, 1995; Monney, 2003; Kohlas and Monney, 2008), which obtains a similar notion in a completely different way with clear semantics. This will be further explored in Chapters 6 and 7.

Chapter Outline

Three alternative approaches are introduced in this chapter:

1. *algebraically*, a conditional Gaussian density can be expressed by the quotient of the full density ϕ divided by the marginal density $\phi^{\downarrow z}$ (in Section 5.1);

2. *geometrically*, the *regression* of the dependent variables x on z can be modelled as a Gaussian linear system (in Section 5.2);

3. *analytically*, a conditional Gaussian density can be represented by a vector and the symmetric non-negative definite "pseudo-concentration matrix" which is the difference of the concentration matrices of the full density and the marginal density (in Section 5.3).

In this chapter, these three approaches will be motivated, but only briefly sketched. They will be worked out in the subsequent Chapters 6–9.

In the following discussion, a Gaussian potential $\phi = (\mu, K)$ and the corresponding Gaussian density $\phi_{\mu,K}$ will often be used interchangeably. For instance, for $\phi = (\mu, K)$, $\phi_{\mu,K}(\mathbf{x})$ will often be abbreviated as $\phi(\mathbf{x})$.

5.1 The Algebraic Approach

A conditional Gaussian density $\phi_{x|z}$ can always be represented by the pair $(\phi, \phi^{\downarrow z})$ of Gaussian potentials, as shown by the following theorem.

THEOREM **5.1**. *For a Gaussian potential $\phi = (\mu, K) \in \mathcal{G}$ with domain $d(\phi) = x \cup z$ such that $x \cap z = \emptyset$, it holds that*

$$\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{\phi(\mathbf{x}, \mathbf{z})}{\phi^{\downarrow z}(\mathbf{z})}. \tag{5.6}$$

⊘

PROOF. Let $\mu(\mathbf{z})$ as in equation (5.2). Then,

$$\begin{aligned}
\phi_{x|z}(\mathbf{x}|\mathbf{z}) &= \phi_{x|\mathbf{z}}(\mathbf{x}) \\
&= \sqrt{\frac{|\det(K^{\downarrow x})|}{(2\pi)^{|x|}}}e^{-\frac{1}{2}(\mathbf{x}-\mu(\mathbf{z}))'K^{\downarrow x}(\mathbf{x}-\mu(\mathbf{z}))} \\
&= \sqrt{\frac{|\det(K^{\downarrow x})|}{(2\pi)^{|x|}}}e^{-\frac{1}{2}\left[\begin{array}{l}(\mathbf{x}'-(\mu^{\downarrow x})',\mathbf{z}'-(\mu^{\downarrow z})')K\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}\\ -(\mathbf{z}'-(\mu^{\downarrow z})')(K^{\downarrow z}-K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z})(\mathbf{z}-\mu^{\downarrow z})\end{array}\right]} \\
&= \frac{\sqrt{\frac{|\det(K)|}{(2\pi)^{|x|+|z|}}}e^{-\frac{1}{2}(\mathbf{x}'-(\mu^{\downarrow x})',\mathbf{z}'-(\mu^{\downarrow z})')K\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}}}{\sqrt{\frac{|\det(K^{\downarrow z}-K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z})|}{(2\pi)^{|z|}}}e^{-\frac{1}{2}(\mathbf{z}'-(\mu^{\downarrow z})')(K^{\downarrow z}-K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z})(\mathbf{z}-\mu^{\downarrow z})}} \\
&= \frac{\phi(\mathbf{x},\mathbf{z})}{\phi^{\downarrow z}(\mathbf{z})}
\end{aligned}$$

because, in light of Theorem 13.3.8 of (Harville, 1997; p.188),

$$\det(K) = \det(K^{\downarrow x}) \cdot \det(K^{\downarrow z} - K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z}).$$ □

Such pairs will be called *conditional Gaussian potentials.*

DEFINITION **5.2**. *Let ϕ be a Gaussian potential and let $z \subseteq d(\phi)$ and $x = d(\phi) - z$. Then, the pair $(\phi, \phi^{\downarrow z})$ is called* conditional Gaussian potential *(CGP) or* conditional *of ϕ for x given z. The first element of the pair is called* numerator *and the second element* denominator. *The variables x are called the* head *and the variables z the* tail *of the conditional Gaussian potential. The set of all conditional Gaussian potentials shall be denoted $\mathcal{G}_c$.* ⊘

Notice that there are clearly different conditional Gaussian potentials which represent the same conditional Gaussian density, for instance $(\phi\otimes\psi, \psi^{\downarrow z}\otimes\psi)$ for $\phi, \psi \in \mathcal{G}$ with $d(\phi) = x \cup z$ and $d(\psi) = z$ since, using Theorem 3.3 and observing that the normalisation constant is the same in the numerator and the denominator,

$$\frac{\phi\otimes\psi(\mathbf{x},\mathbf{z})}{(\phi\otimes\psi)^{\downarrow z}(\mathbf{z})} = \frac{\phi\otimes\psi(\mathbf{x},\mathbf{z})}{\phi^{\downarrow z}\otimes\psi(\mathbf{z})} = \frac{\phi(\mathbf{x},\mathbf{z})\cdot\psi(\mathbf{x})}{\phi^{\downarrow z}(\mathbf{z})\cdot\psi(\mathbf{x})} = \frac{\phi(\mathbf{x},\mathbf{z})}{\phi^{\downarrow z}(\mathbf{z})} = \phi_{x|z}(\mathbf{x}|\mathbf{z}).$$

In other words, the two different (full) densities ϕ and $\psi' = \phi\otimes\psi$ on $x \cup z$ both induce the same conditional Gaussian density on x given z, i.e. $\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \psi'_{x|z}(\mathbf{x}|\mathbf{z})$ for all $\mathbf{x} \in \mathbb{R}^x$, $\mathbf{z} \in \mathbb{R}^z$.

Conditional Gaussian densities and conditional Gaussian potentials will often be used interchangeably, using the conventions of Table 5.1.

abbreviation	for
$\phi(\mathbf{x})$	$\phi_{\mu,K}(\mathbf{x})$
$\phi_{x\|z}$	$(\phi, \phi^{\downarrow z})$
$\phi_{x\|z}(\mathbf{x}\|\mathbf{z})$	$(\phi_{\mu,K})_{x\|z}(\mathbf{x}\|\mathbf{z})$

TABLE 5.1: Notational conventions for a Gaussian potential $\phi = (\mu, K) \in \mathcal{G}$

5.2 The Geometric Approach

The geometric approach can be motivated as follows. In light of Lemma B.1, the conditional Gaussian density $\phi_{x|z}$ is given by

$$\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \phi_{x|\mathbf{z}}(\mathbf{x}) = \phi_{\mu(\mathbf{z}),K^{\downarrow x}}(\mathbf{x}). \tag{5.7}$$

for $\mu(\mathbf{z}) = \mu^{\downarrow x} - {K^{\downarrow x}}^{-1} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z})$, representing the family $\{\phi_{x|\mathbf{z}}\}_{\mathbf{z}\in\mathbb{R}^z}$ of Gaussian densities on x. In light of equation (5.7), the points

$$\Gamma(\omega) = \{(\mathbf{x}, \mathbf{z}) \in \mathbb{R}^{x\cup z} : \omega = \mathbf{x} - \mu^{\downarrow x} + {K^{\downarrow x}}^{-1} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z})\} \tag{5.8}$$

have the *same conditional density* value

$$\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \phi_{x|\mathbf{z}}(\omega).$$

Since the sets $\Gamma(\omega)$ are parallel linear manifolds of dimension $|z|$ in $\mathbb{R}^{x\cup z}$ covering $\mathbb{R}^{x\cup z}$, conditional Gaussian densities represent a (full) distribution over these parallel linear manifolds $\Gamma(\omega)$. This gives a *geometric* flavour to conditional Gaussian densities.

The linear regression equation

$$\mathbf{x} = \mu^{\downarrow x} - {K^{\downarrow x}}^{-1} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z}) + \omega \tag{5.9}$$

can be derived from $\phi_{x|z}$ for each ω. So far, these equations describe only points of the same conditional density, and the ω index the disjoint sets $\Gamma(\omega)$. But what is the motivation for considering these sets $\Gamma(\omega)$ of points of the same conditional density?

Equation (5.9) can be rewritten as

$$\mathbf{x} + {K^{\downarrow x}}^{-1} K^{\downarrow x,z}\mathbf{z} - \omega = y \tag{5.10}$$

for $y = \mu^{\downarrow x} + {K^{\downarrow x}}^{-1} K^{\downarrow x,z}(\mathbf{z} - \mu^{\downarrow z})$. This defines a linear function of $\mathbf{x}$, $\mathbf{z}$ and ω. This mapping may be interpreted in a prescriptive way, mapping the unknown parameter $(\mathbf{x}, \mathbf{z})$ and the unknown disturbance ω to the observation y on the right-hand side. In other words, if the unknown parameter $(\mathbf{x}, \mathbf{z})$ and the unknown disturbance ω were the true one, this would necessarily have generated the observation y. Thus, the set $\Gamma(\omega)$ consists of those $(\mathbf{x}, \mathbf{z})$ which are compatible with the observation y and the assumption ω given the functional model (5.10): If the assumption ω is correct, one of the parameters in $\Gamma(\omega)$ must be the correct one since the observation follows the model (5.10).

In summary, if the regression equations (5.9) are interpreted as a functional model, the sets $\Gamma(\omega)$ get clear semantics. Furthermore, the sets of ω which make a hypothesis necessarily true or which only make a hypothesis possibly true can be used to evaluate hypotheses on the parameters. Such assumption-based reasoning on Gaussian linear models as (5.10) is discussed in detail in the Chapters 6 and 7.

5.3 The Analytic Approach

A conditional Gaussian density can be expressed by

$$\begin{aligned} \phi_{x|z}(\mathbf{x}|\mathbf{z}) \quad &= \quad \phi_{x|\mathbf{z}}(\mathbf{x}) \\ =&\sqrt{\frac{|\det(K^{\downarrow x})|}{(2\pi)^{|x|}}} e^{-\frac{1}{2}(\mathbf{x}'-\mu^{\downarrow x'}+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}})K^{\downarrow x}(\mathbf{x}-\mu^{\downarrow x}+K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}))} \\ =&\sqrt{\frac{|\det(K^{\downarrow x})|}{(2\pi)^{|x|}}} e^{-\frac{1}{2}(\mathbf{x}'-\mu^{\downarrow x'},\mathbf{z}'-\mu^{\downarrow z})'\tilde{K}\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}} \end{aligned} \tag{5.11}$$

in terms of the matrix

$$\tilde{K} = (K - \left(K^{\downarrow z} - K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z}\right)^{\uparrow x\cup z}),$$

which is symmetric but only non-negative definite (and thus singular) in general. This holds since

$$\begin{aligned} &(\mathbf{x}'-\mu^{\downarrow x'}+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}})K^{\downarrow x}(\mathbf{x}-\mu^{\downarrow x}+K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z})) \\ =&(\mathbf{x}'-\mu^{\downarrow x'})K^{\downarrow x}(\mathbf{x}-\mu^{\downarrow x}) \\ &+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x}(\mathbf{x}-\mu^{\downarrow x})+(\mathbf{x}'-\mu^{\downarrow x'})K^{\downarrow x}K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}) \\ &(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x}K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}) \\ =&(\mathbf{x}'-\mu^{\downarrow x'})K^{\downarrow x}(\mathbf{x}-\mu^{\downarrow x}) \\ &+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}(\mathbf{x}-\mu^{\downarrow x})+(\mathbf{x}'-\mu^{\downarrow x'})K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}) \\ &+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}) \\ =&(\mathbf{x}'-\mu^{\downarrow x'},\mathbf{z}'-\mu^{\downarrow z'})K\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}-(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z}(\mathbf{z}-\mu^{\downarrow z}) \\ &+(\mathbf{z}'-\mu^{\downarrow z'})K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z}(\mathbf{z}-\mu^{\downarrow z}) \\ =&(\mathbf{x}'-\mu^{\downarrow x'},\mathbf{z}'-\mu^{\downarrow z'})K\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}-(\mathbf{z}'-\mu^{\downarrow z'})(K^{\downarrow z}-K^{\downarrow z,x}K^{\downarrow x^{-1}}K^{\downarrow x,z})(\mathbf{z}-\mu^{\downarrow z}) \\ =&(\mathbf{x}'-\mu^{\downarrow x'},\mathbf{z}'-\mu^{\downarrow z'})\tilde{K}\begin{pmatrix}\mathbf{x}-\mu^{\downarrow x}\\ \mathbf{z}-\mu^{\downarrow z}\end{pmatrix}. \end{aligned} \tag{5.12}$$

Since $C = \tilde{K}$ appears in the exponent of (5.11) in the same way as a concentration matrix in the formula of an ordinary Gaussian density, C will be called *pseudo-concentration matrix*. On the other hand, the reason for using the *pseudo-mean* *vector* $\nu = \tilde{K}\mu$ cannot be fully understood yet and will become clear only below in Chapter 9. However, the idea can be illustrated by means of a simple special case: A Gaussian potential (μ, K) is represented by the symmetric Gaussian potential $(K\mu, K)$. In this representation, it suffices to add the pseudo-mean vectors and the concentration matrices, which avoids inverting the combined concentration matrix. More generally, it turns out that the combination carried over from Gaussian hints (and equivalently from Gaussian quotients) induces the multiplication of conditional Gaussian densities (irrespective of head and tail) up to a constant factor, and hence induces the addition of pseudo-mean vectors and pseudo-concentration matrices of symmetric Gaussian potentials. Notice that the pair $(\nu, C) = (\tilde{K}\mu, \tilde{K})$ determines the function $\phi_{x|z}$ up to a constant factor since the exponent in equation (5.12) for $\mathbf{u} = \binom{\mathbf{x}}{\mathbf{z}}$ can be developed further as

$$(\mathbf{u}' - \mu')\tilde{K}(\mathbf{u} - \mu) = \mathbf{u}'\tilde{K}\mathbf{u} - 2\mathbf{u}\tilde{K}\mu + \mu'\tilde{K}\mu = \mathbf{u}'C\mathbf{u} - 2\mathbf{u}'\nu + \mu'\tilde{K}\mu,$$

which yields that the terms depending on $\mathbf{u}$ are determined by (ν, C) and the remaining summand $\mu'\tilde{K}\mu$ results in a constant factor of the conditional Gaussian density. Therefore, the pair $(\nu, C) = (\tilde{K}\mu, \tilde{K})$, representing the conditional Gaussian density $\phi_{x|z}$, will be called *symmetric Gaussian potential* (associated with $\phi_{x|z}$). More generally, symmetric Gaussian potentials can be defined as follows.

DEFINITION **5.3**. *Let $x \in D$ be a finite set of variables. A pair $\phi = (\nu, C)$, $\nu \in \mathbb{R}^x$, $C \in \mathbb{R}(x, x)$, C symmetric, is called **symmetric Gaussian potential**, and x is called its* domain*, denoted $d(\phi) = x$. The vector ν is called **pseudo-mean vector** and the matrix c is called **pseudo-concentration matrix**. The set of all symmetric Gaussian potentials is denoted Δ.* ⊘

They will be studied in Chapter 9.

Chapter Synopsis & Discussion

A conditional Gaussian density $\phi_{x|z}$ – the family of densities $\phi_{x|\mathbf{z}}$ obtained by conditioning the same Gaussian distribution with density ϕ on the different values $\mathbf{z}$ of the same set z of variables – can be represented threefold:

- algebraically, by the pair $(\phi, \phi^{\downarrow z})$ of Gaussian potentials since $\phi_{x|z}(\mathbf{x}|\mathbf{z}) = \frac{\phi(\mathbf{x},\mathbf{z})}{\phi^{\downarrow z}(\mathbf{z})}$;
- geometrically, by the Gaussian hint obtained from the regression equation (5.9);
- analytically, by a pseudo-mean vector and a symmetric pseudo-concentration matrix which is the difference of the concentration matrix of ϕ and $\phi^{\downarrow z}$ resulting from the division of ϕ by $\phi^{\downarrow z}$.

In Chapter 7, it will be shown that different conditional Gaussian densities may be related to the same Gaussian hint: The head and tail variables for the focal manifolds can be chosen in several ways. Since the head variables correspond to different axes of integration, the inducing conditional Gaussian densities are equal up to a constant normalisation factor. It will be shown that Gaussian hints and conditional Gaussian densities are in one-to-one correspondence as well as how the operations of combination and marginalisation can be carried over from Gaussian hints to conditional Gaussian potentials. In Chapter 8, the same operations will be introduced in conditional Gaussian potentials in a more general algebraic setting. Finally, in Chapter 9, an equivalent valuation algebra will be defined in terms of symmetric Gaussian potentials. It will be shown that symmetric Gaussian potentials provide a canonical way of representing (equivalent) Gaussian hints and conditional Gaussian potentials.

6
Gaussian Hints

Systems of linear equations with Gaussian disturbances can be analysed using *assumption-based reasoning*: By assuming that a disturbance was underlying the observation, the consequences are logically derived from that assumption. The disturbances which cannot logically have generated the observation are ruled out, and the distribution is conditioned on the possible assumptions. The result is captured in a structure called *hint* (Kohlas and Monney, 1995). Assumption-based reasoning has been used for Gaussian linear systems in (Monney, 2003; Kohlas and Monney, 2008). This theory gives clear semantics to statistical inference from such models, while still reproducing and generalising the results obtained by least-squares and maximum-likelihood estimation.

Chapter Outline

In Section 6.1, the *predictive* and *postdictive* approaches of statistical reasoning are introduced in order to set forth the statistical context of assumption-based reasoning, which is then discussed in Section 6.2. Assumption-based reasoning extracts the possible assumptions and derives their consequences. The result is captured in a structure called *hint*. Hypotheses can then be evaluated *qualitatively* by the supporting and plausible assumptions and *quantitatively* by the probability of these arguments conditioned on the possible assumptions. Different hints coming from independent sources can be combined using *Dempster's Rule*. In Section 6.3, Gaussian linear systems are formally introduced, and Gaussian hints are derived from them by assumption-based reasoning. Marginalisation and combination of Gaussian hints are motivated semantically and defined in Sections 6.4 and 6.5. In order to use the algorithms of Chapter 4 for the solution of the projection problem in the context of Gaussian hints, it is shown in Section 6.6 that Gaussian hints form a valuation algebra. Finally, it is shown in Section 6.7 that Gaussian hints extend the valuation algebra of Gaussian potentials

6.1 Statistical Reasoning

Predictive and Postdictive Probability Statements

A probability $P(E)$ of an event E may be considered in two different ways (Dempster, 1964):

- *predictively*, as a forward-looking measure of uncertainty about a future occurrence of E;
- *postdictively*, after the event E is observed, as a measure of likelihood or non-surprise or plausibility of that event.

> There are two fundamental and contrasting situations under which any probability such as $P(E)$ requires interpretation. If the trial or experiment determining whether or not E obtains has not yet occurred, or if ignorance of the outcome of this trial prevails, then $P(E)$ is to be regarded as a measure of the degree of certainty concerning the eventual establishment of the occurrence of E. On the other hand, if the outcome E of the trial is observed, then a quite different attitude towards $P(E)$ is natural, corresponding to the question: Is it plausible that an event E with probability $P(E)$ should have occurred? [...] I propose to use the terms *predictive* for the first situation and *postdictive* for the second.
> (Dempster, 1964; p.56)

(Dempster, 1964) further distinguishes between *pre-data* and *post-data* statements, i.e. statements made before or after an observation is made:

> For example, if a variable X is assigned a $N(0, 1)$ distribution, i.e., a normal distribution with mean zero and variance unity, then, in advance of observing X, the statement
>
> $$\Pr(X \leq 1.645) = .95 \tag{6.1}$$
>
> should be interpreted *predictively*, i.e., .95 measures a degree of certainty about the event $\Pr(X \leq 1.645)$ before the value of X is established. But after the value of X is observed to be, say, 1.805, the statement (6.1) can only be interpreted *postdictively*, i.e., it is known that an event previously judged to have probability .05 must have occurred. If the probability thus postdictively interpreted is tiny it conveys a feeling of surprise, and, consequently, diffidence or reluctance about accepting the validity of the original predictive probability statement (6.1). On the other hand, if an event E with moderate probability $P(E)$, say $P(E) = .30$, is contemplated before observation and subsequently is observed, then no feeling of surprise is natural and the postdictive interpretation is effectively neutral.
>
> In statistical inference the postdictive interpretation turns up most clearly in the rationale of a significance test. Indeed, the postdictive interpretation given in the above example, with observed $X = 1.805$, may be

> conveyed by stating that the observation is significantly large at the 5 per cent level when the null hypothesis is that X has the $N(0,1)$ distribution. On the other hand, the classical instance of a predictive interpretation in statistical inference is provided by a Bayes posterior probability. For example, if X is regarded as drawn from the $N(\mu,1)$ distribution, if μ is regarded as drawn from a uniform distribution with a very wide range, and if $X = 1.805$ is observed, then μ is assigned a posterior distribution very nearly $N(1.805,1)$. This distribution is to be used for making predictive probability assertions about μ while μ remains unknown. (Dempster, 1964; p.56f.)

Pre-data Predictive Statistical Approaches

Pre-data predictive statistical methods are *parameter-based*, starting from a statistical specification or distribution model

$$(Z, \mathcal{A}, \{P_\theta : \theta \in \Theta\}) \tag{6.2}$$

with outcomes $z \in Z$ and a fixed but unknown distribution P_θ from the parametric family $\{P_\theta : \theta \in \Theta\}$ of distributions on the σ-algebra $\mathcal{A}$ of subsets of Z. An estimator $\hat{\theta}$ of θ has then to be based on some principle that decides which distribution is a better explanation of the observation. In this approach, the parameters θ are hypothesised and some mathematical procedure is used to find an estimator $\hat{\theta}$ under which the observation is least surprising or most likely before data is available (before an observation is made). However, after the observation, only postdictive judgements are possible in the setting of a statistical specification: The observation may have been more or less surprising, resp. less or more likely under different parameters. The surprising observations are then often called *significant*. The same is true for *confidence regions*, which guarantee some overall behaviour of a statistical procedure over the whole observation space Z. Again, no positive confidence in a certain outcome or prediction of a certain outcome is possible, and nothing can be said about the possible error of the estimate after an observation z. Therefore, (Dempster, 1964) suggested using the term *indiffidence region* instead. Pre-data predictive methods provide operational statements about a procedure rather than inference on the parameter. This has been called the position of a *seller* of a statistical procedure (Kohlas and Monney, 2008).

Two classical methods with Gaussian linear models are pre-data predictive: The principle of *maximum likelihood* suggests choosing the parameter under which the observation has the highest likelihood. Another standard approach of linear regression analysis is to find the straight line which fits best a given data set by using the principle of *least squares*. It is then argued that the obtained *estimator* is *unbiased*, and the *Gauss-Markov* theorem asserts that it has the least variance among all unbiased linear estimators.

Post-data Predictive Approaches

In contrast, in a post-data predictive approach, a probability distribution is used to make predictive judgements about the uncertainty of some uncertain outcome or value to be observed.

The classical post-data predictive approach is the Bayesian method: Given a *prior* distribution on the parameter, a *posterior* distribution for the parameter given an observation can be derived by using Bayes' Theorem. The posterior distribution given the observation then allows truly post-data predictive probability statements. However, if nothing is known about θ, then this ignorance cannot be modelled appropriately in this framework and the requirement of a prior becomes apodictic.

In contrast, assumption-based statistical inference (Kohlas and Monney, 1995; Monney, 2003; Kohlas and Monney, 2008) provides an alternative post-data predictive approach. Starting from a functional model and an observation, predictive probabilities can be inferred by *assumption-based reasoning* (Kohlas and Monney, 1995; Monney, 2003; Kohlas and Monney, 2008). It is a framework including and generalising Fisher's fiducial methods, Fraser's structural approach (Fraser, 1968) and Bayesian statistics. On the other hand, the theory gives clear semantics to the likelihood function.

6.2 Assumption-Based Reasoning

Instead of a statistical specification, the starting point of assumption-based reasoning is a functional model (Monney, 2003)[1]

$$(f, \mathcal{A}, P) \tag{6.3}$$

where $f : \Theta \times \Omega \to Z$ and where P is a probability measure on the σ-algebra $\mathcal{A}$ of subsets of Ω. The function f models how an observation is generated given that the parameter θ and a disturbance ω are known. The post-data inference then starts with an observation $z \in Z$: Hypothesising an $\omega \in \Omega$ leads to the set

$$\Gamma_z(\omega) = \{\theta \in \Theta : \ f(\theta, \omega) = z\}, \tag{6.4}$$

the set of parameters compatible with the observation z and the *assumption* ω. The other parameters cannot possibly have generated the observation under no assumption. The set $\Gamma_z(\omega)$ is the smallest set which, under the assumption ω, contains the true parameter with certainty. The assumptions ω for which

$$\Gamma_z(\omega) = \{\theta \in \Theta : \ f(\theta, \omega) = z\} = \emptyset$$

lead to a contradiction since there is no corresponding parameter that could have generated the observation z; so they have to be ruled out as *inadmissible* or *impossible*. Therefore, define

$$v_z = \{\omega \in \Omega : \ \Gamma_z(\omega) \neq \emptyset\}, \tag{6.5}$$

[1]The term had been used in a less general way in the literature before as remarked by (Monney, 2003; p.23).

the set of *admissible assumptions*. This part is the *qualitative* result of the assumption-based inference. *Quantitatively*, the assumption-based inference leads to a conditional probability measure P' on the σ-algebra

$$\mathcal{A}' = \{A \cap v_z : A \in \mathcal{A}\}$$

in the following four cases.

1. If v_z is measurable with respect to $\mathcal{A}$ and if $P(v_z) > 0$, then define

$$P'(A') = \frac{P(A')}{P(v_z)} \tag{6.6}$$

for $A' \in \mathcal{A}' \subseteq \mathcal{A}$. It is readily verified that P' is a probability measure.

2. If v_z is not measurable with respect to $\mathcal{A}$ and $P^*(v_z) > 0$, then define

$$P'(A') = \frac{P^*(A')}{P^*(v_z)} \tag{6.7}$$

for $A' \in \mathcal{A}'$ where P^* is the *outer measure*

$$P^*(A') = \inf\{P(K) : K \supseteq A', K \in \mathcal{A}\}.$$

It has been noted by (Neveu, 1964; p.19) that P' is a probability measure in this second case.

3. Assume that v_z is a null set, i.e. that there is a set $N \in \mathcal{A}$ such that $P(N) = 0$ and $v_z \subseteq N$. Then, there cannot logically be a conditional probability measure given v_z since every subset $B \subseteq v_z$ has the same probability 0. However, the Radon-Nikodým theorem allows for an extension under limited circumstances, see e.g. (Rényi, 1970; Section 5.1). In particular, if $\mathcal{A} = \mathbb{B}^y$ is the Borel σ-algebra on the $|y|$-dimensional real space $\mathbb{R}^y$ and if P has a density f, i.e. if

$$P(B) = \int_B f(\mathbf{y})d\mathbf{y}, \quad B \in \mathbb{B}^y,$$

then an event "$z = \mathbf{z}$" induces the null set $v_z = \mathbb{R}^x \times \{\mathbf{z}\}$,

$$P(v_z) = 0,$$

for $z \subseteq y$, $\mathbf{z} \in \mathbb{R}^z$ and $x = y - z$. If $h(\mathbf{y}) > 0$ for all $\mathbf{y} \in \mathbb{R}^y$, then

$$g(\mathbf{x}) = \frac{h(\mathbf{x}, \mathbf{z})}{\int_{\mathbb{R}^x} h(\mathbf{x}, \mathbf{z})d\mathbf{x}} \tag{6.8}$$

may be interpreted as the conditional density function for x given "$z = \mathbf{z}$" since g defines a probability measure P' on $\mathcal{A}' = \mathbb{B}^x \times \{\mathbf{z}\}$ by

$$P'(A') = \int_{A'^{\downarrow x}} g(\mathbf{x})d\mathbf{x}, \quad A' \in \mathbb{B}^x.$$

The function g is a (normalised) probability density since $g(\mathbf{x}) > 0$ and

$$P'(v_z) = \int_{\mathbb{R}^x} g(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^x} \frac{h(\mathbf{x}, \mathbf{z})}{\int_{\mathbb{R}^x} h(\mathbf{x}, \mathbf{z})d\mathbf{x}} d\mathbf{x} = \frac{\int_{\mathbb{R}^x} h(\mathbf{x}, \mathbf{z})d\mathbf{x}}{\int_{\mathbb{R}^x} h(\mathbf{x}, \mathbf{z})d\mathbf{x}} = 1.$$

4. If P has a positive density f and $v_z = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ consists of a finite number of points, then v_z is null set. The probabilities

$$P'(\mathbf{x}_i) = c \cdot f(\mathbf{x}_i) \tag{6.9}$$

for $i \in 1, \ldots, n$ and

$$c = \sum_{i=1}^{n} f(\mathbf{x}_i)$$

define a probability measure on $\mathcal{A}' = 2^{v_z}$. The motivation for this extension is that $P'(\mathbf{x}_i) = \lim_{h \to 0} P'_h(\mathbf{x}_i)$ where

$$P'_h(\mathbf{x}_i) = c_h \cdot \int_{\mathbf{x}_i - h}^{\mathbf{x}_i + h} f(\mathbf{x}) d\mathbf{x}$$

for

$$c_h = P(\cup_{i=1}^{n} [\mathbf{x}_i - h, \mathbf{x}_i + h]).$$

In these four cases, the result of the assumption-based reasoning on a functional model instance is well defined and summarised in the structure

$$h = (v_z, P', \Gamma'_z, \Theta) \tag{6.10}$$

where $\Gamma'_z : v_z \to 2^{\Theta}$ is the restriction of Γ_z to v_z,

$$\Gamma'_z(\omega) = \Gamma_z(\omega), \qquad \omega \in v_z.$$

Such a structure is called a *hint* (Kohlas and Monney, 1995).

DEFINITION **6.1**. *A* hint *is a quadruple*

$$(\Omega, P, \Gamma, \Theta) \tag{6.11}$$

where

- *P is a probability measure on a σ-algebra of subsets of Ω,*
- *$\Gamma : \Omega \to 2^{\Theta}$ such that $\Gamma(\omega) \neq \emptyset$.*

The elements of the set Ω are called assumptions*, Θ is called the* frame of discernment*, and Γ is called the* focal mapping ⊘

EXAMPLE **6.2** (GAUSSIAN CHANNEL). The description follows (MacKay, 2003; Chapter 11). However, the example is analysed under the completely different perspective of assumption-based reasoning. Consider a *continuous-time channel* with input $x(t)$ and output $z(t) = x(t) + \omega(t)$ over a period $[0, T] \ni t$, with noise $\omega(t)$. In a signal of duration T, a set of n real numbers $\{x_i\}_{i=1}^{n}$ can be transmitted as a weighted combination of orthonormal basis functions $\phi_i(t)$,

$$x(t) = \sum_{i=1}^{n} x_i \phi_i(t), \tag{6.12}$$

where $\int_0^T \phi_i(t)\phi_j(t)\mathrm{d}t = \delta_{ij}$, $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$, $i, j \in \{1, \dots, n\}$. The receiver then gets the n numbers $\{z_i\}_{i=1}^n$ obtained by

$$z_i = \int_0^T \phi_i(t)z(t) = x_i + \int_0^T \phi_i(t)\omega(t)\mathrm{d}t = x_i + \omega_i, \qquad i \in \{1, \dots, n\},$$

with disturbances ω_i. If these disturbances $\omega = \big(\omega_1, \dots, \omega_n\big)' \in \mathbb{R}^n$ are distributed normally with Gaussian density $\phi_{0,K}(\omega)$, the channel is called Gaussian.

When the output $z(t)$ reaches the receiver, the input $x(t)$ should be recovered. However, the transmission error $\omega(t)$ is unknown. Therefore, such a pulse $x(t)$ may be used to encode only two instead of n values, represented by two vectors $\Theta = \{x_0, x_1\} \subset \mathbb{R}^n$. This defines a functional model $f : \Theta \times \mathbb{R}^n \to \mathbb{R}^n$, $f(x, \omega) = x + \omega = z$. After an observation z, the only two possible disturbances are

$$v_z = \{\omega_0 = z - x_0, \omega_1 = z - x_1\}. \tag{6.13}$$

Observe that $\omega_0 \neq \omega_1$ since $z - x_0 \neq z - x_1$. Conditioning on the event v_z yields the conditional probabilities

$$p'(\omega_j) = \frac{\phi_{0,K}(\omega_j)}{\phi_{0,K}(\omega_0) + \phi_{0,K}(\omega_1)}, \qquad j \in \{0, 1\} \tag{6.14}$$

defining the probability measure

$$P'(A) = \sum_{\omega \in A} p'(\omega), \qquad A \subseteq v_z.$$

Of course, v_z is a null set. The focal sets are

$$\Gamma_z(\omega_j) = \{x_j\}, \qquad j \in \{0, 1\}. \tag{6.15}$$

Hence, the received information is captured in the hint $h = (v_z, P', \Gamma_z, \Theta)$. ⊘

Evaluating Hypotheses

A hint $(\Omega, P, \Gamma, \Theta)$ can be used to evaluate a *hypothesis* $H \subseteq \Theta$ regarding the true parameter $\theta^* \in \Theta$. Every ω such that $\Gamma(\omega) \subseteq H$ is an argument in favour of the hypothesis H: The hypothesis is necessarily true under the *assumption* ω as it contains all logically possible parameters $\Gamma(\omega)$. The assumptions under which H is *necessarily* true are then grouped together in the set

$$u(H) = \{\omega \in \Omega : \ \Gamma(\omega) \subseteq H\}; \tag{6.16}$$

the set $u(H)$ represents the *qualitative* evaluation of the hypothesis H and its elements are said to support the hypothesis. If $u(H)$ is P-measurable (i.e. $u(H) \in \mathcal{A}$, $P : \mathcal{A} \to [0, 1]$), then the degree of support

$$sp(H) = P(u(H)) \tag{6.17}$$

gives the quantitative evaluation of H. The function sp can be extended in the following semantically motivated way: Since sp is supposed to measure the strength of the arguments which support H, it seems reasonable to assign the least upper bound of the probabilities $P(A)$ of all sets A containing only assumptions supporting H, i.e.

$$sp(H) = \sup_{A \in \mathcal{A}: A \subseteq u(H)} P(A) = P_*(u(H)) \tag{6.18}$$

where P_* is an *inner measure*. This function is indeed an extension since the supremum is attained by $u(H)$ if $u(H)$ is measurable. However, it has been noted that (6.18) is not the only possible extension of *support* or *belief functions* (Shafer, 1979; Kohlas and Monney, 1995).

On the other hand, an argument ω disproves a hypothesis H if the hypothesis cannot be true under that assumption. The assumptions disproving H are gathered in

$$w(H) = \{\omega \in \Omega : \ \Gamma(\omega) \cap H = \emptyset\}. \tag{6.19}$$

Since clearly

$$w(H) = \{\omega \in \Omega : \ \Gamma(\omega) \subseteq H^c\}, \tag{6.20}$$

$w(H)$ contains the assumptions supporting the complementary hypothesis $H^c = \Theta - H$. No assumptions support both H and its complement H^c, but some assumptions may well support neither H nor H^c. Therefore, for all $H \subseteq \Theta$,

$$u(H) \cap w(H) = \emptyset \quad \text{and} \quad u(H) \cup w(H) \subseteq \Omega \tag{6.21}$$

and

$$sp(H) + sp(H^c) \leq 1. \tag{6.22}$$

An assumption ω casts doubt on a hypothesis H if it supports its complement H^c. Conversely, if an assumption ω does not disprove the complementary hypothesis H^c, then H remains *possible* or *plausible* under ω. These assumptions form the set

$$v(H) = \{\omega \in \Omega : \ \Gamma(\omega) \not\subseteq H^c\}. \tag{6.23}$$

Clearly,

$$u(H) \subseteq v(H) \tag{6.24}$$

and

$$v(H) = \{\omega \in \Omega : \ \Gamma(\omega) \cap H \neq \emptyset\}. \tag{6.25}$$

Their strength is measured by the *degree of plausibility*

$$pl(H) = 1 - sp(H^c). \tag{6.26}$$

Since an assumption may neither disprove H^c nor $(H^c)^c = H$, it follows that

$$pl(H) + pl(H^c) \geq 1. \tag{6.27}$$

The functions $sp : 2^\Theta \to [0,1]$ and $pl : 2^\Theta \to [0,1]$ are called support and plausibility function, respectively. They have the following further properties.

Theorem **6.3**. (1) *$pl(H) = \inf_{A \in \mathcal{A}: A \supseteq v(H)} P(A) = P^*(v(H))$.*

(2) $sp(\emptyset) = pl(\emptyset) = 0$, $sp(\Theta) = pl(\Theta) = 1$.

(3) $sp(H) = 1 - pl(H^c)$, $pl(H) = 1 - sp(H^c)$.

(4) $sp(H) \leq pl(H)$.

(5) If $H_1 \subseteq H_2$, then $sp(H_1) \leq sp(H_2)$ and $pl(H_1) \leq pl(H_2)$. ⊘

Proof. (1) See for instance (Halmos, 1950). Assertions (2)–(5) are proved in (Kohlas and Monney, 1995; Theorem 3.2) for Θ finite. The general proof is essentially the same without this assumption. □

Degrees of plausibility are clearly post-data postdictive statements about the parameter. Moreover, in the discrete case, a functional model always induces a statistical specification by

$$p_\theta(\mathbf{z}) = P(\{\omega : f(\theta, \omega) = \mathbf{z}\}) \tag{6.28}$$

if Θ is discrete. However, different functional models may induce the same statistical specification as observed by (Monney, 2003; Kohlas and Monney, 2008). Therefore, *a functional model in general contains more information than a statistical specification alone.* The assumption-based approach also gives a clear semantics to the likelihoods

$$l_{\mathbf{z}}(\theta) = p_\theta(\mathbf{z}) = c \cdot P'(\{\omega : f(\theta, \omega) = \mathbf{z}\}) = c \cdot pl(\{\theta\}) \tag{6.29}$$

for a positive constant c not depending on θ accounting for the conditioning to the admissible assumptions: The likelihoods are proportional to the degrees of plausibility of the corresponding singleton hypothesis. In the continuous case, if the focal sets are all disjoint and cover Θ, if h is a continuous density on Ω and if f is a continuous function, then the functional model induces the statistical specification given by

$$h_\theta(\mathbf{z}) = h(f^{-1}(\theta, \mathbf{z})) \tag{6.30}$$

where $\omega = f^{-1}(\theta, \mathbf{z})$ is uniquely determined by $f(\theta, \omega) = \mathbf{z}$. Let $\bar{h}$ be the conditional density on v_z, and let the plausibility density be defined by

$$\bar{h}_\theta(\mathbf{z}) = \bar{h}(f^{-1}(\theta, \mathbf{z})). \tag{6.31}$$

Then, the likelihood function

$$l_{\mathbf{z}}(\theta) = h_\theta(\mathbf{z}) \tag{6.32}$$

is proportional to the plausibility density, i.e.

$$l_{\mathbf{z}}(\theta) = c \cdot \bar{h}(f^{-1}(\theta, \mathbf{z})) \tag{6.33}$$

for a positive constant c.

EXAMPLE **6.4** (DECODING MESSAGES OVER A GAUSSIAN CHANNEL). The redundancy in a pulse over the Gaussian channel of Example 6.2 can be used for reliable decoding. Here, the degrees of support are

$$sp_z(\{x_j\}) = p'(\omega_j), \qquad j \in \{0,1\}, \tag{6.34}$$

and

$$sp_z(\{x_0, x_1\}) = p'(\omega_0) + p'(\omega_1) = 1, \tag{6.35}$$

i.e. the input has been x_0 or x_1 with certainty. Intuitively, one should take x_0 if $sp_z(\{x_0\}) > sp_z(\{x_1\})$ and x_1 if $sp_z(\{x_1\}) > sp_z(\{x_0\})$. If $sp_z(\{x_0\}) = sp_z(\{x_1\})$, no decision can be taken. It can be verified that this induces the following decision rule:

$$\begin{aligned} a(z) > 0 &\rightarrow x_0, \\ a(z) < 0 &\rightarrow x_1, \\ a(z) = 0 &\rightarrow \text{no decision} \end{aligned}$$

where $a(z) = z'K(x_0 - x_1)$.

What is the reliability of this decoding scheme? Assume that one has decided for x_0. Then,

$$1 - \alpha = sp_z(\{x_0\}) > sp_z(\{x_1\}) = \alpha. \tag{6.36}$$

The decision was right if ω_0 was the correct error and wrong if ω_1 was the correct error. This leads to the hint $\mathcal{E} = (v_z, P', \Gamma_{\mathcal{E}}, \Delta)$ on the decisions $\Delta = \{right, wrong\}$ with $\Gamma_{\mathcal{E}}(\omega_0) = \{right\}$ and $\Gamma_{\mathcal{E}}(\omega_1) = \{wrong\}$. The support of the decision being right is

$$sp_{\mathcal{E}}(\{right\}) = p'(\omega_0) = 1 - \alpha.$$

However, the decision being wrong remains plausible with degree

$$pl_{\mathcal{E}}(\{wrong\}) = p'(\omega_1) = \alpha.$$

This shows that the probability of rejecting x_0 wrongly and accepting x_0 wrongly are both bounded by α. The same argument can be applied in the case of a decision for x_1. Assumption-based decision rules have been studied in more generality in (Kohlas and Monney, 2008). ⊘

Precise Hints

DEFINITION **6.5**. *A hint is called* precise *if all its focal sets are singletons.* ⊘

Assumptions in a precise hint lead to mutually contradictory most precise answers (neglecting that two assumptions may lead to the same singleton focal set).

THEOREM **6.6**. *Let sp and pl denote the support and plausibility functions of a precise hint. Then, $sp(H) = pl(H)$ for all $H \subseteq \Theta$ and sp is a probability measure on 2^{Θ}.* ⊘

Proof. The general proof is essentially the same as that of Theorem 3.8 of (Kohlas and Monney, 1995) for finite Θ. □

Although the mapping $\Gamma : \Omega \to 2^{\Theta}$ of a precise hint is formally a random variable, its interpretation is different: The set Ω contains all possible assumptions and the function Γ derives their consequences.

Equivalence of Hints

Definition **6.7**. *Two hints h_1, h_2 on the same domain are called* equivalent, *$h_1 \cong h_2$, if and only if they induce the same plausibility and support functions.* ⊘

It has to be remarked that the support function unambiguously determines the plausibility function and vice-versa. Equivalent hints are obtained by renaming and regrouping assumptions with the same focal set.

Combining Hints

Since information may come in junks, several pieces of information have to be aggregated or combined. Combining two hints $h_1 = (\Omega_1, P_1, \Gamma_1, \Theta)$ and $h_2 = (\Omega_2, P_2, \Gamma_2, \Theta)$ yields a new hint

$$h_1 \otimes h_2 = (v, P', \Gamma, \Theta),$$

which is obtained by *Dempster's Rule* or *product-intersection* rule (Dempster, 1967): Define the mapping $\Gamma : v \to 2^{\Theta}$ by

$$\Gamma(\omega_1, \omega_2) = \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2)$$

where

$$v = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \; \Gamma_1(\omega_1) \cap \Gamma_2(\omega_2) \neq \emptyset\};$$

further let

$$P(A_1 \times A_2) = P_1(A_1) \cdot P_2(A_2), \qquad A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$$

and define P' to be the conditional distribution of P given v in the same way as in Section 6.2. The qualitative interpretation of this rule is the following: In order to derive a θ from both ω_1 and ω_2, that θ has to be in the intersection $\Gamma_1(\omega_1) \cap \Gamma(\omega_2)$. If the intersection is empty, no θ can be derived from both ω_1 and ω_2. Hence, h corresponds to the joint functional model of any two functional models inducing h_1 and h_2, respectively. Furthermore, the two random variables Γ_1 and Γ_2 have independent distribution, and the joint probability measure P has to be conditioned on the set v of assumptions admissible in light of the joint functional model. See also (Kohlas and Monney, 2008).

EXAMPLE **6.8** (GAUSSIAN CHANNEL WITH A PRIOR). Continuing Example 6.4, assume now that there is additional information on the input in the form of probabilities $q_0(x_0)$, $q_0(x_1)$. This piece of information is captured in the precise hint

$$h_0 = (\Theta, Q_0, \Gamma_0, \Theta)$$

where

$$\Gamma_0(x_i) = \{x_i\}, \qquad i \in \{0, 1\},$$

and Q_0 is the probability measure

$$Q_0(A) = \sum_{x_i \in A} q_0(x_i), \qquad A \subseteq \Theta.$$

How does this additional information influence the decision-making? The new piece of information has to combined to the updated hint $h' = h \otimes h_0 = (v', Q, \Gamma_z, \Theta)$ where $v' = \{(\omega_1, x_1), (\omega_2, x_2)\}$ and

$$Q(A) = \sum_{(x_i, \omega_i) \in A} p'(\omega_i) \cdot q_0(x_i).$$

It can be verified that this induces the following updated decision rule:

$$a'(z) > 0 \to x_0,$$
$$a'(z) < 0 \to x_1,$$
$$a'(z) = 0 \to \text{no decision}$$

where $a'(z) = z'K(x_0 - x_1) + \ln \frac{q_0(x_0)}{q_0(x_1)}$. This rule corresponds to the previous rule in the case $q_0(x_0) = 0.5 = q_0(x_1)$. In fact, $p'(\omega_i)$ can be interpreted as the conditional probability of z given x_i and q_0 is then a prior on the x_i. This reproduces the results of the Bayesian approach of (MacKay, 2003). ⊘

6.3 Assumption-Based Reasoning on Gaussian Linear Systems

Algorithms for the assumption-based inference on Gaussian linear systems will now be derived. This leads to Gaussian hints. Again, domains are supposed to be in a lattice $D \subseteq 2^r$ of finite subsets of a set r of variables.

Gaussian Linear Systems and Gaussian Hints

Gaussian linear systems and Gaussian hints will now be defined formally.

DEFINITION **6.9**. *Let $x \in D$ and let $m \in \mathbb{N}$ be a non-negative integer. A Gaussian linear system (GLS) on $x \in D$ is a triplet*

$$g = (A, z, K)$$

where

- *$A \in \mathbb{R}(m, x)$ of rank $r \leq m, |x|$,*
- *$z \in \mathbb{R}^m$,*
- *and $K \in \mathbb{R}(m, m)$ symmetric and positive definite.*

The matrix A is called the design matrix*, the vector z the* observation vector*, and K the* concentration matrix*. Let the set of all Gaussian linear system be denoted $\mathfrak{L}$. The domain $d(g) = x$ is called the* label *of g. This defines an operation $d : \mathfrak{L} \to D$.* ⊘

A Gaussian linear system is an instance of a *functional model* (Monney, 2003) $(f, \mathcal{A}, P)$ for statistical inference about x (i.e. $\Theta = \mathbb{R}^x$): When an experiment is performed, the outcome depends only on the value $\mathbf{x}$ of the parameter x and the random "disturbance" ω; the outcome is then given by $f : \mathbb{R}^x \times \mathbb{R}^m \to \mathbb{R}^m$,

$$f(\mathbf{x}, \omega) = A\mathbf{x} + \omega.$$

Further, $\mathcal{A} = \mathbb{B}^x \subseteq 2^{\mathbb{R}^x}$ and $P = \Phi_{0,K}$ (the probability measure with Gaussian density $\phi_{0,K}$). Given the result of such an experiment, some ω cannot have generated the observation and become *impossible* or *inadmissible* in light of the functional model, i.e. the set

$$\Gamma_z(\omega) = \{\mathbf{x} : \; A\mathbf{x} + \omega = z\}$$

may be empty for some ω. Although it is not known which *assumption* ω was underlying the experiment, one may ask what the consequences $\Gamma_z(\omega)$ of the assumption ω are. Since the experiment is described by the functional model, the assumptions ω which lead to the contradiction $\Gamma_z(\omega) = \emptyset$ have to be ruled out and the distribution has to be conditioned on the admissible assumptions

$$v_z = \{\omega : \exists \mathbf{x} : \omega = z - A\mathbf{x}\}.$$

How do these focal sets look like? Let $r = \mathrm{r}(A)$ be the rank of A. Then, for an admissible assumption $\omega \in v_z$, the *focal set* $\Gamma_z(\omega)$ is a linear manifold of dimension $|x| - r$. In other words, the distribution conditioned on v_z is a distribution over focal sets which are parallel linear manifolds in $\mathbb{R}^x$. On the one extreme, if $r = 0$, then the unique linear manifold of the partition is $\mathbb{R}^x$ itself; on the other extreme, if $r = |x|$, then the focal sets are all the singletons made up of one point of $\mathbb{R}^x$, i.e. the Gaussian linear system represents a Gaussian distribution over the points of $\mathbb{R}^x$.

How can the admissible disturbances be described? In a Gaussian linear system, all assumptions are admissible if and only if the design matrix A has full row rank since then and only then, for each ω, there is a linear combination of the columns of A such that $z - A\mathbf{x} = \omega$. The set of admissible disturbances is given by

$$v_z = \mathcal{C}(A) + z = \{\omega : \; \omega = A\mathbf{x} + z, \; \mathbf{x} \in \mathbb{R}^x\},$$

the r-dimensional column space $\mathcal{C}(A)$ of the design matrix A which is translated by z.

Example **6.10**. An example is shown in Figure 6.1: The event v_z is a straight line in the (ω_1, ω_2)-space for a design matrix A of rank $\mathrm{r}(A) = 1$. Here, the event v_z has probability $P(v_z) = 0$. So in this case the conditional probability is assured by the Radon-Nikodým Theorem (see above p. 107). ⊘

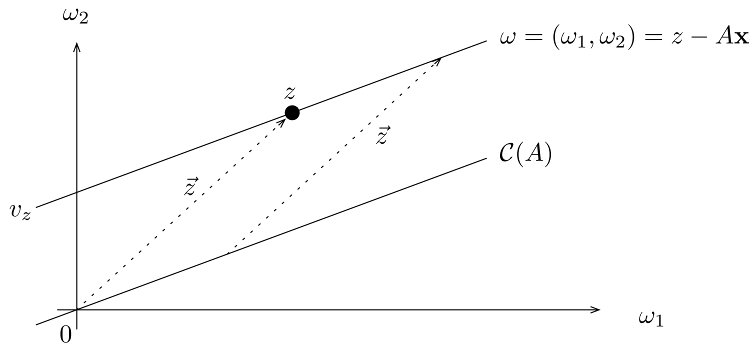


FIGURE 6.1: The admissible disturbances are given by a straight line v_z in the space of the disturbances $\omega = (\omega_1, \omega_2)$.

In order to compute the distribution conditioned on v_z, a change of coordinates is suitable: The Gaussian linear system has to be transformed such that the first rows have full rank and the remaining rows are 0. In these new coordinates, the admissible disturbances are given by any value in the first components and a constant value in the remaining components. An example in the two-dimensional space is shown in Figure 6.2: In the new coordinates (ξ_1, ξ_2), the event v_z is given by the equation $\xi_2 = c$. Now, Lemma B.1 can be applied to compute the conditional Gaussian

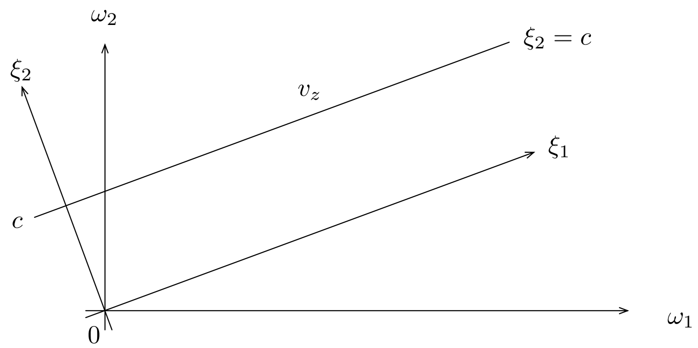


FIGURE 6.2: In the new coordinates (ξ_1, ξ_2), the admissible disturbances v_z are given by $\xi_2 = c$.

distribution of the admissible disturbances.
More generally, let $B_1 \in \mathbb{R}(m, r)$ be a basis of the column space of A. Then, there is a matrix $\Lambda \in \mathbb{R}(r, x)$ such that

$$A = B_1 \Lambda.$$

Further, let

$$B = (B_1, B_2)$$

be a regular matrix of $\mathbb{R}(m, m)$.

Definition **6.11**. *Let $A \in \mathbb{R}(m, x)$ and let $B = (B_1, B_2) \in \mathbb{R}(m, m)$ be a regular matrix such that $\mathcal{C}(B_1) = \mathcal{C}(A)$. Then, B is called a* permissible basis *for A.*[2] ⊘

Then, the Gaussian linear system can be transformed by the regular matrix $T = B^{-1}$. Let T be partitioned into $T_1 \in \mathbb{R}(r, m)$ and $T_2 \in \mathbb{R}(m - r, m)$ such that

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}.$$

Then,

$$TA = \begin{pmatrix} T_1 A \\ T_2 A \end{pmatrix} = \begin{pmatrix} T_1 A \\ T_2 B_1 A \end{pmatrix} = \begin{pmatrix} T_1 A \\ 0_{m-r,x} \end{pmatrix}. \tag{6.37}$$

Here, $T_1 A$ has full row rank r since $T_2 A = 0_{m-r,x}$ and T being regular imply that

$$\mathrm{r}(T_1 A) = \mathrm{r}(TA) = \mathrm{r}(A)$$

in light of Corollary 8.3.3 of (Harville, 1997; p.83). Therefore,

$$\Gamma_z(\omega) = \{\mathbf{x} : T_1 A\mathbf{x} + T_1\omega = T_1 z,\ T_2\omega = T_2 z\}. \tag{6.38}$$

Here, according to Appendix B.2, the transformed disturbances $T\omega$ have the density $\phi_{0,T^{-1\prime}KT^{-1}}$. It holds that

$$T^{-1\prime} K T^{-1} = \begin{pmatrix} B_1' K B_1 & B_1' K B_2 \\ B_2' K B_1 & B_2' K B_2 \end{pmatrix}.$$

In these transformed coordinates, it is easy to capture the admissible disturbances, namely

$$v_z = \{\omega :\ T_2\omega = T_2 z\}. \tag{6.39}$$

Theorem **6.12**. *Let (A, z, K) be a Gaussian linear system, $A \in \mathbb{R}(m, x)$ of rank $k = \mathrm{r}(A)$, $z \in \mathbb{R}(m)$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Let $B \in \mathbb{R}(m, m)$ be a permissible basis and define $T = B^{-1}$. Partition*

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

$T_1 \in \mathbb{R}(k, m)$, $T_2 \in \mathbb{R}(m - k, m)$, and

$$B = (B_1, B_2),$$

$B_1 \in \mathbb{R}(m, k)$, $B_2 \in \mathbb{R}(m, m - k)$. The result of the assumption-based inference is given by the Gaussian linear system

$$(T_1 A, T_1 z + (B_1' K B_1)^{-1}(B_1' K B_2) T_2 z, B_1' K B_1), \tag{6.40}$$

[2]The matrix A is not required to have full column rank as in the definition of a permissible basis of (Monney, 2003; p.82).

or, in terms of the covariance matrix $\Sigma = K^{-1}$,

$$T\Sigma T' = \begin{pmatrix} T_1\Sigma T_1' & T_1\Sigma T_2' \\ T_2\Sigma T_1' & T_2\Sigma T_2' \end{pmatrix},$$

by

$$(T_1A, T_1z - T_1\Sigma T_2'(T_2\Sigma T_2')^{-1}T_2z, T_1\Sigma T_1' - T_1\Sigma T_2'(T_2\Sigma T_2')^{-1}T_2\Sigma T_1'). \qquad (6.41)$$

⊘

PROOF. In light of Lemma B.1, the conditional distribution of $T_1\omega$ given $T_2\omega = T_2z$ has concentration

$$B_1'KB_1 = (T_1\Sigma T_1' - T_1\Sigma T_2'(T_2\Sigma T_2')^{-1}T_2\Sigma T_1')^{-1}$$

and mean

$$-(B_1'KB_1)^{-1}(B_1'KB_2)T_2z = T_1\Sigma T_2'(T_2\Sigma T_2')^{-1}T_2z. \qquad \square$$

The Gaussian linear system obtained by assumption-based reasoning is called a *Gaussian hint*.

DEFINITION **6.13**. *A* **Gaussian hint** *is a Gaussian linear system* (A, z, K) *on* $x \in D$ *where* $A \in \mathbb{R}(m, x)$ *has full row rank* $m = \mathrm{r}(A)$. *The set of all Gaussian hints shall be denoted by* $\mathcal{H}$, *where* $\mathcal{H} \subseteq \mathfrak{L}$. ⊘

Formally, the triplet (A, z, K) is of course not a hint as defined in equation (6.11): More precisely, it is taken as an abbreviation for the hint

$$(\mathbb{R}^m, \Phi_{0,K}, \Gamma, \mathbb{R}^x) \qquad (6.42)$$

where

- $\Phi_{0,K}$ is the Gaussian distribution with concentration matrix K,
- $\Gamma : \mathbb{R}^m \to 2^{\mathbb{R}^x}$, $\Gamma(\omega) = \{\mathbf{x} \in \mathbb{R}^x : A\mathbf{x} + \omega = z\}$.

EXAMPLE **6.14** (MEASUREMENT MODEL). A real-valued variable X with unknown value $\mathbf{x}^* \in \mathbb{R}$ is measured m times with results z_i ($i \in \{1, \ldots, m\}$) and the errors ω_i are assumed to be independent and identically distributed according to $\mathcal{N}(0, \sigma^2)$. This situation induces the following functional linear model on $x = \{X\}$:

$$\mathbf{x} + \omega_i = z_i, \qquad i = 1, \ldots, m.$$

This functional model is captured in the Gaussian linear system (A, z, K),

$$A = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix}, \quad K = \begin{pmatrix} \frac{1}{\sigma^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma^2} \end{pmatrix}.$$

In order to get directly directly $\mathbf{x}$ in the first row, a transformation matrix is chosen such that $T_1 A = \begin{pmatrix}1 & 0 & \cdots & 0\end{pmatrix}'$, $T_2 A = 0$ and such that $T_1\omega$ and $T_2\omega$ are independent, for instance

$$T = \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ 1 & -1 & & \\ \vdots & & \ddots & \\ 1 & & & -1 \end{pmatrix}.$$

Applying T yields

$$\begin{cases} \mathbf{x} + \xi_1 = & \frac{1}{m}\sum_{i=1}^m z_i, \\ \xi_i = & z_1 - z_i, \quad i \in \{2,\ldots,m\} \end{cases}$$

where the transformed disturbances ξ are distributed normally with mean 0 and variance-covariance

$$TK^{-1}T' = \begin{pmatrix} \frac{\sigma^2}{m} & 0 & \cdots & 0 \\ 0 & 2\sigma^2 & \sigma^2 & \cdots \\ \vdots & \sigma^2 & \ddots & \ddots \\ 0 & \vdots & \ddots & 2\sigma^2. \end{pmatrix}$$

Here, T corresponds to the permissible basis

$$T^{-1} = \begin{pmatrix} 1 & \frac{1}{m} & \cdots & \cdots & \frac{1}{m} \\ 1 & -\frac{m-1}{m} & \ddots & \iddots & \vdots \\ \vdots & \frac{1}{m} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{1}{m} \\ 1 & \frac{1}{m} & \cdots & \frac{1}{m} & -\frac{m-1}{m} \end{pmatrix}.$$

Since the distribution of the ξ_1 is independent from ξ_i for $i \in \{2,\ldots,m\}$, the conditional distribution of ξ_1 given $\xi_i = z_1 - z_i$ for $i \in \{2,\ldots,m\}$ has mean 0 and variance $\frac{\sigma^2}{m}$. So the more observations there are, the more *support* one gets for x being close to the sample mean $\bar{z} = \frac{1}{m}\sum_{i=1}^m z_i$. From the equation $\xi_1 = \bar{z} - \theta$, the predictive *fiducial density* of θ given the observation vector $z \in \mathbb{R}^m$ is then

$$t(\theta; z) = \phi_{0,\frac{\sigma^2}{m}}(\bar{z} - \theta) = c \cdot e^{-\frac{1}{2}(\bar{z}-\theta)^2} = c'\phi_{0,\sigma^2}(z_1 - \theta)\cdots\phi_{0,\sigma^2}(z_m - \theta) \qquad (6.43)$$

for appropriate normalisation constants $c, c' > 0$. This shows that the fiducial density is proportional to the product of the fiducial densities from the individual measurements z_i, $i \in \{1,\ldots,m\}$. However, this fiducial density must be interpreted carefully: The measurement model does not take into account any outliers due to improper measurements. Assumption-based reasoning relies on the premise that all eventualities are modelled *explicitly*. Therefore, if improper measurements are possible, their impact must be stated in the model as well.

Although the above results are the same as obtained by least-squares estimation, the interpretation is radically different: The distribution of x reflects the strength of the supporting (and plausible) arguments implied by the model. No optimisation is involved in the reasoning process. ⊘

Equivalent Gaussian Hints and Equivalent Gaussian Linear Systems

Different Gaussians hints may have the same focal sets and the same distribution over these focal sets. Such *equivalent* Gaussian hints capture essentially the same information, with respect to a different basis of the assumption space. Since, as defined above, the hint inferred from a Gaussian linear system depends on the choice of a permissible basis, it has also to be verified that inference leads to equivalent Gaussian hints.

DEFINITION **6.15**. *Let* $h_1 = (A_1, z_1, K_1)$ *and* $h_2 = (A_2, z_2, K_2)$ *be two Gaussian hints on the same set* $x \in D$ *of variables, where* $A_1, A_2 \in \mathbb{R}(m,x)$, $z_1, z_2 \in \mathbb{R}^m$, $K_1, K_2 \in \mathbb{R}(m,m)$ *symmetric and positive definite. Then,* h_1 *and* h_2 *are* equivalent, *written*

$$h_1 \cong h_2,$$

if and only if there is a regular matrix $B \in \mathbb{R}(m,m)$ *such that*

$$(A_2, z_2, K_2) = (BA_1, Bz_1, B^{-1'}K_1B^{-1}).$$

The set of equivalent Gaussian hints is denoted $\mathcal{H}(h)$*, and the quotient set of all equivalence classes of Gaussian hints is* $\mathcal{H}/\cong \{\mathcal{H}(h) : h \in \mathcal{H}\}$. ⊘

The relation $\cong$ is an equivalence relation in $\mathcal{H}$ as shown by (Monney, 2003; Theorem 22, p.62). The following two lemmata show that equivalent Gaussian hints essentially represent the same information, although they do not represent the same hint in the sense of (6.42): They represent the same information since they have the same focal sets and the same distribution over these focal sets.

LEMMA **6.16**. *Let* $h_1 = (A_1, z_1, K_1)$ *and* $h_2 = (A_2, z_2, K_2)$ *be two equivalent Gaussian hints on the same domain* $x \in D$*, where* $A_1, A_2 \in \mathbb{R}(m,x)$, $z_1, z_2 \in \mathbb{R}^m$, $K_1, K_2 \in \mathbb{R}(m,m)$ *symmetric and positive definite. Define* $\Gamma_1, \Gamma_2 : \mathbb{R}^m \to 2^{\mathbb{R}^x}$ *by*

$$\Gamma_1(\omega) = \{\mathbf{x} \in \mathbb{R}^m : \ A_1\mathbf{x} + \omega = z_1\},$$
$$\Gamma_2(\omega) = \{\mathbf{x} \in \mathbb{R}^m : \ A_2\mathbf{x} + \omega = z_2\}.$$

Then, there is a regular matrix $B \in \mathbb{R}(m,m)$ *such that for* $\omega \in \mathbb{R}^m$

$$\Gamma_1(\omega) = \Gamma_2(B\omega) \tag{6.44}$$

and

$$\phi_{0,K_1}(\omega) = \phi_{0,K_2}(B\omega) \cdot |\det(B)|\,. \tag{6.45}$$

⊘

PROOF. Let $B \in \mathbb{R}(m,m)$ be the regular matrix that establishes the equivalence $h_1 \cong h_2$, i.e.

$$A_2 = BA_1, \quad z_2 = Bz_1, \quad K_2 = B^{-1'}K_1B^{-1}$$

Let $\omega \in \mathbb{R}^m$. Then, since B is regular,

$$\begin{aligned}\Gamma_1(\omega) &= \{\mathbf{x} \in \mathbb{R}^m : \; A_1\mathbf{x} + \omega = z_1\} \\ &= \{\mathbf{x} \in \mathbb{R}^m : \; B^{-1}A_2\mathbf{x} + \omega = B^{-1}z_2\} \\ &= \{\mathbf{x} \in \mathbb{R}^m : \; A_2\mathbf{x} + B\omega = z_2\} \\ &= \Gamma_2(B\omega).\end{aligned}$$

Using

$$\det(B'KB) = \det(B') \cdot \det(K) \cdot \det(B) = \det(B)^2 \cdot \det(K)$$

[in light of Theorem 13.3.4 of (Harville, 1997; p.187) and Lemma 13.2.1 of (Harville, 1997; p.181)], it also follows that

$$\begin{aligned}\phi_{0,K_1}(\omega) &= \sqrt{\frac{|\det(K_1)|}{(2\pi)^{|x|}}} \cdot e^{-\frac{1}{2}\omega' K_1 \omega} \\ &= \sqrt{\frac{|\det(B'K_2B)|}{(2\pi)^{|x|}}} \cdot e^{-\frac{1}{2}(B\omega)' B^{-1'} K_1 B^{-1}(B\omega)} \\ &= \sqrt{\frac{\det(B)^2 \cdot |\det(K_2)|}{(2\pi)^{|x|}}} \cdot e^{-\frac{1}{2}(B\omega)' K_2 (B\omega)} \\ &= \phi_{0,K_2}(B\omega) \cdot |\det(B)|.\end{aligned}$$

□

The previous lemma shows that equivalent Gaussian hints have the same focal sets and the same distribution over these focal sets. Moreover, the following theorem shows that Gaussian hints are equivalent if and only if they generate the same support and plausibility functions. This shows that the notion of equivalence of Gaussian hints is compatible with the more general notion of equivalence of hints defined in Section 6.2 and justifies the use of the same symbol $\cong$.

Theorem **6.17**. *Let $h_1 = (A_1, z_1, K_1)$ and $h_2 = (A_2, z_2, K_2)$ be Gaussian hints on the same domain. Then, they are equivalent if and only they generate the same support and plausibility functions.* ⊘

Proof. For the "only if" part, see Theorem 23 of (Monney, 2003; p.64).
In order to prove the "if" part, assume that the support functions sp_1 of h_1 and sp_2 of h_2 are equal, i.e. that $sp_1(H) = sp_2(H)$ for all $H \subseteq \mathbb{R}^x$. Let $\Gamma_1 : \mathbb{R}^{m_1} \to 2^{\mathbb{R}^x}$, $\Gamma_1(\omega) = \{\mathbf{x} : A_1\mathbf{x} + \omega = z_1\}$ and $\Gamma_2 : \mathbb{R}^{m_2} \to 2^{\mathbb{R}^x}$, $\Gamma_2(\xi) = \{\mathbf{x} : A_2\mathbf{x} + \xi = z_2\}$.
Let $U \in \mathbb{B}^{m_2}$ be the unit ball in $\mathbb{R}^{m_2}$ and let $H_U = \Gamma_2(U) = \bigcup_{\xi \in U} \Gamma_2(\xi)$. Since the focal sets $\Gamma_2(\xi)$ form a partition of $\mathbb{R}^x$ and since H_U is the union of such partitioning elements, it follows that $sp_2(H_U) = P_2(U) > 0$. Define $S(\omega) = \{\xi : \Gamma_2(\xi) \cap \Gamma_1(\omega) \neq \emptyset\}$.
It is now shown that $\dim(\Gamma_2(\xi)) = \dim(\Gamma_1(\omega))$. Conversely, assume $\dim(\Gamma_2(\xi)) > \dim(\Gamma_1(\omega))$. This situation is depicted in Figure 6.3(a). Assume $\Gamma_1(\omega) \subseteq H_U$. Then, $S(\omega) \subseteq U$ since H_U is the union of partitioning elements $\Gamma_2(\xi)$. Since $\dim(\Gamma_2(\xi)) = \dim(\Gamma_1(\omega))$ and since the focal sets are parallel of the same dimension, for a fixed ω, there are then different ξ_1, ξ_2 such that $\Gamma_2(\xi_1) \cap \Gamma_1(\omega) \neq \emptyset$ and $\Gamma_2(\xi_2) \cap \Gamma_1(\omega) \neq \emptyset$.

Define $S_0(\omega) = \{\xi = \xi_1 + s\cdot(\xi_1 - \xi_2) : \Gamma_2(\xi_1)\cap\Gamma_1(\omega) \neq \emptyset, \Gamma_2(\xi_2)\cap\Gamma_1(\omega) \neq \emptyset, s \in \mathbb{R}\}$. Then $S_0(\omega) \subseteq S(\omega)$ since for $\mathbf{x}_1 \in \Gamma(\xi_1)\cap\Gamma_1(\omega)$ and $\mathbf{x}_2 \in \Gamma(\xi_2)\cap\Gamma_1(\omega)$ it holds that $A_2(\mathbf{x}_1 + s\cdot(\mathbf{x}_1 - \mathbf{x}_2)) + (\xi_1 + s\cdot(\xi_1 - \xi_2)) = (s+1)\cdot(A_2\mathbf{x}_1 + \xi_1) - s\cdot(A_2\mathbf{x}_2 + \xi_2) = z_2$. Hence, $\emptyset \neq S_0(\omega) \subseteq U$. However, this is not possible since the vectors in $S_0(\omega)$ are unbounded whereas the vectors in U are bounded. A similar argument shows that $\dim(\Gamma_2(\xi)) < \dim(\Gamma_1(\omega))$ also leads to a contradiction. Hence, indeed $\dim(\Gamma_2(\xi)) = \dim(\Gamma_1(\omega))$.
It will now be shown that h_1 and h_2 have the same focal sets, i.e. that $\mathrm{im}(\Gamma_1) = \mathrm{im}(\Gamma_2)$. Conservely, assume $\mathrm{im}(\Gamma_1) \neq \mathrm{im}(\Gamma_2)$. This situation is depicted in Figure 6.3(b). Assume $\Gamma_1(\omega) \subseteq H_U$. Again, for fixed ω, there are then different ξ_1, ξ_2 such that $\Gamma_2(\xi_1) \cap \Gamma_1(\omega) \neq \emptyset$ and $\Gamma_2(\xi_2) \cap \Gamma_1 \neq \emptyset$. A similar argument as above shows that $\mathrm{im}(\Gamma_1) = \mathrm{im}(\Gamma_2)$.
Since $\mathrm{im}(\Gamma_1) = \mathrm{im}(\Gamma_2)$, Lemma A.2 (4) shows that there is a regular matrix T such that $A_2 = TA_1$ and $z_2 = Tz_1$. Therefore, $\Gamma_2(T\omega) = \Gamma_1(\omega)$ for all $\omega \in \mathbb{R}^{m_1} = \mathbb{R}^{m_2}$. Hence, for all $A \in \mathbb{B}^x$,

$$\begin{aligned}
\int_{\omega\in A} \phi_{0,K_1}(\omega)d\omega &= sp_1(\Gamma_1(A)) \\
&= sp_2(\Gamma_2(A)) \\
&= \int_{\xi\in TA} \phi_{0,K_2}(\xi)d\xi \\
&= \int_{\omega\in A} |\det(T)| \cdot \phi_{0,K_2}(T\omega)d\omega.
\end{aligned}$$

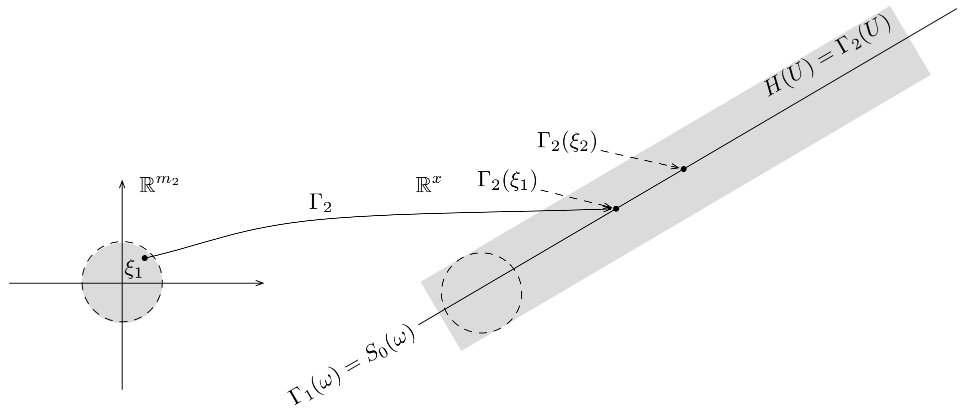
Since ϕ_{0,K_1} and $|\det(T)|\cdot\phi_{0,K_2}$ are both the derivative of P_1 with respect to ω, it follows that $\phi_{0,K_1}(\omega) = |\det(T)|\cdot\phi_{0,K_2}(T\omega)$ for all ω. Hence,

$$\begin{aligned}
\sqrt{\frac{|\det(K_1)|}{(2\pi)^{|x|}}}\cdot e^{-\frac{1}{2}\omega' K_1 \omega} &= \phi_{0,K_1}(\omega) \\
&= |\det(T)|\cdot\phi_{0,K_2}(T\omega) \\
&= \sqrt{\frac{|\det(K_2)|}{(2\pi)^{|x|}}}\cdot e^{-\frac{1}{2}(T\omega)' K_2 (T\omega)} \\
&= \sqrt{\frac{|\det(K_2)|}{(2\pi)^{|x|}}}\cdot e^{-\frac{1}{2}\omega'(T'K_2T)\omega}.
\end{aligned}$$

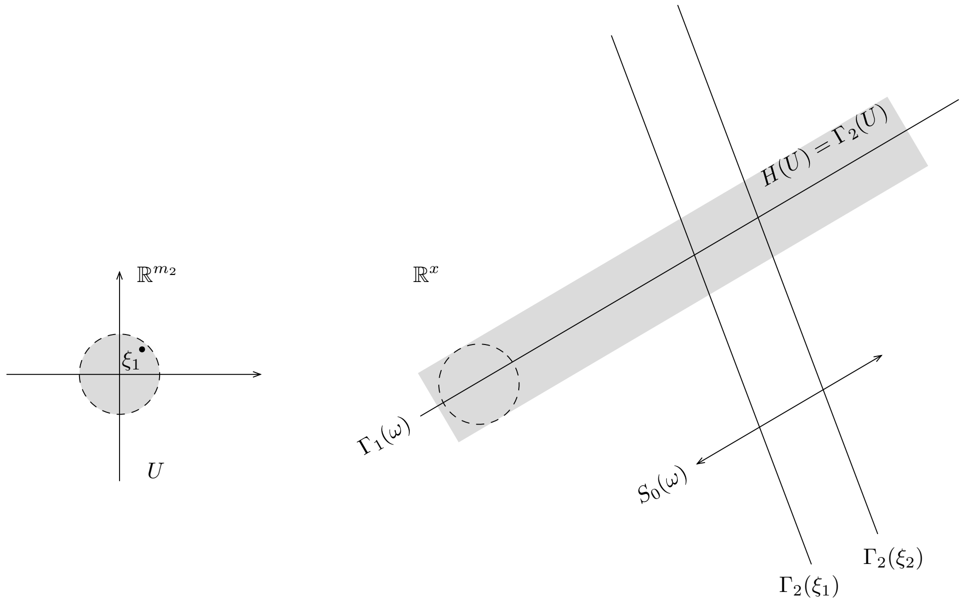
Since the first factor is constant (not depending on ω) and since e is strictly monotone, it follows that $\omega' K\omega = \omega'(T'K_2T)\omega$ for all ω. Hence, in light of Lemma A.4, $K_1 = T'K_2T$ and, equivalently, $K_2 = T^{-1'}K_1T^{-1}$. □

This shows that equivalent Gaussian hints capture essentially the same information with respect to different bases of the assumption space. Geometrically, equivalent Gaussian hints represent the same distribution over the same focal sets.

As developed so far, the assumption-based inference starts with a permissible basis B and then defines the matrix $T = B^{-1}$. Then, $T_2A = 0$. It is now shown



(a) Assumption $\Gamma_1(\omega) \subseteq H_U$ and $\dim(\Gamma_2(\xi)) > \dim(\Gamma_1(\omega))$



(b) Assumption $\mathrm{im}(\Gamma_1) \neq \mathrm{im}(\Gamma_2)$, $\Gamma_1(\omega) \subseteq H_U$ and $\dim(\Gamma_2(\xi)) = \dim(\Gamma_1(\omega))$

FIGURE 6.3: Gaussian hints are equivalent if they have the same support function.

that any matrix having the properties of T corresponds to a permissible basis and can thus be used for the assumption-based inference.

DEFINITION **6.18**. *Let $A \in \mathbb{R}(m,x)$ of rank $k = \mathrm{r}(A)$. Then, a regular matrix $T \in \mathbb{R}(m,m)$,*

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \quad T_1 \in \mathbb{R}(k,x), \quad T_2 \in \mathbb{R}(m-k,x),$$

is called **admissible** *if $T_2 A = 0_{m-k,x}$.*[3] ⊘

The following lemma shows that the inverse of an admissible matrix constitutes a permissible basis.

LEMMA **6.19**. *Let T be an admissible matrix for a Gaussian linear system (A, z, K), $A \in \mathbb{R}(m,x)$ of rank $k = \mathrm{r}(A)$, $z \in \mathbb{R}(m)$, $K \in \mathbb{R}(m,m)$ symmetric and positive definite. Partition*

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

$T_1 \in \mathbb{R}(k,x)$, $T_2 \in \mathbb{R}(m-k,x)$ such that $T_2 A = 0_{m-k,x}$. Let $B = T^{-1}$ and

$$B = (B_1, B_2),$$

$B_1 \in \mathbb{R}(m,k)$, $B_2 \in \mathbb{R}(m,m-k)$. Then, B is a permissible basis, i.e. B_1 is a basis of the column space $\mathcal{C}(A)$ of A. ⊘

PROOF. Since T is regular, the column space of A and its rank are preserved when premultiplied by T,

$$\mathcal{C}(TA) = \mathcal{C}(A), \qquad \mathrm{r}(TA) = \mathrm{r}(A) = k,$$

according to Lemma of 8.3.2 (Harville, 1997; p.83). Since

$$TA = \begin{pmatrix} T_1 A \\ 0_{m-k,x} \end{pmatrix},$$

the matrix

$$\Lambda_1 = \begin{pmatrix} I_k \\ 0_{m-k,k} \end{pmatrix}$$

is a basis of $\mathcal{C}(TA)$. Hence, $\mathcal{C}(TA) = \mathcal{C}(\Lambda_1)$. Again, since T^{-1} is regular,

$$\mathcal{C}(T^{-1}\Lambda_1) = \mathcal{C}(T^{-1}TA) = \mathcal{C}(A).$$

Since the columns of $B_1 = T^{-1}\Lambda_1$ are linearly independent, B_1 is a basis of $\mathcal{C}(A)$. Hence, B is indeed a permissible basis. □

[3] As in the case of a permissible basis, this definition of an admissible matrix is more general than that of (Monney, 2003; p.83).

The following theorem shows that the assumption-based inference with different permissible bases (or, equivalently, different admissible matrices) leads to equivalent Gaussian hints.

Theorem **6.20**. *Let $T, \tilde{T}$ be admissible matrices for a Gaussian linear system (A, z, K), $A \in \mathbb{R}(m, x)$ of rank $k = \mathrm{r}(A)$, $z \in \mathbb{R}(m)$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Then, assumption-based inference leads to equivalent Gaussian hints.* ⊘

Proof. Partition

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \quad \tilde{T} = \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix}$$

$T_1, \tilde{T}_1 \in \mathbb{R}(k, m)$, $T_2, \tilde{T}_2 \in \mathbb{R}(m-k, x)$ such that $T_2 A = \tilde{T}_2 A = 0_{m-k,x}$. Let $B = T^{-1}$ and $\tilde{B} = \tilde{T}^{-1}$. Partition

$$B = (B_1, B_2), \quad \tilde{B} = (\tilde{B}_1, \tilde{B}_2),$$

$B_1, \tilde{B}_1 \in \mathbb{R}(m, k)$, $B_2, \tilde{B}_2 \in \mathbb{R}(m, m-k)$. According to Lemma 6.19, B and $\tilde{B}$ are permissible bases, hence B_1 and $\tilde{B}_1$ are bases of the column space $\mathcal{C}(A)$ of A. Similarly, B_2 and $\tilde{B}_2$ are bases of the null space $\mathcal{N}(A)$ of A. Therefore, every column of $\tilde{B}_1$ is a unique linear combination of columns of B_1, hence there is a unique matrix $R_1 \in \mathbb{R}(k, k)$ such that

$$B_1 R_1 = \tilde{B}_1.$$

Similarly, there is a unique matrix $R_2 \in \mathbb{R}(m-k, m-k)$ such that

$$B_2 R_2 = \tilde{B}_2.$$

Observe that

$$k \geq \mathrm{r}(R_1) \geq \mathrm{r}(B_1 R_1) = \mathrm{r}(\tilde{B}_1) = k$$

in light of Corollary 4.4.5 of (Harville, 1997; p.37), hence $\mathrm{r}(R_1) = k$. Thus, R_1 is regular. It can be shown similarly that R_2 is regular as well. Furthermore,

$$\tilde{B} \;=\; B \begin{pmatrix} R_1 & 0_{k,m-k} \\ 0_{m-k,k} & R_2 \end{pmatrix} \;=\; (B_1 R_1, B_2 R_2)$$

and

$$\begin{aligned} \tilde{T} = \tilde{B}^{-1} \;&=\; \begin{pmatrix} {R_1}^{-1} & 0_{k,m-k} \\ 0_{m-k,k} & {R_2}^{-1} \end{pmatrix} B^{-1} \;=\; \begin{pmatrix} {R_1}^{-1} & 0_{k,m-k} \\ 0_{m-k,k} & {R_2}^{-1} \end{pmatrix} T \\ &= \begin{pmatrix} {R_1}^{-1} T_1 \\ {R_2}^{-1} T_2 \end{pmatrix}, \end{aligned}$$

using result (8.2.8) of (Harville, 1997; p.82) for the inverse of a product of regular matrices and result (8.5.2) of (Harville, 1997; p.88) for the inverse of a block-diagonal matrix. Hence,

$$\tilde{T}_1 = {R_1}^{-1} T_1, \quad \tilde{B}_1 = B_1 R_1. \tag{6.46}$$

Then, according to Theorem 6.12, the hint inferred using T is

$$h = (T_1 A, T_1 z - (B_1' K B_1)^{-1} B_1' K z, B_1' K B_1)$$

and the hint inferred using $\tilde{T}$ is

$$\tilde{h} = (\tilde{T}_1 A, \tilde{T}_1 z - (\tilde{B}_1' K \tilde{B}_1)^{-1} \tilde{B}_1' K z, \tilde{B}_1' K \tilde{B}_1).$$

Using (6.46),

$$\tilde{T}_1 A = {R_1}^{-1}(T_1 A),$$

$$\tilde{B}_1' K \tilde{B}_1 = (B_1 R_1)' K (B_1 R_1) = R_1'(B_1' K B_1) R_1,$$

and

$$\begin{aligned} \tilde{T}_1 z - (\tilde{B}_1' K \tilde{B}_1)^{-1} \tilde{B}_1' K z &= ({R_1}^{-1} T_1) z - {R_1}^{-1} (B_1' K B_1)^{-1} {R_1'}^{-1} R_1' B_1' K z \\ &= {R_1}^{-1}(T_1 z - (B_1' K B_1)^{-1} B_1' K z), \end{aligned}$$

using result (8.2.8) of (Harville, 1997; p.82) for the inverse of a product of regular matrices, result (8.2.3) of (Harville, 1997; p.82) for the commutativity of transposition and inversion of regular matrices and result (1.2.13) of (Harville, 1997; p.5) for the transpose of a product of matrices. Hence, ${R_1}^{-1}$ establishes the equivalence $h \cong \tilde{h}$. □

In light of Theorem 6.20, the hints inferred from a Gaussian linear system p by different permissible bases are equivalent. Since $h \in \mathcal{H}(h)$ for $h \in \mathcal{H}$, the following (extensions of existing) definitions are sound.

DEFINITION **6.21**. *For a Gaussian linear system $p \in \mathfrak{L}$, define $\mathcal{H}(p)$ to be set of equivalent Gaussian hints induced by the Gaussian linear system. This defines an* inference operator *$\mathcal{H} : \mathfrak{L} \to \mathcal{H}/\cong$. Furthermore, two Gaussian linear systems $p, p' \in \mathfrak{L}$ are said to be equivalent, $p \cong p'$ if and only if they induce equivalent Gaussian hints, i.e.*

$$p \cong p' \iff \mathcal{H}(p) = \mathcal{H}(p'). \tag{6.47}$$

⊘

This situation is shown in Figure 6.4: Gaussian hints are a subset of Gaussian linear systems, $\mathcal{H} \subseteq \mathfrak{L}$; the grey-shaded area shows the equivalence class of a Gaussian linear system p; the dark grey-shaded subarea $\mathcal{H}(p)$ is the subset of induced Gaussian hints; the Gaussian linear system p' is an equivalent Gaussian linear system which leads to the same Gaussian hint up to equivalence, i.e. $\mathcal{H}(p) = \mathcal{H}(p')$.

The following lemma gives a sufficient criterion for two Gaussian linear systems to be equivalent if they have the same number of equations.

LEMMA **6.22**. *Let $p = (A, z, K)$ and $\tilde{p} = (\tilde{A}, \tilde{z}, \tilde{K})$ be Gaussian linear systems of m equations over the same domain $x \in D$, $A, \tilde{A} \in \mathbb{R}(m, x)$, $z, \tilde{z} \in \mathbb{R}^m$, $K, \tilde{K} \in \mathbb{R}(m, m)$. Let $n = |x|$. If there is a regular matrix $P \in \mathbb{R}(m, m)$ such that*

$$(\tilde{A}, \tilde{z}, \tilde{K}) = (PA, Pz, {P^{-1}}' K P^{-1}),$$

then

$$p \cong \tilde{p}.$$

⊘

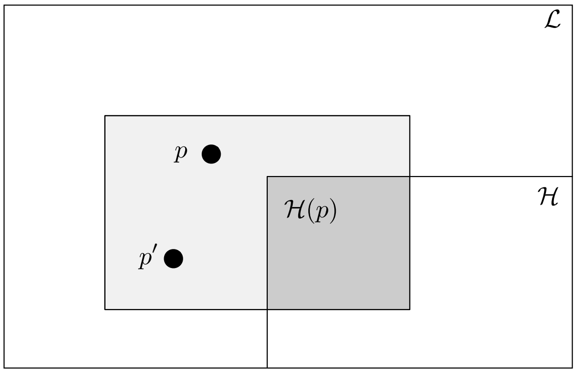


Figure 6.4: Gaussian linear systems are equivalent if the derived Gaussian hints are equivalent.

Proof. Let B be a permissible basis for p and define $T = B^{-1}$. Further, define

$$\tilde{B} = PB$$

and

$$\tilde{T} = B^{-1}P^{-1} = TP^{-1}.$$

Observe that $\tilde{A} = PA$ and P being regular imply that $\mathrm{r}(\tilde{A}) = \mathrm{r}(A)$ in light of Lemma 8.3.2 of (Harville, 1997; p.83). Define $k = \mathrm{r}(A)$. Partition

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \quad \tilde{T} = \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix}$$

where $T_1, \tilde{T}_1 \in \mathbb{R}(k, m)$ and $T_2, \tilde{T}_2 \in \mathbb{R}(m-k, m)$ and

$$B = (B_1, B_2), \quad \tilde{B} = (\tilde{B}_1, \tilde{B}_2)$$

where $B_1, \tilde{B}_1 \in \mathbb{R}(m, k)$ and $B_2, \tilde{B}_2 \in \mathbb{R}(m, m-k)$. Then,

$$\begin{pmatrix} \tilde{T}_1\tilde{A} \\ 0_{m-k,x} \end{pmatrix} = \begin{pmatrix} \tilde{T}_1\tilde{A} \\ \tilde{T}_2\tilde{A} \end{pmatrix} = \tilde{T}\tilde{A} = TP^{-1}PA = TA = \begin{pmatrix} T_1 A \\ T_2 A \end{pmatrix} = \begin{pmatrix} T_1 A \\ 0_{m-k,x} \end{pmatrix}$$

shows that $\tilde{T}$ is an admissible matrix for $\tilde{p}$. Hence, in light of Lemma 6.19, $\tilde{B}$ is a permissible basis for $\tilde{p}$. Then, in light of Theorem 6.12, the hint inferred from p is given by

$$h = (T_1 A, T_1 z + (B_1' K B_1)^{-1}(B_1' K B_2) T_2 z, B_1' K B_1) \in \mathcal{H}(p)$$

and the hint inferred from $\tilde{p}$ by

$$\tilde{h} = (\tilde{T}_1\tilde{A}, \tilde{T}_1\tilde{z} + (\tilde{B}_1'\tilde{K}\tilde{B}_1)^{-1}(\tilde{B}_1'\tilde{K}\tilde{B}_2)\tilde{T}_2\tilde{z}, \tilde{B}_1'\tilde{K}\tilde{B}_1) \in \mathcal{H}(\tilde{p}).$$

Since

$$\tilde{T}_1\tilde{A} = T_1P^{-1}PA = T_1A,$$
$$\tilde{B}_1'\tilde{K}\tilde{B}_1 = (PB_1)'(P'^{-1}KP^{-1})(PB_1) = B_1'P'P'^{-1}KP^{-1}PB_1 = B_1'KB_1,$$

and

$$\begin{aligned}&\tilde{T}_1\tilde{z} + (\tilde{B}_1'\tilde{K}\tilde{B}_1)^{-1}(\tilde{B}_1'\tilde{K}\tilde{B}_2)\tilde{T}_2\tilde{z}\\ =&T_1P^{-1}Pz + (B_1'KB_1)^{-1}(B_1'((PB_1)'P'^{-1}KP^{-1}(PB))T_2P^{-1}Pz\\ =&T_1z + (B_1'KB_1)^{-1}(B_1'KB_2)T_2z,\end{aligned}$$

it follows that $h \cong \tilde{h}$, i.e. $\mathcal{H}(p) = \mathcal{H}(\tilde{p})$. □

How an admissible matrix can be constructed for a Gaussian linear system is discussed in depth in (Monney, 2003). Another algorithm based on singular-value decomposition has been discussed and implemented in (Eichenberger, 2004) .

Inference if the Design Matrix has full Column Rank

If the design matrix of a Gaussian linear system has full column rank n, then the focal sets are linear manifolds of dimension 0. Hence, the result of the assumption-based inference is a precise hint representing a distribution over the points (more precisely the singleton subsets) of the parameter space. The result can then be obtained in a very simple way as shown by the following theorem.

THEOREM **6.23**. *Let $p = (A, z, K)$ be a Gaussian linear system, $A \in \mathbb{R}(m, x)$ of rank $\mathrm{r}(A) = |x| = n$, $z \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Then, the result of the assumption-based inference is given by the Gaussian hint*

$$(I_x, (A'KA)^{-1}A'Kz, A'KA) \in \mathcal{H}(p). \tag{6.48}$$

⊘

Equation (6.48) is short-hand for

$$(I_{n,x}, I_{n,x}(A'KA)^{-1}A'Kz, I_{n,x}A'KAI_{x,n}), \tag{6.49}$$

i.e. the rows are indexed by variables in (6.48) instead of numbers in (6.49).

PROOF. Notice that $m \geq n$. Since A has full column rank n, there is a permutation π of $\{1, \ldots, m\}$ such that the n rows $\pi(1), \ldots, \pi(n)$ are linearly independent. Define the regular permutation matrix $P \in \mathbb{R}(m, m)$ by $P(i, j) = \delta_{\pi(i),j}$ for $i, j \in \{1, \ldots, m\}$ (for $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$). Consider the transformed Gaussian linear system

$$\tilde{p} = (\tilde{A}, \tilde{z}, \tilde{K}) = (PA, Pz, P'^{-1}KP^{-1}), \tag{6.50}$$

which is equivalent to the original Gaussian linear system (A, z, K) in light of Lemma 6.22, i.e. $\mathcal{H}(p) = \mathcal{H}(\tilde{p})$. Partition

$$\tilde{A} = \begin{pmatrix}\tilde{A}_1\\ \tilde{A}_2\end{pmatrix}, \quad \tilde{z} = \begin{pmatrix}\tilde{z}_1\\ \tilde{z}_2\end{pmatrix},$$

$\tilde{A}_1 \in \mathbb{R}(n,x)$, $\tilde{A}_2 \in \mathbb{R}(m-n,x)$, $\tilde{z}_1 \in \mathbb{R}^n$, $\tilde{z}_2 \in \mathbb{R}^{m-n}$. By the definition of P, the first n rows of PA are linearly independent, i.e. $\tilde{A}_1$ is regular. Then, the matrix

$$B = \begin{pmatrix} \tilde{A}_1 & 0 \\ \tilde{A}_2 & I_{m-n} \end{pmatrix}$$

is a permissible basis and $T = B^{-1}$,

$$T = \begin{pmatrix} \tilde{A}_1^{-1} & 0 \\ -\tilde{A}_2\tilde{A}_1^{-1} & I_{m-n} \end{pmatrix},$$

is an admissible matrix. According to Theorem 6.12, the hint inferred from (6.50) is given by

$$(I_x, \tilde{A}_1^{-1}\tilde{z}_1 + (\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K \begin{pmatrix} 0 \\ -\tilde{A}_2\tilde{A}_1^{-1}\tilde{z}_1 + \tilde{z}_2 \end{pmatrix}, \tilde{A}'K\tilde{A}).$$

Here,

$$\begin{aligned}
&\tilde{A}_1^{-1}\tilde{z}_1 + (\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K \begin{pmatrix} 0 \\ -\tilde{A}_2\tilde{A}_1^{-1}\tilde{z}_1 + \tilde{z}_2 \end{pmatrix} \\
=&\tilde{A}_1^{-1}\tilde{z}_1 + (\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K \begin{pmatrix} -\tilde{A}_1\tilde{A}_1^{-1}\tilde{z}_1 + \tilde{z}_1 \\ -\tilde{A}_2\tilde{A}_1^{-1}\tilde{z}_1 + \tilde{z}_2 \end{pmatrix} \\
=&\tilde{A}_1^{-1}\tilde{z}_1 - (\tilde{A}'K\tilde{A})^{-1}(\tilde{A}'K\tilde{A})\tilde{A}_1^{-1}\tilde{z}_1 + (\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K\tilde{z} \\
=&(\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K\tilde{z}.
\end{aligned}$$

Hence,

$$\tilde{h} = (I_x, (\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K\tilde{z}, \tilde{A}'K\tilde{A}) \in \mathcal{H}(\tilde{p}).$$

Further,

$$\tilde{A}'K\tilde{A} = A'P'P'^{-1}KP^{-1}PA = A'KA$$

and

$$\begin{aligned}
(\tilde{A}'K\tilde{A})^{-1}\tilde{A}'K\tilde{z} &= (A'KA)^{-1}A'P'P'^{-1}KP^{-1}Pz \\
&= (A'KA)^{-1}A'Kz.
\end{aligned}$$

Therefore, $\tilde{h} \in \mathcal{H}(\tilde{p})$ shows that $\mathcal{H}(p) = \mathcal{H}(\tilde{p})$. □

Notice that Theorem 6.23 can be applied to all equivalent Gaussian linear systems, and (6.48) is then a *canonical representation* of the result of the assumption-based inference. This will be discussed more thoroughly in Section 6.7.

Example **6.24** (Measurement Model Revisited). Reconsider the measurement model of Example 6.14. Assume that the errors ω_i are still independent but with possible different known variance σ_i^2 for $i \in \{1, \ldots, m\}$. Since $A = (1, 1, \ldots, 1)'$, Theorem 6.23 can be applied. The result of the assumption-based inference is thus the weighted sample mean

$$\left[\sum_{i=1}^{m} \frac{1}{\sigma_i^2}\right]^{-1} \sum_{i=1}^{m} \frac{z_i}{\sigma_i^2} \tag{6.51}$$

with variance

$$\left[\sum_{i=1}^{m} \frac{1}{\sigma_i^2}\right]. \tag{6.52}$$

Again, these results are the same as obtained by least-squares estimation, with the same different interpretation as remarked in Example 6.14. The results of Example 6.14 are of course reproduced for $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma_2$. ⊘

6.4 Marginalisation of Gaussian Hints

In order to extract the consequences of an assumption in a hint with respect to a smaller domain, its focal sets have to be projected to that domain. This idea is now applied to Gaussian hints. Let

$$h = (A, \mu, K) \in \mathcal{H}$$

be a Gaussian hint on $x \cup z$, $x \cap z = \emptyset$ with $A \in \mathbb{R}(m, x \cup z)$, $A_1 \in \mathbb{R}(m, x)$, $A_2 \in \mathbb{R}(m, z)$,

$$A = (A_1, A_2),$$

$\mu \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Then, if one is not interested in the variables z, the focal function $\Gamma : \mathbb{R}^m \to 2^{\mathbb{R}^{x \cup z}}$ becomes $\Gamma^{\downarrow x} : \mathbb{R}^m \to 2^{\mathbb{R}^x}$,

$$\Gamma^{\downarrow x}(\omega) = \{\mathbf{u}^{\downarrow x} : \ \mathbf{u} \in \Gamma(\omega) \subseteq \mathbb{R}^{x \cup z}\} \tag{6.53}$$
$$= \{\mathbf{x} \in \mathbb{R}^x : \ \exists \mathbf{z} \in \mathbb{R}^z \text{ s.t. } A_1\mathbf{x} + A_2\mathbf{z} + \omega = \mu\}. \tag{6.54}$$

In other words, $\mathbf{x} \in \mathbb{R}^x$ is a consequence of ω if and only if there is a complement $\mathbf{z} \in \mathbb{R}^z$ such that $(\mathbf{x}, \mathbf{z})$ is a consequence of ω. Further, an assumption cannot become impossible through marginalisation since $\Gamma(\omega) \subseteq \Gamma^{\downarrow x}(\omega) \times \mathbb{R}^z$. Further, $\Gamma^{\downarrow x}(\omega) = \Gamma^{\downarrow x}(\omega')$ does not imply $\Gamma(\omega) = \Gamma(\omega')$. Then the assumptions with the same focal set can be grouped together, which leads to an other equivalent hint.
How can now the projection these focal sets and their distribution be described? Can the result be described by a Gaussian hint again? The answers to both questions are affirmative. The strategy is the following: Find a transformation $T = \binom{T_1}{T_2}$ such that $TA_2 = \binom{0}{T_2 A_2}$ and such that $T_2 A_2$ has full row rank. Then,

$$\begin{aligned}\Gamma^{\downarrow x}(\omega) &= \{\mathbf{x} \in \mathbb{R}^x : \exists \mathbf{z} \in \mathbb{R}^z \text{ s.t. } T_1 A_1 \mathbf{x} + T_1 \omega = T_1 \mu, T_2 A_1 \mathbf{x} + T_2 A_2 \mathbf{z} + T_2 \omega = T_2 \mu\} \\ &= \{\mathbf{x} \in \mathbb{R}^x : \exists \mathbf{z} \in \mathbb{R}^z \text{ s.t. } T_1 A_1 \mathbf{x} + T_1 \omega = T_1 \mu\}\end{aligned}$$

since T_2A_2 has full row rank. Given such a transformation, the projection is characterised by the assumptions $T_1\omega$ and their marginal distribution, which is Gaussian. Therefore, it suffices to show how such a transformation can be found.

Let $k = \mathrm{r}(A_2)$ and let $B = (B_1, B_2) \in \mathbb{R}(m, m)$ be a regular matrix such that the submatrix $B_2 \in \mathbb{R}(m, k)$ is a basis of the column space $\mathcal{C}(A_2)$. Then, there is a matrix $C_2 \in \mathbb{R}(k, z)$ such that

$$A_2 = B_2C_2.$$

Define $T = B^{-1}$ and partition

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

such that $T_1 \in \mathbb{R}(m-k, m)$, $T_2 \in \mathbb{R}(k, m)$. Transformation by T leads to the equivalent hint

$$h' = (TA, T\mu, B'KB),$$

where

$$TA = \begin{pmatrix} T_1A_1 & T_1A_2 \\ T_2A_1 & T_2A_2 \end{pmatrix} = \begin{pmatrix} T_1A_1 & T_1B_2C_2 \\ T_2A_1 & T_2A_2 \end{pmatrix} = \begin{pmatrix} T_1A_1 & 0_{m-k,z} \\ T_2A_1 & T_2A_2 \end{pmatrix}. \tag{6.55}$$

The focal function of h' is

$$\begin{aligned} \Gamma'(\xi) &= \{\mathbf{u} \in \mathbb{R}^{x\cup z} : \ TA\mathbf{u} + \xi = T\mu\} \\ &= \{\mathbf{u} \in \mathbb{R}^{x\cup z} : \ T_1A_1\mathbf{x} + \xi_1 = T_1\mu, T_2A_1\mathbf{x} + T_2A_2\mathbf{z} + \xi_2 = T_2\mu\} \end{aligned}$$

for $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \in \mathbb{R}^m$, $\xi_1 \in \mathbb{R}^{m-k}$, $\xi_2 \in \mathbb{R}^k$. Then,

$$\begin{aligned} \Gamma(\omega) &= \{\mathbf{u} \in \mathbb{R}^{x\cup z} : \ A\mathbf{u} + \omega = \mu\} \\ &= \{\mathbf{u} \in \mathbb{R}^{x\cup z} : \ TA\mathbf{u} + T\omega = T\mu\} \\ &= \Gamma'(T\omega). \end{aligned}$$

Since T is regular and since A has full row rank m, the matrix TA has full row rank. In particular, T_1A and T_2A have full row rank. Therefore, for all $\mathbf{x} \in \mathbb{R}^x$ and all $\omega \in \mathbb{R}^m$ there is a $\mathbf{z} \in \mathbb{R}^z$ such that $T_2A_1\mathbf{x} + T_2A_2\mathbf{z} + T_2\omega = T_2\mu$. Therefore, using (6.54) and (6.55),

$$\begin{aligned} \Gamma^{\downarrow x}(\omega) \quad &= \quad \Gamma'^{\downarrow x}(T\omega) \\ =&\{\mathbf{x} \in \mathbb{R}^x : \ \exists \mathbf{z} \in \mathbb{R}^z \text{ s.t. } T_1A_1\mathbf{x} + T_1\omega = T_1\mu, T_2A_1\mathbf{x} + T_2A_2\mathbf{z} + T_2\omega = T_2\mu\} \\ =&\{\mathbf{x} \in \mathbb{R}^x : \ \exists \mathbf{z} \in \mathbb{R}^z \text{ s.t. } T_1A_1\mathbf{x} + T_1\omega = T_1\mu\} \\ =&\{\mathbf{x} \in \mathbb{R}^x : \ T_1A_1\mathbf{x} + T_1\omega = T_1\mu\}. \end{aligned}$$

These focal sets $\Gamma^{\downarrow x}(\omega)$ are parallel affine linear manifolds of dimension

$$|x| - \mathrm{r}(T_1A_1) = |x| - (m-k) = |x| - (\mathrm{r}(A) - \mathrm{r}(A_2)).$$

Here, according to Appendix B.3, the marginal distribution of the disturbances $\xi_1 = T_1\omega$ has mean 0_{m-k} and concentration $(T_1K^{-1}T_1')^{-1}$. Therefore, the Gaussian hint

$$(T_1A_1, T_1\mu, (T_1K^{-1}T_1')^{-1})$$

represents the information contained in h about x. As observed above, T_1A has full row rank.

Definition **6.25**. *Let $A = (A_1, A_2) \in \mathbb{R}(m, x \cup z)$, $A_1 \in \mathbb{R}(m, x)$, $A_2 \in \mathbb{R}(m, z)$. Let $k = \mathrm{r}(A_2)$. A matrix $T_1 \in \mathbb{R}(m-k, m)$ of full row rank $m-k$ such that $T_1A_2 = 0$ is called a* projection matrix *for A to x.* ⊘

The following lemma shows that using different projection matrices on equivalent Gaussian hints yields equivalent Gaussian hints again.

Lemma **6.26**. *Let $h = (A, z, K)$ and $\tilde{h} = (\tilde{A}, \tilde{z}, \tilde{K})$ be equivalent Gaussian hints, $A, \tilde{A} \in \mathbb{R}(m, x)$, $z, \tilde{z} \in \mathbb{R}^m$, $K, \tilde{K} \in \mathbb{R}(m, m)$,*

$$A = (A_1, A_2), \quad \tilde{A} = (\tilde{A}_1, \tilde{A}_2),$$

$A_1, \tilde{A}_1 \in \mathbb{R}(m, x)$, $k = \mathrm{r}(A_2)$. Let $C \in \mathbb{R}(m, m)$ be the regular matrix that establishes the equivalence $h \cong \tilde{h}$, i.e.

$$CA = \tilde{A}, \quad Cz = \tilde{z}, \quad C^{-1'}KC^{-1} = \tilde{K}.$$

Further, let T_1 and $\tilde{T}_1$ be projection matrices for h and $\tilde{h}$ to $x \subseteq d(h) = d(\tilde{h})$, respectively. Then,

$$(T_1A_1, T_1z, (T_1K^{-1}T_1')^{-1}) \cong (\tilde{T}_1\tilde{A}_1, \tilde{T}_1\tilde{z}, (\tilde{T}_1\tilde{K}^{-1}\tilde{T}_1')^{-1}).$$

Furthermore, $\tilde{T}_1$ is also a projection matrix for h to x. ⊘

Proof. Observe that

$$CA_2 = \tilde{A}_2,$$

and, since C is regular,

$$\mathrm{r}(\tilde{A}_2) = \mathrm{r}(CA_2) = \mathrm{r}(A_2) = k$$

in light of Lemma 8.3.2 of (Harville, 1997; p.83). Therefore, $\tilde{T}_1$ being a projection matrix for $\tilde{A}$, it has the same dimensions and rank as T_1, i.e. $\tilde{T}_1 \in \mathbb{R}(m-k, m)$ and $\mathrm{r}(\tilde{T}_1) = m - k$. Also, it follows from

$$T_1A_2 = 0 = \tilde{T}_1\tilde{A}_2 = \tilde{T}_1CA_2 \quad \iff \quad A_2'T_1' = 0 = \tilde{A}_2'\tilde{T}_1' = A_2'C'\tilde{T}_1'$$

that

$$\mathcal{C}(T_1'), \mathcal{C}(C'\tilde{T}_1') \subseteq \mathcal{N}(A_2'). \tag{6.56}$$

Furthermore,

$$\dim(\mathcal{N}(A_2')) = m - \mathrm{r}(A_2') = m - \mathrm{r}(A_2) = m - k$$

in light of Lemma 11.3.1 of (Harville, 1997; p.142). On the other hand, since C' is regular,

$$\mathrm{r}(T_1') = \mathrm{r}(T_1) = m-k = \mathrm{r}(\tilde{T}_1) = \mathrm{r}(\tilde{T}_1') = \mathrm{r}(C'\tilde{T}_1').$$

Hence, using (6.56) and $\mathrm{r}(T_1') = \dim(\mathcal{N}(A_2')) = \mathrm{r}(C'\tilde{T}_1')$, it follows that

$$\mathcal{C}(T_1') = \mathcal{N}(A_2') = \mathcal{C}(C'\tilde{T}_1').$$

Therefore,

$$\mathcal{C}(T_1') = \mathcal{C}(C'\tilde{T}_1') \quad \Longleftrightarrow \quad \mathcal{R}(T_1) = \mathcal{R}(\tilde{T}_1 C),$$

i.e. every row of T_1 is a linear combination of the rows of $\tilde{T}_1 C$. Hence, there is a matrix $D \in \mathbb{R}(m-k, m-k)$ such that

$$T_1 = \tilde{T}_1 C D. \tag{6.57}$$

In light of Corollary 4.4.5 of (Harville, 1997; p.37),

$$m-k = \mathrm{r}(T_1) = \mathrm{r}(\tilde{T}_1 CD) \leq \mathrm{r}(\tilde{T}_1 C) =\leq \mathrm{r}(\tilde{T}_1) = m-k.$$

From $\mathrm{r}(\tilde{T}_1 CD) = m-k = \mathrm{r}(\tilde{T}_1 C)$, it follows that $\mathrm{r}(D) \geq m-k$ in light of Corollary 4.4.5 of (Harville, 1997; p.37). Hence, $D \in \mathbb{R}(m-k, m-k)$ is regular. Therefore, multiplying both sides of (6.57) on the right-hand side by $D^{-1}C^{-1}$ yields

$$T_1 D^{-1} C^{-1} = \tilde{T}_1. \tag{6.58}$$

Since D is regular,

$$\mathcal{R}(T_1 D^{-1}) = \mathcal{R}(T_1), \quad \mathrm{r}(T_1 D^{-1}) = \mathrm{r}(T_1)$$

in light of Lemma 8.3.2 of (Harville, 1997; p.83). Hence, there is matrix $E \in \mathbb{R}(m-k, m-k)$ such that

$$E T_1 = T_1 D^{-1}, \tag{6.59}$$

which is regular since

$$m-k = \mathrm{r}(T_1 D^{-1}) = \mathrm{r}(E T_1) \leq \mathrm{r}(E) \leq m-k$$

in light of Lemma 11.3.1 of (Harville, 1997; p.142). Using (6.58) and (6.59),

$$\tilde{T}_1 \tilde{A}_1 = T_1 D^{-1} C^{-1} C A_1 = E(T_1 A_1), \quad \tilde{T}_1 \tilde{z} = T_1 D^{-1} C^{-1} C z = E(T_1 z),$$

and

$$\begin{aligned}
(\tilde{T}_1 \tilde{K}^{-1} \tilde{T}_1')^{-1} &= (T_1 D^{-1} C^{-1} C K^{-1} C' C^{-1\prime} D^{-1\prime} T_1')^{-1} \\
&= (E T_1 K^{-1} C' C'^{-1} T_1' E')^{-1} \\
&= E^{-1\prime} (T_1 K^{-1} T_1')^{-1} E^{-1}.
\end{aligned}$$

So the regular matrix E establishes that indeed

$$(T_1 A, T_1 \mu, (T_1 K^{-1} T_1')^{-1}) \cong (\tilde{T}_1 A, \tilde{T}_1 \mu, (\tilde{T}_1 K^{-1} \tilde{T}_1')^{-1}).$$

Furthermore, since $D^{-1}C^{-1}$ is regular, by the same argument as above, there is a regular matrix $\tilde{E} \in \mathbb{R}(m-k, m-k)$ such that

$$T_1 D^{-1} C^{-1} = \tilde{E} T_1.$$

Then,

$$\tilde{T}_1 A_2 = T_1 D^{-1} C^{-1} A_2 = \tilde{E} T_1 A_2 = \tilde{E} 0 = 0$$

shows that $\tilde{T}_1$ is also a projection matrix for h to x. □

In light of the second assertion of Lemma 6.26, a projection matrix P for a particular representative $h \in \mathcal{H}$ is also a projection matrix for any $\tilde{h} \in \mathcal{H}(h)$.

DEFINITION **6.27**. *The marginalisation of Gaussian hints* $\downarrow : \mathcal{H} \times D \to \mathcal{H}$, $(h, x) \mapsto h^{\downarrow x}$ *for* $x \subseteq d(h)$ *is defined up to equivalence by*

$$h^{\downarrow x} \cong (T_1 A_1, T_1 \mu, (T_1 K^{-1} T_1')^{-1}) \tag{6.60}$$

where T_1 *is any projection matrix for* A *to* x. ⊘

Marginalisation of Gaussian Linear Systems

The marginalisation of Gaussian hints can easily be generalised to Gaussian linear systems since projection matrices also exist for matrices A which do not have full row rank (as can be verified by the same argument as for Gaussian hints above without the assumption of full row rank). Applying a projection matrix P to a Gaussian linear system (A, z, K) yields the Gaussian linear system

$$(PA_1, Pz, (PK^{-1}P')^{-1}).$$

The following lemma shows that different projection matrices lead to equivalent Gaussian linear systems.

LEMMA **6.28**. *Let* $p = (A, z, K)$ *be a Gaussian linear system on* x, $A \in \mathbb{R}(m, x)$, $z \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$ *symmetric and positive definite. Let* $s \subseteq x$ *and partition*

$$A = (A_1, A_2),$$

$A_1 \in \mathbb{R}(m, s)$, $A_2 \in \mathbb{R}(m, x - s)$. *Let* $r = \mathrm{r}(A)$, $k_1 = \mathrm{r}(A_1)$, $k_2 = \mathrm{r}(A_2)$ *and let* $P, \tilde{P} \in \mathbb{R}(m-k, m)$ *be projection matrices for* p *to* s, *i.e. they are matrices of full row rank* $\mathrm{r}(P) = m - k = \mathrm{r}(\tilde{P})$ *and* $PA_2 = 0 = \tilde{P}A_2$. *Then,*

$$\mathrm{r}(PA) = \mathrm{r}(\tilde{P}A) = k - k_2.$$

Further, applying P *and* $\tilde{P}$ *to* p *and* $\tilde{p}$ *yields*

$$p_s = (PA_1, Pz, (PK^{-1}P')^{-1})$$

and

$$\tilde{p}_s = (\tilde{P}A_1, \tilde{P}z, (\tilde{P}K^{-1}\tilde{P}')^{-1}),$$

respectively. Then, $\mathcal{H}(p_s) = \mathcal{H}(\tilde{p}_s)$ ⊘

Proof. Since

$$PA_2 = 0 = \tilde{P}A_2 = 0 \quad \Longleftrightarrow \quad A_2'P' = 0 = A_2'\tilde{P}',$$

it follows that

$$\mathcal{C}(P'), \mathcal{C}(\tilde{P}') \subseteq \mathcal{N}(A_2').$$

Furthermore, $\dim(\mathcal{N}(A_2')) = m - \mathrm{r}(A_2') = m - \mathrm{r}(A_2) = m - k_2 = \dim(\mathcal{N}(A_2))$ in light of Lemma 11.3.1 of (Harville, 1997; p.142). Since

$$\mathrm{r}(P') = \mathrm{r}(P) = m - k_2 = \mathrm{r}(\tilde{P}) = \mathrm{r}(\tilde{P}'),$$

it follows that $\mathcal{C}(P') = \mathcal{N}(A_2') = C(\tilde{P}')$. Hence, $\mathcal{R}(P) = \mathcal{R}(\tilde{P})$. Let $C \in \mathbb{R}(k_2, m)$ be a basis of $\mathcal{R}(A_2)$. Then, the rows of

$$D = \begin{pmatrix} P \\ C \end{pmatrix} \quad \text{and} \quad \tilde{D} = \begin{pmatrix} \tilde{P} \\ C \end{pmatrix}$$

are linearly independent, so D and $\tilde{D}$ are regular. Since D is regular, $\mathrm{r}(DA_2) = \mathrm{r}(A_2) = k_2$. Since

$$DA_2 = \begin{pmatrix} 0_{m-k_2,x-s} \\ CA_2 \end{pmatrix},$$

it follows that $\mathrm{r}(CA_2) = \mathrm{r}(DA_2) = k_2$. Then, since DA is block-triangular,

$$DA = \begin{pmatrix} PA \\ CA \end{pmatrix} = \begin{pmatrix} PA_1 & 0 \\ CA_1 & CA_2 \end{pmatrix},$$

and since CA_2 has full row rank $\mathrm{r}(CA_2) = k_2$, it follows that

$$\mathrm{r}(PA_1) = \mathrm{r}(DA) - \mathrm{r}(CA_2) = k - k_2$$

in light of Lemma 8.5.3 of (Harville, 1997; p.90). In the same way, it can be proved that $\mathrm{r}(\tilde{P}A_1) = k - k_2$.
Since $\mathcal{R}(P) = \mathcal{R}(\tilde{P})$, there is a matrix $B \in \mathbb{R}(m - k_2, m - k_2)$ such that

$$\tilde{P} = BP.$$

Then, $m - k_2 = \mathrm{r}(\tilde{P}) = \mathrm{r}(BP) \leq \mathrm{r}(B) \leq m - k_2$ shows that B is regular. Further, the matrix B establishes

$$\tilde{P}A_1 = B(PA_1), \qquad \tilde{P}z = B(Pz)$$

and

$$(\tilde{P}K^{-1}\tilde{P}')^{-1} = ((BP)K^{-1}(BP)')^{-1} = B^{-1'}(PK^{-1}P')^{-1}B^{-1}.$$

Hence, B establishes that p_s and $\tilde{p}_s$ are equivalent Gaussian linear systems in light of Lemma 6.22, i.e. $\mathcal{H}(p_s) = \mathcal{H}(\tilde{p}_s)$. □

Since applying different projection matrices to a Gaussian linear system yields equivalent Gaussian linear systems, projection of Gaussian linear systems can be defined up to equivalence by

$$(A, z, K)^{\downarrow s} \cong (PA, Pz, (PK^{-1}P')^{-1}) \tag{6.61}$$

where P is any projection matrix to the domain s. Applying this definition to a Gaussian hint $h \in \mathcal{H} \subseteq \mathfrak{L}$ yields a Gaussian hint $h^{\downarrow s} \in \mathcal{H} \subseteq \mathfrak{L}$. Therefore, it is sound to use the same symbol $\downarrow$ also for the marginalisation of general Gaussian linear systems.

The following lemma shows that projection and inference commute: The hint obtained from the projected Gaussian linear system is equivalent to the one obtained by projecting the hint obtained from the original Gaussian linear system. In other words, the projections of equivalent Gaussian linear systems are equivalent.

LEMMA **6.29**. *Let $p = (A, z, K)$ be a Gaussian linear system on x, $A \in \mathbb{R}(m, x)$, $z \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Let $s \subseteq x$ and partition*

$$A = (A_1, A_2),$$

$A_1 \in \mathbb{R}(m, s)$, $A_2 \in \mathbb{R}(m, x - s)$. Let $r = \mathrm{r}(A)$, $k_1 = \mathrm{r}(A_1)$, $k_2 = \mathrm{r}(A_2)$. Let

- *h be the hint inferred from p by an admissible matrix T,*
- *$h^{\downarrow s}$ be the projection of h by a projection matrix P,*
- *p_s be the projection of p by a projection matrix $\tilde{P}$, and let*
- *h_s be the hint inferred from p_s by an admissible matrix $\tilde{T}$.*

Then, h_s is equivalent to the projection of h to s,

$$h_s \cong h^{\downarrow s}.$$

⊘

The situation of this lemma is shown in Figure 6.5. Inferences are shown by a horizontal arrow labelled by an admissible matrix, projections by a vertical arrow labelled by the projection matrix used.

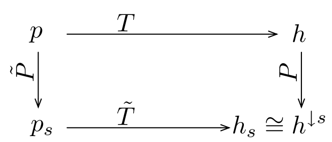


FIGURE 6.5: Projection and inference of Gaussian linear systems commute.

Proof. On the one hand, let $T \in \mathbb{R}(m,m)$ be an admissible matrix for p,

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

$T_1 \in \mathbb{R}(k,m)$, $T_2 \in \mathbb{R}(m-k,m)$, $T_2A = 0$ and

$$B = T^{-1} = (B_1, B_2),$$

$B_1 \in \mathbb{R}(m,k)$, $B_2 \in \mathbb{R}(m,m-k)$. This leads to the Gaussian hint $h = (A_h, z_h, K_h)$,

$$A_h = T_1A,\ z_h = T_1z + (B_1'KB_1)^{-1}(B_1'KB_2)T_2z,\ K_h = B_1'KB_1.$$

Further, let $P \in \mathbb{R}(k-k_2,k)$ be a projection matrix such that

$$PT_1A_2 = 0 \tag{6.62}$$

and $\mathrm{r}(T_1A_2) = \mathrm{r}(A_2) = k_2$. Then, the projection of h to s is $h_s = (A_s, z_s, K_s)$ where

$$A_s = PT_1A_1,\ z_s = P(T_1z + (B_1'KB_1)^{-1}(B_1'KB_2)T_2z),\ K_s = (P(B_1'KB_1)^{-1}P')^{-1}$$

On the other hand, let $\tilde{P} \in \mathbb{R}(m-k_2,m)$ be a projection matrix for p to s, i.e. of rank $\mathrm{r}(P) = m-k_2$ such that $\tilde{P}A_2 = 0$. Define

$$p_s = (\tilde{P}A_1, \tilde{P}z, (\tilde{P}K^{-1}\tilde{P}')^{-1}).$$

Further, let $h_s = (\tilde{A}_s, \tilde{z}_s, \tilde{K}_s)$ be the hint inferred from p_s,

$$\tilde{A}_s = \tilde{T}_1\tilde{P}A_1,\ \tilde{K}_s = (\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_1),$$

$$\tilde{z}_s = \tilde{T}_1\tilde{P}z + \left(\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_1\right)^{-1}(\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_2)\tilde{T}_2\tilde{P}z)$$

under an admissible matrix $\tilde{T} \in \mathbb{R}(m-k_2, m-k_2)$,

$$\tilde{T} = \begin{pmatrix} \tilde{T}_1 \\ \tilde{T}_2 \end{pmatrix},$$

$\tilde{T}_1 \in \mathbb{R}(k-k_2, m-k_2)$, $\tilde{T}_2 \in \mathbb{R}(m-k, m-k_2)$, $\tilde{T}_2A = 0$ and

$$\tilde{B} = \tilde{T}^{-1} = (\tilde{B}_1, \tilde{B}_2),$$

$\tilde{B}_1 \in \mathbb{R}(m-k_2, k-k_2)$, $\tilde{B}_2 \in \mathbb{R}(m-k_2, m-k)$.
It is now going to be shown that there are regular matrices $R_1 \in \mathbb{R}(k-k_2, k-k_2)$, $R_2 \in \mathbb{R}(k_2,k_2)$ and $R_3 \in \mathbb{R}(m-k, m-k)$ such that

$$(\tilde{T}_1\tilde{P}) = R_1(PT_1), \quad \tilde{P}_2 = R_2(P_2T_1), \quad \tilde{T}_2\tilde{P} = R_3T_2. \tag{6.63}$$

Let $P_2 \in \mathbb{R}(k_2,k)$ and $\tilde{P}_2 \in \mathbb{R}(k_2,m)$ be matrices such that

$$\begin{pmatrix} P \\ P_2 \end{pmatrix} \in \mathbb{R}(k,k)$$

and

$$\begin{pmatrix} \tilde{P} \\ \tilde{P}_2 \end{pmatrix} \in \mathbb{R}(m,m)$$

are regular. Further, define $C, \tilde{C} \in \mathbb{R}(m,m)$, $C_1, \tilde{C}_1 \in \mathbb{R}(k-k_2,m)$, $C_2, \tilde{C}_2 \in \mathbb{R}(k_2,m)$ and $C_3, \tilde{C}_3 \in \mathbb{R}(m-k,m)$ by

$$C = \begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} PT_1 \\ P_2T_1 \\ T_2 \end{pmatrix}, \quad \tilde{C} = \begin{pmatrix} \tilde{C}_1 \\ \tilde{C}_2 \\ \tilde{C}_3 \end{pmatrix} = \begin{pmatrix} \tilde{T}_1\tilde{P} \\ \tilde{P}_2 \\ \tilde{T}_2\tilde{P} \end{pmatrix},$$

which are both regular since they are the product of regular matrices,

$$\begin{pmatrix} PT_1 \\ P_2T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} P & \\ P_2 & \\ & I_k \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

and

$$\begin{pmatrix} \tilde{T}_1\tilde{P} \\ \tilde{P}_2 \\ \tilde{T}_2\tilde{P} \end{pmatrix} = \begin{pmatrix} I_{k-k_2} & & \\ & & I_{k-2} \\ & I_{m-k} & \end{pmatrix} \begin{pmatrix} \tilde{T}_1 & \\ \tilde{T}_2 & \\ & I_k \end{pmatrix} \begin{pmatrix} \tilde{P} \\ \tilde{P}_2 \end{pmatrix}.$$

Then, using (6.62),

$$CA = \begin{pmatrix} PT_1 \\ P_2T_1 \\ T_2 \end{pmatrix} (A_1, A_2) = \begin{pmatrix} PT_1A_1 & 0 \\ P_2T_1A_1 & P_2T_1A_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_1A_1 & 0 \\ C_2A_1 & C_2A_2 \\ 0 & 0 \end{pmatrix}$$

and

$$\tilde{C}A = \begin{pmatrix} \tilde{T}_1\tilde{P} \\ \tilde{P}_2 \\ \tilde{T}_2\tilde{P} \end{pmatrix} (A_1, A_2) = \begin{pmatrix} \tilde{T}_1\tilde{P}A_1 & 0 \\ \tilde{P}_2A_1 & \tilde{P}_2A_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \tilde{C}_1A_1 & 0 \\ \tilde{C}_2A_1 & \tilde{C}_2A_2 \\ 0 & 0 \end{pmatrix}.$$

Observe that

$$A'C_3' = 0 = A'\tilde{C}_3',$$

hence $\mathcal{C}(C_3), \mathcal{C}(\tilde{C}_3') \subseteq \mathcal{N}(A')$. Then, since

$$\mathrm{r}(C_3') = \mathrm{r}(C_3) = m - k = \mathrm{r}(\tilde{C}_3) = \mathrm{r}(\tilde{C}_3')$$

and since

$$\dim(\mathcal{N}(A')) = m - r(A) = m - k$$

in light of Lemma 11.3.1 of (Harville, 1997; p.142), it follows that

$$\mathcal{C}(C_3') = \mathcal{N}(A') = \mathcal{C}(\tilde{C}_3').$$

Hence,

$$\mathcal{R}(C_3) = \mathcal{R}(\tilde{C}_3). \tag{6.64}$$

Similarly,

$$A_2'(C_1', C_3') = 0 = A_2'(\tilde{C}_1', \tilde{C}_3')$$

implies that

$$\mathcal{C}(\begin{pmatrix} C_1', C_3' \end{pmatrix}), \mathcal{C}(\begin{pmatrix} \tilde{C}_1', \tilde{C}_3' \end{pmatrix}) \subseteq \mathcal{N}(A_2').$$

Then, since C and $\tilde{C}$ are regular,

$$\mathrm{r}(\begin{pmatrix} C_1', C_3' \end{pmatrix}) = m - k_2 = \mathrm{r}(\begin{pmatrix} \tilde{C}_1', \tilde{C}_3' \end{pmatrix})$$

and

$$\dim(\mathcal{N}(A_2')) = m - \mathrm{r}(A_2') = m - \mathrm{r}(A_2) = m - k_2$$

in light of Lemma 11.3.1 of (Harville, 1997; p.142). Hence,

$$\mathcal{C}(\begin{pmatrix} C_1', C_3' \end{pmatrix}) = \mathcal{N}(A_2') = \mathcal{C}(\begin{pmatrix} \tilde{C}_1', \tilde{C}_3' \end{pmatrix}),$$

and thus

$$\mathcal{R}(\begin{pmatrix} C_1 \\ C_3 \end{pmatrix}) = \mathcal{R}(\begin{pmatrix} \tilde{C}_1 \\ \tilde{C}_3 \end{pmatrix}). \tag{6.65}$$

Since the rows of C and $\tilde{C}$ are linearly independent and since they have the same row space $\mathcal{R}(C) = \mathbb{R}(1, m) = \mathcal{R}(\tilde{C})$, it follows from equations (6.64) and (6.65) that

$$\mathcal{R}(C_1) = \mathcal{R}(\tilde{C}_1). \tag{6.66}$$

Similarly, since the rows of C and $\tilde{C}$ are linearly independent, it follows from equation (6.65) that

$$\mathcal{R}(C_2) = \mathcal{R}(\tilde{C}_2). \tag{6.67}$$

Hence, in light of (6.66), the rows of $\tilde{C}_1$ are linear combinations of the rows of C_1. Therefore, there is a matrix $R_1 \in \mathbb{R}(k - k_2, k - k_2)$ such that

$$\tilde{T}_1 \tilde{P} = \tilde{C}_1 = R_1 C_1 = R_1 (P T_1)$$

which is regular since $k - k_2 = \mathrm{r}(\tilde{C}_1) = \mathrm{r}(R_1 C_1) \leq \mathrm{r}(R_1) \leq k - k_2$. Hence, there is indeed such a regular matrix R_1 as claimed. The existence of R_2 and R_3 can be derived in the same way from (6.67) and (6.64), respectively. This shows that the matrices R_1, R_2, R_3 satisfying (6.63) exist.
Finally, it is going to be shown that R_1 establishes the equivalence $h_s \cong h^{\downarrow s}$. Observe that

$$\begin{pmatrix} B_1' K B_1 & B_1' K B_2 \\ B_2' K B_1 & B_2' K B_2 \end{pmatrix} = B' K B = (T K^{-1} T')^{-1} = \begin{pmatrix} T_1 K^{-1} T_1' & T_1 K^{-1} T_2' \\ T_2 K^{-1} T_1' & T_2 K^{-1} T_2' \end{pmatrix}^{-1} \tag{6.68}$$

and

$$\begin{pmatrix} \tilde{B}_1' (\tilde{P} K^{-1} \tilde{P}')^{-1} \tilde{B}_1 & \tilde{B}_1' (\tilde{P} K^{-1} \tilde{P}')^{-1} \tilde{B}_2 \\ \tilde{B}_2' (\tilde{P} K^{-1} \tilde{P}')^{-1} \tilde{B}_1 & \tilde{B}_2' (\tilde{P} K^{-1} \tilde{P}')^{-1} \tilde{B}_2 \end{pmatrix} = \begin{pmatrix} \tilde{T}_1 (\tilde{P} K^{-1} \tilde{P}') \tilde{T}_1' & \tilde{T}_1 (\tilde{P} K^{-1} \tilde{P}') \tilde{T}_2' \\ \tilde{T}_2 (\tilde{P} K^{-1} \tilde{P}') \tilde{T}_1' & \tilde{T}_2 (\tilde{P} K^{-1} \tilde{P}') \tilde{T}_2' \end{pmatrix}^{-1}. \tag{6.69}$$

Firstly, using (6.63), $\tilde{T}_1\tilde{P}A_1 = R_1(PT_1A_1)$, i.e. the design matrices of h_s and $h^{\downarrow s}$ are related by R_1. Secondly,

$$\begin{aligned}
&\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_1 \\
&\overset{(1)}{=}(\tilde{T}_1\tilde{P}K^{-1}\tilde{P}'\tilde{T}_1' - \tilde{T}_1\tilde{P}K^{-1}\tilde{P}'\tilde{T}_2'(\tilde{T}_2\tilde{P}K^{-1}\tilde{P}'\tilde{T}_2')^{-1}\tilde{T}_2\tilde{P}K^{-1}\tilde{P}'\tilde{T}_1')^{-1} \\
&\overset{(2)}{=}(R_1PT_1K^{-1}T_1'P'R_1' - R_1PT_1K^{-1}T_2'R_3'(R_3T_2K^{-1}T_2'R_3')^{-1}R_3T_2K^{-1}T_1'P'R_1')^{-1} \\
&\overset{(3)}{=}(R_1P\left(T_1K^{-1}T_1' - T_1K^{-1}T_2'R_3'R_3'^{-1}(T_2K^{-1}T_2')^{-1}R_3^{-1}R_3T_2K^{-1}T_1'\right)P'R_1')^{-1} \\
&\overset{(4)}{=}(R_1P(T_1K^{-1}T_1' - T_1K^{-1}T_2'(T_2K^{-1}T_2')^{-1}T_2K^{-1}T_1')P'R_1')^{-1} \\
&\overset{(5)}{=}R_1^{-1\prime}(P(B_1'KB_1)^{-1}P')^{-1}R_1^{-1},
\end{aligned}$$

applying (A.3) to (6.69) in (1), using (6.63) in (2), and applying (A.3) to (6.68) in (5). This shows that the concentrations matrices are also related by R_1. Thirdly,

$$\begin{aligned}
&\tilde{T}_1\tilde{P}z + \left(\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_1\right)^{-1}\left(\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_2\right)\tilde{T}_2\tilde{P}z \\
&\overset{(1)}{=}\tilde{T}_1\tilde{P}z - \tilde{T}_1\tilde{P}K^{-1}\tilde{P}'\tilde{T}_2'(\tilde{T}_2\tilde{P}K^{-1}\tilde{P}'\tilde{T}_2')^{-1}\tilde{T}_2\tilde{P}z \\
&\overset{(2)}{=}R_1PT_1z - R_1PT_1K^{-1}T_1'P'R_1'(R_3T_2K^{-1}T_2'R_3')^{-1}R_3T_2z \\
&\overset{(3)}{=}R_1PT_1z - R_1PT_1K^{-1}T_2'R_3'R_3'^{-1}(T_2K^{-1}T_2')^{-1}R_3^{-1}R_3T_2z \\
&\overset{(4)}{=}R_1P(T_1z - (T_1K^{-1}T_2'(T_2K^{-1}T_2')^{-1})T_2z) \\
&\overset{(5)}{=}R_1P(T_1z + (B_1'KB_1)^{-1}(B_1'KB_2)T_2z),
\end{aligned}$$

using (B.13) in (1) [with $K_{11} = (\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_1)$ and $K_{12} = (\tilde{B}_1'(\tilde{P}K^{-1}\tilde{P}')^{-1}\tilde{B}_2)$], then (6.63) in (2), and again (B.13) in (5) [with $\Sigma_{12} = T_1K^{-1}T_2'$ and $\Sigma_{22} = T_2K^{-1}T_2'$]. This shows that the mean vectors are related by R_1, as well. Hence, indeed $h_s \cong h^{\downarrow s}$. □

6.5 Combination of Gaussian Hints

Dempster's Rule may be applied to Gaussian hints $h_1 = (A_1, z_1, K_1)$ and $h_2 = (A_2, z_2, K_2)$ (of possibly different domains x and y) where $A_1 \in \mathbb{R}(m_1, x)$, $A_2 \in \mathbb{R}(m_2, y)$. Using the convention $\mathbb{R}^x \times \mathbb{R}^y = \mathbb{R}^{x\cup y}$ (if $x \cap y = \emptyset$), this leads to focal sets

$$\begin{aligned}
\Gamma(\omega_1, \omega_2) &= (\Gamma_1(\omega_1) \times \mathbb{R}^{y-x}) \cap (\Gamma_2(\omega_2) \times \mathbb{R}^{x-y}) && (6.70)\\
&= \{\mathbf{u} \in \mathbb{R}^{x\cup y} : A_1\mathbf{u}^{\downarrow x} + \omega_1 = z_1, A_2\mathbf{u}^{\downarrow y} + \omega_2 = z_2\} && (6.71)\\
&= \{\mathbf{u} \in \mathbb{R}^{x\cup y} :\ A\mathbf{u} + \omega = z\} && (6.72)
\end{aligned}$$

where

$$A = \begin{pmatrix} A_1{}^{\downarrow x-y} & A_1{}^{\downarrow x\cap y} & 0 \\ 0 & A_2{}^{\downarrow x\cap y} & A_2{}^{\downarrow y-x} \end{pmatrix}, \qquad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \tag{6.73}$$

and the disturbances $\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \in \mathbb{R}^{m_1+m_2}$ have density $\phi_{0,K}$ where

$$K = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}. \tag{6.74}$$

This defines the operation of joining Gaussian linear systems $\oplus : \mathfrak{L} \times \mathfrak{L} \to \mathfrak{L}$ by

$$(A_1, z_1, K_1) \oplus (A_2, z_2, K_2) = (A, z, K). \tag{6.75}$$

The following lemma shows that the Gaussian hint inferred from a joint Gaussian linear system can also be obtained from the joint system of the hints from the original Gaussian linear systems. The situation is shown in Figure 6.6.

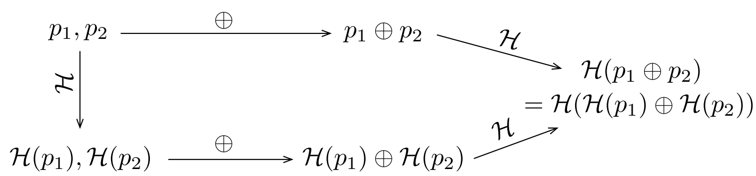


FIGURE 6.6: Joining Gaussian linear systems and their Gaussian hints

LEMMA **6.30**. *Let $p_1, p_2 \in \mathfrak{L}$ be Gaussian linear systems. Then,*

$$\mathcal{H}(p_1 \oplus p_2) = \mathcal{H}\big(\mathcal{H}(p_1) \oplus \mathcal{H}(p_2)\big). \tag{6.76}$$

⊘

PROOF. First, two simplifying assumptions can be made.

- Let $p_1 = (A_1, z_1, K_1)$ and $p_2 = (A_2, z_2, K_2)$ of domains x_1 and x_2, $A_1 \in \mathbb{R}(m_1, x_1)$, $A_2 \in \mathbb{R}(m_2, x_2)$, $z_1 \in \mathbb{R}^{m_1}$, $z_2 \in \mathbb{R}^{m_2}$, $K_1 \in \mathbb{R}(m_1, m_1)$ and $K_2 \in \mathbb{R}(m_2, m_2)$. Then extend the design matrices vacuously by defining $\tilde{A}_1 = (A_1, 0_{m_1, x_2 - x_1}) \in \mathbb{R}(m_1, x)$ and $\tilde{A}_2 = (A_2, 0_{m_2, x_1 - x_2}) \in \mathbb{R}(m_2, x_2)$. Further, let $\tilde{p}_1 = (\tilde{A}_1, z_1, K_1)$ and $\tilde{p}_2 = (\tilde{A}_2, z_2, K_2)$. It is readily verified that $p_1 \oplus p_2 = \tilde{p}_1 \oplus \tilde{p}_2$ and $\mathcal{H}(p_1) \oplus \mathcal{H}(p_2) = \mathcal{H}(\tilde{p}_1) \oplus \mathcal{H}(\tilde{p}_2)$. Therefore, assume without loss of generality that $x = x_1 = x_2$.

- The transformation by an admissible matrices T_1 and T_2 yields equivalent systems $\tilde{p}_1 = (\tilde{A}_1, z_1, K_1)$ and $\tilde{p}_2 = (\tilde{A}_2, z_2, K_2)$. Then, the matrix $T = \left(\begin{smallmatrix} T_1 & 0 \\ 0 & T_2 \end{smallmatrix}\right)$ establishes the equivalence of $p_1 \oplus p_2$ and $\tilde{p}_1 \oplus \tilde{p}_2$. Therefore, it can be assumed without loss of generality that $A_1 = \left(\begin{smallmatrix} A_{11} \\ 0 \end{smallmatrix}\right)$ and $A_2 = \left(\begin{smallmatrix} A_{21} \\ 0 \end{smallmatrix}\right)$ such that A_{11} and A_{21} have full row rank.

Further, in order to simplify the proof, variance-covariance matrices will be used instead of concentration matrices. Therefore, define $\varSigma_1 = {K_1}^{-1}$ and $\varSigma_2 = {K_2}^{-1}$. Consider first the right-hand side of equation (6.76). According to the rank of A_1

and A_2, partition $z_1 = \binom{z_{11}}{z_{12}}$, $z_2 = \binom{z_{21}}{z_{22}}$, $\Sigma_1 = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} \Sigma_{33} & \Sigma_{34} \\ \Sigma_{43} & \Sigma_{44} \end{pmatrix}$. Then, using Theorem 6.12,

$$\begin{aligned} h_1 &= (A_{11}, z_{11} + \Sigma_{12}{\Sigma_{22}}^{-1} z_{12}, \hat{\Sigma}_{11}), \quad \hat{\Sigma}_{11} = \Sigma_{11} - \Sigma_{12}{\Sigma_{22}}^{-1}\Sigma_{21}, \text{ and} \\ h_2 &= (A_{21}, z_{21} + \Sigma_{34}{\Sigma_{44}}^{-1} z_{22}, \hat{\Sigma}_{22}), \quad \hat{\Sigma}_{22} = \Sigma_{33} - \Sigma_{34}{\Sigma_{44}}^{-1}\Sigma_{43} \end{aligned}$$

are Gaussian hints inferred from p_1 and p_2, respectively. Let $T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$ be an admissible matrix for the joint system $h_1 \oplus h_2$. Applying T yields the equivalent system $(\bar{A}, \bar{z}, \bar{\Sigma}) \cong h_1 \oplus h_2$

$$\bar{A} = \begin{pmatrix} T_{11}A_{11} + T_{12}A_{21} \\ 0 \end{pmatrix}, \quad \bar{z} = \begin{pmatrix} T_{11}(z_{11} + \Sigma_{12}{\Sigma_{22}}^{-1} z_{12}) + T_{12}(z_{21} + \Sigma_{34}{\Sigma_{44}}^{-1} z_{22}) \\ T_{21}(z_{11} + \Sigma_{12}{\Sigma_{22}}^{-1} z_{12}) + T_{22}(z_{21} + \Sigma_{34}{\Sigma_{44}}^{-1} z_{22}) \end{pmatrix}$$

$$\bar{\Sigma} = \begin{pmatrix} T_{11}\hat{\Sigma}_{11}T'_{11} + T_{12}\hat{\Sigma}_{22}T'_{12} & T_{11}\hat{\Sigma}_{11}T'_{21} + T_{12}\hat{\Sigma}_{22}T'_{22} \\ T_{21}\hat{\Sigma}_{11}T'_{11} + T_{22}\hat{\Sigma}_{22}T'_{12} & T_{21}\hat{\Sigma}_{11}T'_{21} + T_{22}\hat{\Sigma}_{22}T'_{22}. \end{pmatrix}$$

On the other hand, consider now the left-hand side of equation (6.76). The matrix

$$\tilde{T} = \begin{pmatrix} T_{11} & 0 & T_{12} & 0 \\ T_{21} & 0 & T_{22} & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}.$$

is an admissible matrix for the joint system $p_1 \oplus p_2$. Indeed, applying $\tilde{T}$ yields the equivalent system $(\tilde{A}, \tilde{z}, \tilde{\Sigma})$ given by

$$\tilde{A} = \begin{pmatrix} T_{11}A_{11} + T_{12}A_{21} \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \tilde{z} = \begin{pmatrix} T_{11}z_{11} + T_{12}z_{21} \\ T_{21}z_{11} + T_{22}z_{21} \\ z_{12} \\ z_{22} \end{pmatrix}$$

$$\tilde{\Sigma} = \begin{pmatrix} T_{11}\Sigma_{11}T'_{11} + T_{12}\Sigma_{33}T'_{12} & T_{11}\Sigma_{11}T'_{21} + T_{12}\Sigma_{33}T'_{22} & T_{11}\Sigma_{12} & T_{12}\Sigma_{34} \\ T_{21}\Sigma_{11}T'_{11} + T_{22}\Sigma_{33}T'_{22} & T_{21}\Sigma_{11}T'_{21} + T_{22}\Sigma_{33}T'_{22} & T_{21}\Sigma_{12} & T_{22}\Sigma_{34} \\ \Sigma_{21}T'_{11} & \Sigma_{21}T'_{21} & \Sigma_{22} & 0 \\ \Sigma_{43}T'_{12} & \Sigma_{43}T'_{22} & 0 & \Sigma_{44} \end{pmatrix}$$

In order to prove the lemma, it suffices to show that $(\bar{A}, \bar{z}, \bar{\Sigma})$ and $(\tilde{A}, \tilde{z}, \tilde{\Sigma})$ induce the same Gaussian hint. Since the assumption-based inference in these two systems consists in conditioning to the admissible assumptions and since conditioning can be done step-wise, it suffices to show that partial conditioning to $z_{12} = 0$ and $z_{22} = 0$ in $(\tilde{A}, \tilde{z}, \tilde{\Sigma})$ produces $(\bar{A}, \bar{z}, \bar{\Sigma})$. Indeed, using Lemma B.1 for conditioning, the first parts of the observation are equal,

$$\begin{aligned} & T_{11}z_{11} + T_{12}z_{21} - T_{11}\Sigma_{12}{\Sigma_{22}}^{-1}z_{12} - T_{12}\Sigma_{34}{\Sigma_{44}}^{-1}z_{22} \\ =& T_{11}(z_{11} + \Sigma_{12}{\Sigma_{22}}^{-1}z_{12}) + T_{12}(z_{21} + \Sigma_{34}{\Sigma_{44}}^{-1}z_{22}), \end{aligned}$$

and the first diagonal block of the variance-covariance matrices are equal,

$$\begin{aligned} & T_{11}\Sigma_{11}T'_{11} + T_{12}\Sigma_{33}T'_{12} - T_{11}\Sigma_{12}{\Sigma_{22}}^{-1}\Sigma_{21}T'_{11} - T_{12}\Sigma_{34}{\Sigma_{44}}^{-1}\Sigma_{43}T'_{12} \\ =& T_{11}(\Sigma_{11} - \Sigma_{12}{\Sigma_{22}}^{-1}\Sigma_{21}) + T_{12}(\Sigma_{33} - \Sigma_{34}{\Sigma_{44}}^{-1}\Sigma_{43}) \\ =& T_{11}\hat{\Sigma}_{11}T'_{11} + T_{12}\hat{\Sigma}_{22}T'_{12}. \end{aligned}$$

The equality of the other components of the observation vector and the variance-covariance matrix can be proved in the same way. □

COROLLARY **6.31**. *Let $p_1, \tilde{p}_1$ and $p_2, \tilde{p}_2$ be equivalent Gaussian linear systems, respectively, i.e.*

$$\mathcal{H}(p_1) = \mathcal{H}(\tilde{p}_1), \quad \mathcal{H}(p_2) = \mathcal{H}(\tilde{p}_2).$$

Then, the joint Gaussian linear systems $p_1 \oplus p_2$ and $\tilde{p}_1 \oplus \tilde{p}_2$ induce the equivalent Gaussian hints, i.e.

$$\mathcal{H}(p_1 \oplus p_2) = \mathcal{H}(\tilde{p}_1 \oplus \tilde{p}_2).$$

⊘

PROOF. Let $h_1, h_2, \tilde{h}_1, \tilde{h}_2$ be Gaussian hints in $\mathcal{H}(p_1)$, $\mathcal{H}(p_2)$, $\mathcal{H}(\tilde{p}_1)$, and $\mathcal{H}(\tilde{p}_2)$, respectively. Let B_1 and B_2 be the regular matrices that establish the equivalences $h_1 \cong \tilde{h}_1$ and $h_2 \cong \tilde{h}_2$. Then, in light of Lemma 6.22, the matrix $B = \begin{pmatrix} B_1 & \\ & B_2 \end{pmatrix}$ establishes the equivalence $h_1 \oplus h_2 \cong \tilde{h}_1 \oplus \tilde{h}_2$, i.e. $\mathcal{H}(p_1) \oplus \mathcal{H}(p_2) = \mathcal{H}(\tilde{p}_1) \oplus \mathcal{H}(\tilde{p}_2)$. Therefore, using Lemma 6.30,

$$\mathcal{H}(p_1 \oplus p_2) = \mathcal{H}(h_1 \oplus h_2) = \mathcal{H}(\tilde{h}_1 \oplus \tilde{h}_2) = \mathcal{H}(\tilde{p}_1 \oplus \tilde{p}_2).$$

□

In light of Corollary 6.31, the combination of Gaussian hints can be defined up to equivalence.

DEFINITION **6.32**. *The combination $\otimes : \mathcal{H} \times \mathcal{H} \to \mathcal{H}, (h_1, h_2) \mapsto h_1 \otimes h_2$ of Gaussian hints can be defined up to equivalence. by the Gaussian hint inferred from the joint Gaussian linear system of h_1 and h_2, i.e.*

$$h_1 \otimes h_2 \in \mathcal{H}(h_1 \oplus h_2). \tag{6.77}$$

or, equivalently,

$$h_1 \otimes h_2 \cong h_1 \oplus h_2, \tag{6.78}$$

or also

$$\mathcal{H}(h_1 \otimes h_2) = \mathcal{H}(h_1 \oplus h_2). \tag{6.79}$$

⊘

As the joint design matrix of two Gaussian hints $h_1, h_2 \in \mathcal{H} \subseteq \mathfrak{L}$ is need not have full row rank, the two operations $\oplus$ on Gaussian linear systems and $\otimes$ on Gaussian hints only have to be distinguished. If $h_1 \oplus h_2$ is a Gaussian hint, then of course $\mathcal{H} \ni h_1 \otimes h_2 \cong h_1 \oplus h_2 \in \mathfrak{L}$.

6.6 Valuation Algebra of Gaussian Hints

The following to observations summarise the operations defined on Gaussian hints.

- Equivalent hints are related by a regular matrix. The regular matrices of the same dimension form a group.

- Combination and marginalisation have been defined only up to equivalence, i.e. there is a family of combination and marginalisation operators compatible with $\cong$.

Of course, it would be possible to fix a particular algorithm of the operations and then to derive a quotient valuation algebra as in Section 2.4. However, since equivalent Gaussian hints represent the same information, the choice of a canonical representation and of a canonical algorithm is in general not important. However, in order to keep the notation simple, representatives will be used for their equivalence classes; the abbreviations of Table 6.1 will be used.

representative	for
$\mathcal{H}$	$\mathcal{H}/\cong$
h	$\mathcal{H}(h)$
$h_1 =_{id} h_2$	$h_1 = h_2$
$h_1 = h_2$	$h_1 \cong h_2$
$h_1 \otimes h_2 = h$	$h_1 \otimes h_2 \cong h$
$h^{\downarrow s} = h_s$	$h^{\downarrow s} \cong h_s$

TABLE 6.1: Abbreviations for the operations on Gaussian hints defined only up to equivalence

The situation of Gaussian hints can be generalised as follows: Assume the elements of the equivalence classes $[\phi]_\theta$ in Φ are related by a group $G_{[\phi]_\theta}$ of transformations, i.e.

$$\psi \in [\phi]_\theta \iff \exists g \in G_{[\phi]_\theta} \text{ s.t. } \psi = g(\phi).$$

Further, assume that a family of combination and marginalisation operators is defined which are all compatible with θ. Without defining combination and marginalisation in Φ exactly, this results in a valuation algebra $(\Phi/\theta, D, d, \otimes, \mathcal{M}, \downarrow)$ directly. Such a (well-defined) valuation algebra will also be called a quotient valuation algebra although the underlying valuation algebra is not specified. Here, the simplified notation of Table 2.1 cannot be used. Instead, it is often more convenient to work on representatives of the equivalence classes without naming the congruence. The abbreviations of Table 6.2 will be used for both kinds of a quotient valuation algebra (i.e. induced or not).

abbreviation	for
$(\Phi, D, d, \otimes, \mathcal{M}, \downarrow, \theta)$	$(\Phi/\theta, D, d, \otimes, \mathcal{M}, \downarrow)$
$\phi =_{id} \psi$	$\phi = \psi$
$\phi = \psi$	$[\phi]_\theta = [\psi]_\theta$
$d(\phi)$	$d([\phi]_\theta)$
$\phi \otimes \psi$	$[\phi]_\theta \otimes [\psi]_\theta$
$\mathcal{M}(\phi)$	$\mathcal{M}([\phi]_\theta)$
$\psi = \phi^{\downarrow s}$	$\psi \in {[\phi]_\theta}^{\downarrow s}$

TABLE 6.2: Abbreviations for quotient valuation algebras working on representatives

In order to apply the local computation algorithms of Chapter 4 to Gaussian hints, it will now be shown that Gaussian hints with the operations defined on them form a valuation algebra.

THEOREM **6.33**. *Gaussian hints $(\mathcal{H}, D, d, \otimes, \downarrow)$ form a stable valuation algebra.* ⊘

PROOF. It has to be verified that the operations satisfy the axioms (A1)-(A7) imposed on a valuation algebra.

(A1) Combination of Gaussian hints $h_1, h_2, h_3 \in \mathcal{H}$ is associative since

$$\begin{aligned}\mathcal{H}((h_1 \otimes h_2) \oplus h_3) &= \mathcal{H}((h_1 \oplus h_2) \oplus (h_3)) \\ &= \mathcal{H}((h_1) \oplus (h_2 \oplus h_3)) = \mathcal{H}(h_1 \oplus (h_2 \otimes h_3))\end{aligned}$$

in light of Lemma 6.30.

In order to prove that the combination of Gaussian hints is commutative, let $h_1 = (A_1, z_1, K_1)$ and $h_2 = (A_2, z_2, K_2)$ be Gaussian hints with $A_1 \in \mathbb{R}(m_1, x)$, $A_2 \in \mathbb{R}(m_2, y)$. Define

$$(A_{12}, z_{12}, K_{12}) = h_1 \oplus h_2, \quad (A_{21}, z_{21}, K_{21}) = h_2 \oplus h_1.$$

Define

$$B = \begin{pmatrix} 0 & I_{m_2} \\ I_{m_1} & 0 \end{pmatrix},$$

which is regular since its inverse is

$$B^{-1} = \begin{pmatrix} 0 & I_{m_1} \\ I_{m_2} & 0 \end{pmatrix}.$$

Observe that

$$A_{21} = BA_{12}, \quad z_{21} = BA_{12}, \quad K_{21}B^{-1'}K_{12}B^{-1},$$

i.e. the equivalence of $h_1 \oplus h_2$ and $h_2 \oplus h_1$ is established by B. Hence, using Lemma 6.22, $h_1 \otimes h_2 = \mathcal{H}(h_1 \oplus h_2) = \mathcal{H}(h_2 \oplus h_1) = h_2 \otimes h_1$. This shows that combination is also commutative.

(A2) $d(h_1 \otimes h_2) = d(\mathcal{H}(h_1 \oplus h_2)) = d(h_1 \oplus h_2) = d(h_1) \cup d(h_2)$.

(A3) The marginalisation axiom follows by definition (6.60) of marginalisation.

(A4) Let $h = (A, z, K) \in \mathcal{H}$ be a Gaussian hint with $A \in \mathbb{R}(m, x)$, $z \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Further, let $s \subseteq t \subseteq x$ and let T_t and T_s be projection matrices for h to t and for $h^{\downarrow t}$ to s, respectively. Then, in light of the definition of marginalisation (6.60), it suffices to prove that $T = T_s T_t$ is a projection matrix for h to s. Partition

$$A = (A_1, A_2, A_3)$$

such that $A_1 \in \mathbb{R}(m, s)$, $A_2 \in \mathbb{R}(m, t - s)$, $A_3 \in \mathbb{R}(m, x - t)$. Then,

$$h^{\downarrow t} = ((T_t A_1, T_t A_2), T_t \mu, (T_t K^{-1} T_t')^{-1}),$$

and, using the transitivity axiom,

$$h^{\downarrow s} = h^{\downarrow t \downarrow s} = (T_s T_t A_1, T_s T_t \mu, (T_s T_t K^{-1} T_t' T_s')^{-1}),$$

Observe that

$$T_t A_3 = 0$$

and

$$T_s(T_t A_2) = 0.$$

Hence, on the one hand,

$$T_s T_t (A_2, A_3) = T_s (T_t A_2, 0) = (T_s(T_t A_2), T_s 0) = 0.$$

On the other hand, T_t having full row rank implies that $\mathrm{r}(T) = \mathrm{r}(T_s T_t) = \mathrm{r}(T_s)$ in light of Lemma 8.3.2 of (Harville, 1997; p.83), hence T has full row rank. It holds that

- $T_t \in \mathbb{R}(m - \mathrm{r}(A_3), m)$,
- $T_s \in \mathbb{R}(m - \mathrm{r}(A_3) - \mathrm{r}(T_t A_2), m - \mathrm{r}(A_3))$, and
- $\mathrm{r}(A_2, A_3) = \mathrm{r}(A_3) + \mathrm{r}(T_t A_2)$.

Thus,

$$\mathrm{r}(T) = \mathrm{r}(T_s) = m - \mathrm{r}(A_3) - \mathrm{r}(T_t A_2) = m - \mathrm{r}(A_2, A_3).$$

Therefore, $T = T_s T_t$ is indeed a projection matrix for h to s.

(A5) Let $h_1 = (A_1, z_1, K_1)$ and $h_2 = (A_2, z_2, K_2)$ be Gaussian hints, $A_1 \in \mathbb{R}(m_1, x)$, $z_1 \in \mathbb{R}^{m_1}$, $A_2 \in \mathbb{R}(m_2, y)$, $z_2 \in \mathbb{R}^{m_2}$, $K_1 \in \mathbb{R}(m_1, m_1)$ and $K_2 \in \mathbb{R}(m_2, m_2)$ both symmetric positive definite. Let $(A, z, K) = h_1 \oplus h_2$ be the joint Gaussian linear system and let s be a domain such that $x \subseteq s \subseteq x \cup y$.

Then, let $P \in \mathbb{R}(m_2 - k, m_2)$ be a projection matrix for A_2 to $s \cap y$, i.e. $PA_{22} = 0$ and $\mathrm{r}(A_{22}) = k$ for

$$A_2 = (A_{21}, A_{22}),$$

$A_{21} \in \mathbb{R}(m_2, y \cap s)$, $A_{22} \in \mathbb{R}(m_2, y - s)$. Then,

$$h_2{}^{\downarrow s \cap y} = (PA_{21}, Pz_2, (PK^{-1}P')^{-1}).$$

Furthermore, let

$$g = (\begin{pmatrix} A_1{}^{\uparrow s} \\ PA_{21}{}^{\Rightarrow s} \end{pmatrix}, \begin{pmatrix} z_1 \\ Pz_2 \end{pmatrix}, \begin{pmatrix} K_1 & \\ & (PK_2{}^{-1}P')^{-1} \end{pmatrix}).$$

On the one hand,

$$h_1 \oplus h_2{}^{\downarrow s \cap y} = g. \tag{6.80}$$

On the other hand, define $\tilde{P} \in \mathbb{R}(m_1 + m_2 - k, m_1 + m_2)$,

$$\tilde{P} = \begin{pmatrix} I_{m_1} & \\ & P \end{pmatrix}.$$

It is now shown that $\tilde{P}$ is a projection matrix for the Gaussian linear system $h_1 \oplus h_2$ to s. Indeed, on the on hand,

$$\tilde{P}\begin{pmatrix} A_1{}^{\uparrow s} & 0 \\ PA_{21}{}^{\Rightarrow s} & A_{22} \end{pmatrix} = \begin{pmatrix} A_1{}^{\uparrow s} & 0 \\ PA_{21}{}^{\Rightarrow s} & 0 \end{pmatrix}. \tag{6.81}$$

On the other hand, the block-diagonal matrix $\tilde{P}$ has rank

$$\mathrm{r}(\tilde{P}) = \mathrm{r}(I_{m_1}) + \mathrm{r}(P) = m_1 + (m_2 - k),$$

where

$$\mathrm{r}(\begin{pmatrix} 0_{m_1, y-s} \\ A_{22} \end{pmatrix}) = \mathrm{r}(A_{22}) = k.$$

Hence, $\tilde{P}$ is indeed a projection matrix for $h_1 \oplus h_2$ to s. It then follows from equations (6.80) and (6.81) that

$$h_1 \oplus h_2{}^{\downarrow s \cap y} = g = (h_1 \oplus h_2)^{\downarrow s}.$$

Hence, in light of Lemma 6.29,

$$(h_1 \otimes h_2)^{\downarrow s} = h_1 \otimes h_2{}^{\downarrow s \cap y}.$$

(A6) The domain axiom is also verified by the definition of marginalisation.

(A7) Finally, $e = (\diamond, \diamond, \diamond)$ is an identity element since joining e with any Gaussian hint h leads to the Gaussian linear system $h \oplus e = h = e \oplus h$. Hence, $h \otimes e = h = e \otimes h$. □

6.7 Precise Gaussian Hints and Gaussian Potentials

It is now shown that Gaussian hints are more general than Gaussian potentials: Gaussian potentials can be represented by precise Gaussian hints and the operations of valuation algebras in both representations are compatible. It is readily verified that the focal sets of a Gaussian hint are all singletons if and only if the design matrix is regular. Therefore, a Gaussian hint (A, z, K) is called precise if A is regular. Define $\mathcal{H}_0$ to be the set of all precise Gaussian hints,

$$\mathcal{H}_0 = \{(A, z, K) \in \mathcal{H} : \ A \text{ regular}\}. \tag{6.82}$$

This definition is sound since $h_1 \in \mathcal{H}_0$ and $h_1 = h_2$ imply that $h_2 \in \mathcal{H}_0$.

The following lemma shows that precise Gaussian hints have a *canonical representation.*

Lemma **6.34**. *Let $h_1 = (A_1, z_1, K_1)$, $h_2 = (A_2, z_2, K_2)$ be precise Gaussian hints on x. Let*

$$h_1' = (I_x, A_1{}^{-1} z_1, A_1' K_1 A_1) \quad \textit{and} \quad h_2' = (I_x, A_2{}^{-1} z_1, A_2' K_2 A_2).$$

Then,

$$h_1 = h'_1, \qquad h_2 = h'_2 \tag{6.83}$$

and

$$h_1 = h_2 \iff h'_1 = h'_2. \tag{6.84}$$

⊘

PROOF. Since h_1, h_2 are precise, it follows that A_1, A_2 are regular, thus invertible, ${A_1}^{-1}, {A_2}^{-1} \in \mathbb{R}(x, m)$. Hence, the equivalence of h_1 and h'_1 is established by ${A_1}^{-1}$, and, similarly, the equivalence of h_2 and h'_2 by ${A_2}^{-1}$. Thus, this proves the first claim. The second claim holds since I_x is the only matrix that could establish $h'_1 = h'_2$ ($\iff h_1 = h'_1 = h'_2 = h_2$). □

This relates precise Gaussian hints to Gaussian potentials by the mapping $p : \mathcal{H} \to \mathcal{G}$ defined by

$$p(A, z, K) = (A^{-1}z, A'KA). \tag{6.85}$$

The mapping $p : \mathcal{H}_0 \to \mathcal{G}$, $(A, z, K) \mapsto p(A, z, K)$ is well defined in light of Lemma 6.34. The following theorem is a reformulation of Lemma 6.34 in terms of the mapping p; it shows that the equivalence classes of precise Gaussian hints mapping to the same Gaussian potential coincide with the classes of equivalent precise hints.

THEOREM **6.35**. *For $h_1, h_2 \in \mathcal{H}_0$*

$$h_1 = h_2 \iff p(h_1) = p(h_2). \tag{6.86}$$

⊘

PROOF. The "if" part follows from Lemma 6.34. Let $h_1 = (A_1, z_1, K_1)$ and $h_2 = (A_2, z_2, K_2)$. Let $(\mu, K) = p(h_1) = p(h_2)$ and

$$h_0 = (I_{m,x}, \mu, K)$$

for $m = |x|$. Then, the equivalence $h_1 = h_0$ is established by ${A_1}^{-1}$ and the equivalence $h_0 = h_2$ by A_2, hence $h_1 = h_2$ shows the "only if" part. □

Furthermore, p is compatible with the operations of valuation algebras; more formally, it is now shown that p is a *surjective valuation algebra homomorphism*, i.e. that p is

1. *surjective*: for every $\phi \in \mathcal{G}$ there is a $h \in \mathcal{H}_0$ such that $p(h) = \phi$,
2. *compatible with combination*: $p(h_1 \otimes h_2) = p(h_1) \otimes p(h_2)$ for $h_1, h_2 \in \mathcal{H}_0$, and
3. *compatible with marginalisation*: $p(h^{\downarrow s}) = p(h)^{\downarrow s}$ for $h \in \mathcal{H}_0$, $s \subseteq d(h)$.

It has to be remarked that these three conditions and surjectivity imply that the identity element of $\mathcal{H}_0$ is mapped to the identity element of $\mathcal{G}$.

THEOREM **6.36**. *$p : \mathcal{H}_0 \to \mathcal{G}$ is a surjective valuation algebra homomorphism.* $\oslash$

PROOF. For $\phi = (\mu, K) \in \mathcal{G}$, $x = d(\phi)$, $m = |x|$, it holds that

$$h = (I_{m,x}, I_{m,x}\mu, I_{m,x}KI_{x,m}) \in \mathcal{H}_0$$

and

$$p(h) = (I_{x,m}I_{m,x}\mu, I_{x,m}I_{m,x}KI_{x,m}I_{m,x}) = (\mu, K) = \phi,$$

hence p is surjective.

Let h_1, h_2 be precise Gaussian hints on $x_1, x_2 \in D$ and let h_1' and h_2' be their canonical representations. Let $h_1' = (I_{m_1,x}, z_1, K_1)$ and $h_2' = (I_{m_2,x_2}, z_2, K_2)$ for $m_1 = |x_1|$, $m_2 = |x_2|$. Let $m = m_1 + m_2$ and $x = x_1 \cup x_2$. Notice that $|x_1| + |x_2| = m \geq |x| = |x_1| \cup |x_2|$,

$$|x_1 - x_2| = m - m_2, \quad |x_1 \cap x_2| = m_1 + m_2 - m, \quad |x_2 - x_1| = m - m_1$$

and

$$m_1 + m_2 = (m - m_2) + 2 \cdot (m_1 + m_2 - m) + (m - m_1).$$

Define $n = |x| = |x_1 \cup x_2|$,

$$l_1 = |x_1 - x_2|, \quad l_2 = |x_2 - x_1|, \quad \text{and} \quad l_{12} = |x_1 \cap x_2|.$$

Partition

$$z_1 = \begin{pmatrix} z_{1.1} \\ z_{1.2} \end{pmatrix}, \quad z_2 = \begin{pmatrix} z_{2.1} \\ z_{2.2} \end{pmatrix}$$

such that $z_{1.1} \in \mathbb{R}^{l_1}$, $z_{1.2}, z_{2.1} \in \mathbb{R}^{l_{12}}$, $z_{2.2} \in \mathbb{R}^{l_2}$, and

$$K_1 = \begin{pmatrix} K_{1.11} & K_{1.12} \\ K_{1.21} & K_{1.22} \end{pmatrix}, \quad K_2 = \begin{pmatrix} K_{2.11} & K_{2.12} \\ K_{2.21} & K_{2.22} \end{pmatrix}$$

such that $K_{1.11} \in \mathbb{R}(l_1, l_1)$, $K_{1.12} \in \mathbb{R}(l_1, l_{12})$, $K_{1.21} \in \mathbb{R}(l_{12}, l_1)$, $K_{1.22} \in \mathbb{R}(l_{12}, l_{12})$, $K_{2.11} \in \mathbb{R}(l_{12}, l_{12})$, $K_{2.12} \in \mathbb{R}(l_{12}, l_2)$, $K_{2.21} \in \mathbb{R}(l_2, l_{12})$ and $K_{2.22} \in \mathbb{R}(l_2, l_2)$. Then, $h_1' \otimes h_2'$ is the hint inferred from the Gaussian linear system (A, z, K) where $A \in \mathbb{R}(m_1 + m_2, x)$, $z \in \mathbb{R}(m_1 + m_2)$ and $K \in \mathbb{R}(m_1 + m_2, m_1 + m_2)$ are given by

$$A = \begin{pmatrix} I_{l_1,x_1-x_2} & 0_{l_1,x_1\cap x_2} & 0_{l_1,x_2-x_1} \\ 0_{l_{12},x_1-x_2} & I_{l_{12},x_1\cap x_2} & 0_{l_{12},x_2-x_1} \\ 0_{l_{12},x_1-x_2} & I_{l_{12},x_1\cap x_2} & 0_{l_{12},x_2-x_1} \\ 0_{l_2,x_1-x_2} & 0_{l_2,x_1\cap x_2} & I_{l_2,x_2-x_1} \end{pmatrix},$$

$$z = \begin{pmatrix} z_{1.1} \\ z_{1.2} \\ z_{2.1} \\ z_{2.2} \end{pmatrix}$$

and

$$K = \begin{pmatrix} \begin{matrix} K_{1.11} & K_{1.12} \\ K_{1.21} & K_{1.22} \end{matrix} & 0_{m_1,m_2} \\ 0_{m_2,m_1} & \begin{matrix} K_{2.11} & K_{2.12} \\ K_{2.21} & K_{2.22} \end{matrix} \end{pmatrix}.$$

Then, in light of Theorem 6.23,

$$h'_1 \otimes h'_2 = (I_{n,x}, (A'KA)^{-1}A'Kz, A'KA),$$

where

$$A'KA = \begin{pmatrix} K_{1.11} & K_{1.12} & 0 \\ K_{1.21} & K_{1.22} + K_{2.11} & K_{2.12} \\ 0 & K_{2.21} & K_{2.22} \end{pmatrix}.$$

and

$$A'Kz = \begin{pmatrix} K_{1.11}z_{1.1} + K_{1.12}z_{1.2} \\ K_{1.21}z_{1.1} + K_{1.22}z_{1.2} + K_{2.11}z_{2.1} + K_{2.12}z_{2.2} \\ K_{2.21}z_{2.1} + K_{2.22}z_{2.2} \end{pmatrix}$$

Hence,

$$p(h'_1 \otimes h'_2) = (I_{x,n}(A'KA)^{-1}A'Kz, I_{x,n}A'KAI_{n,x}).$$

On the other hand,

$$\begin{aligned} p(h'_1) &= (I_{x_1,m_1}z_1, I_{x_1,m_1}K_1I_{m_1,x_1}), \\ p(h'_2) &= (I_{x_2,m_2}z_2, I_{x_2,m_2}K_2I_{m_2,x_2}). \end{aligned}$$

Define $(\tilde{\mu}, \tilde{K}) = p(h'_1) \otimes p(h'_2)$, where

$$\tilde{K} = (I_{x_1,m_1}K_1I_{m_1,x_1})^{\uparrow x} + (I_{x_2,m_2}K_2I_{m_2,x_2})^{\uparrow x}$$

and

$$\begin{aligned} \tilde{\mu} &= \tilde{K}^{-1}\left((I_{x_1,m_1}K_1I_{m_1,x_1})^{\uparrow x}(I_{m_1,x_1}z_1)^{\uparrow x} + (I_{x_2,m_2}K_2I_{m_2,x_2})^{\uparrow x}(I_{m_2,x_2}z_2)^{\uparrow x}\right) \\ &= \tilde{K}^{-1}\left((I_{x_1,m_1}K_1z_1)^{\uparrow x} + (I_{x_2,m_2}K_2z_2)^{\uparrow x}\right). \end{aligned}$$

Then, observing that $A'KA = \tilde{K}$ and $(A'KA)^{-1}A'Kz = \tilde{K}A'Kz = \tilde{\mu}$, it follows that

$$p(h'_1 \otimes h'_2) = p(h'_1) \otimes p(h'_2).$$

Since $h_1 = h'_1$ and $h_2 = h'_2$ and thus $h_1 \otimes h_2 = h'_1 \otimes h'_2$, it follows in light of Lemma 6.34 that

$$p(h_1) = p(h'_1), \quad p(h_2) = p(h'_2), \quad p(h_1 \otimes h_2) = p(h'_1 \otimes h'_2),$$

hence

$$p(h_1 \otimes h_2) = p(h'_1 \otimes h'_2) = p(h'_1) \otimes p(h'_2) = p(h_1) \otimes p(h_2),$$

so p is indeed compatible with combination.

Let h be a precise Gaussian hint on $x \in D$, let $h_0 = (I_x, \mu, K)$ be its canonical version, and let $s \subseteq x$. Since marginalisation of Gaussian hints is compatible with the equivalence of Gaussian hints and in light of Lemma 6.34,

$$\begin{aligned} p(h^{\downarrow s}) &= p({h_0}^{\downarrow s}) = p(PI_x, P\mu, (PK^{-1}P')^{-1}) \\ &= p(I_s, \mu^{\downarrow s}, (K^{-1\downarrow s})^{-1}) = (\mu^{\downarrow s}, (K^{-1\downarrow s})^{-1}) = p(h)^{\downarrow s} \end{aligned}$$

where $P \in \mathbb{R}(s, x)$,

$$P(S, X) = \begin{cases} 0 & \text{if } S \neq X \\ 1 & \text{if } S = X, \end{cases}$$

is an elimination matrix of full row rank $|s|$. Hence, p is compatible with marginalisation. □

It has been shown that Gaussian potentials can be seen as canonical representations of precise Gaussian hints and that the operations on them are compatible. This is summarised in Figure 6.7.

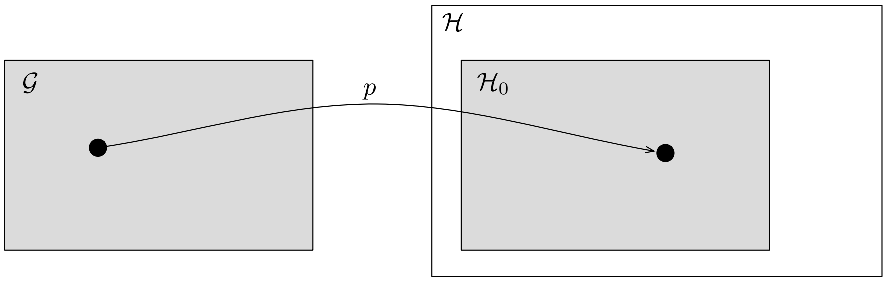


Figure 6.7: Precise Gaussian hints form a subalgebra of Gaussian hints. Gaussian potentials and precise Gaussian hints form isomorphic valuation algebras.

Chapter Synopsis & Discussion

Assumption-based reasoning and the theory of hints provide a framework for statistical inference (Kohlas and Monney, 1995). Assumption-based inference on Gaussian linear systems leads to Gaussian hints, as discussed in depth by (Monney, 2003; Kohlas and Monney, 2008). An implementation of an alternative inference algorithm based on singular-value decomposition was developed in (Eichenberger, 2004). Gaussian hints form a valuation algebra, where marginalisation and combination have a geometric interpretation: Marginalisation essentially corresponds to the projection and combination to the intersection of focal sets. Gaussian potentials correspond to precise Gaussian hints.

7
Gaussian Hints and Conditional Gaussian Densities

A conditional Gaussian density represents a family of Gaussian densities on the same set x of variables. These densities are indexed by the values of a set z of variables (disjoint from x). In this setting, probabilities can only be asserted if the value of each variable of z is known. However, a conditional Gaussian density induces a Gaussian linear system of regression equations from the tail on the head variables. The induced system is already a Gaussian hint. Thereby, *different* conditional Gaussian densities (with different head and tail variables) may be linked to the *same* Gaussian hint (up to equivalence).

Chapter Outline

It will be shown that

- every conditional Gaussian density induces a Gaussian hint (in Section 7.1),
- every Gaussian hint represents a conditional Gaussian density (in Section 7.2), and
- how different conditional Gaussian densities inducing the same Gaussian hint are related (in Section 7.3).

The pivotal role of Gaussian hints is depicted in Figure 7.1: Different conditional Gaussian densities $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ may be related to the same Gaussian hint h, and a conditional Gaussian density may be represented by different conditional Gaussian potentials. Further, combination of Gaussian hints and elimination of variables in Gaussian hints will be carried over to conditional Gaussian densities (in Sections 7.4 and 7.5).

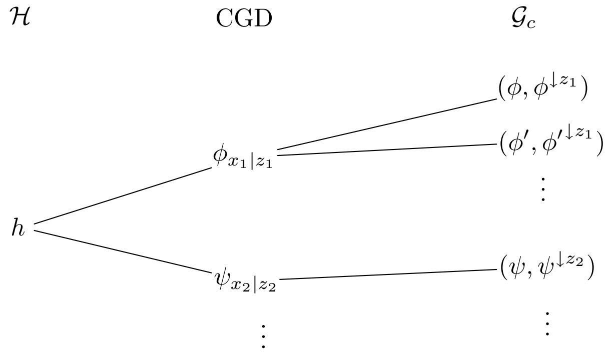


FIGURE 7.1: Different conditional Gaussian densities may be related to the same Gaussian hint h, and different conditional Gaussian potentials may represent the same conditional Gaussian density.

7.1 From Conditional Gaussian Densities to Gaussian Hints

As sketched in Section 7.1, a conditional $\phi_{x|z}$ of a Gaussian density $\phi_{\mu,K}$ induces the Gaussian hint

$$\mathcal{H}(\phi_{x|z}) = (A, \bar{\mu}, \bar{K}) \tag{7.1}$$

on $x \cup z$ where

$$A = \left(I_{m,x}, I_{m,x}{K^{\downarrow x}}^{-1}K^{\downarrow x,z}\right), \quad \bar{\mu} = I_{m,x}(\mu^{\downarrow x} + {K^{\downarrow x}}^{-1}K^{\downarrow x,z}\mu^{\downarrow z}) \tag{7.2}$$

and

$$\bar{K} = I_{m,x}K^{\downarrow x}I'_{m,x}. \tag{7.3}$$

It has to be verified that another indexation $J_{m,x}$ yields the same Gaussian hint (up to equivalence). Indeed, the indexations $I_{m,x}$ and $J_{m,x}$ are related by a regular permutation matrix $B \in \mathbb{R}(m,m)$ such that $J_{m,x} = BI_{m,x}$. It also holds that

$$B\bar{\mu} = BI_{m,x}(\mu^{\downarrow x} + {K^{\downarrow x}}^{-1}K^{\downarrow x,z}\mu^{\downarrow z}) = J_{m,x}(\mu^{\downarrow x} + {K^{\downarrow x}}^{-1}K^{\downarrow x,z}\mu^{\downarrow z})$$

and

$$B\bar{K}B' = I_{m,x}K^{\downarrow x}I'_{m,x}B' = J_{m,x}K^{\downarrow x}J'_{m,x} = {J'_{m,x}}^{-1}K^{\downarrow x}{J_{m,x}}^{-1}.$$

Recalling the conventions of Tables 5.1 and 6.1, $\mathcal{H}$ is a mapping

$$\mathcal{H} : \mathcal{G}_c \to \mathcal{H}. \tag{7.4}$$

In order to simplify the notation, the following abbreviation will be used for (7.1):

$$\mathcal{H}(\phi_{x|z}) = \left[\left(I_x, {K^{\downarrow x}}^{-1}K^{\downarrow x,z}\right), \mu^{\downarrow x} + {K^{\downarrow x}}^{-1}K^{\downarrow x,z}\mu^{\downarrow z}, K^{\downarrow x}\right]. \tag{7.5}$$

Here, the rows of the design matrix are indexed by the head variables of $\phi_{x|z}$. Let $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ be two conditional Gaussian densities, and let

$$\mathcal{H}(\phi_{x_1|z_1}) = (A_1, z_1, K_1), \quad \mathcal{H}(\psi_{x_2|z_2}) = (A_2, z_2, K_2),$$

where $A_1 \in \mathbb{R}(x, x \cup z_1)$ and $A_2 \in \mathbb{R}(\tilde{x}, \tilde{x} \cup z_2)$. In this notation, the two conditional Gaussian densities $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ induce the same Gaussian hint (up to equivalence) if and only there is a regular matrix $T \in \mathbb{R}(x_2, x_1)$ such that

$$A_2 = TA_1, \quad z_2 = Tz_1, \quad K_2 = T^{-1\prime}K_1T^{-1}.$$

In summary, the mapping $\mathcal{H}$ associates a Gaussian hint to every conditional Gaussian density.

7.2 From Gaussian Hints to Conditional Gaussian Densities

The following lemma shows that every Gaussian hint represents at least one conditional Gaussian density. The proof is constructive.

Lemma **7.1**. *Let $h = (A, \mu, K) \in \mathcal{H}$ be a Gaussian hint on $y = d(h)$, $A \in \mathbb{R}(m, y)$. Then, there is a subset $x \subseteq y$ of cardinality $|x| = m$ and a Gaussian potential $\phi \in \mathcal{G}$ of domain $d(\phi) = y$ such that*

$$\mathcal{H}(\phi_{x|z}) = h$$

where $z = y - x$. Let $x \subseteq y$ of cardinality $|x| = m$ such that the submatrix $A_1 \in \mathbb{R}(m, x)$ of $A = (A_1, A_2)$ is regular. Then, there is a Gaussian potential ϕ of domain $d(\phi) = y$ such that

$$\mathcal{H}(\phi_{x|z}) = h.$$

⊘

Proof. In light of Theorem 4.4.10 of (Harville, 1997; p.39), there is a subset $x \subseteq y$ of cardinality $|x| = m$ such that A_1 has full column rank m and thus is regular. Let x be any such subset $x \subseteq y$ of cardinality $|x| = m$ such that A_1 has full column rank m. Then, transformation by the regular matrix $T = A_1{}^{-1} \in \mathbb{R}(x, m)$ yields an equivalent representation $h = ((I_x, B), \tilde{\mu}, \tilde{K})$ where

$$B = TA^{\downarrow z}, \quad \tilde{\mu} = T\mu, \quad \tilde{K} = T^{-1\prime}KT^{-1}.$$

Define

$$\bar{\mu} = \begin{pmatrix} \tilde{\mu} \\ 0_z \end{pmatrix}, \qquad \bar{K} = \begin{pmatrix} \tilde{K} & \tilde{K}B \\ B'\tilde{K} & I_z + B'KB \end{pmatrix}.$$

By Lemma A.7, $\bar{K}$ is a symmetric and positive definite matrix. Hence, $\phi = (\bar{\mu}, \bar{K})$ is a Gaussian potential such that

$$\mathcal{H}(\phi_{x|z}) = \left((I_x, \tilde{K}^{-1}\tilde{K}B), \tilde{\mu} + B \cdot 0_z, \tilde{K}\right) = \left((I_x, B), \tilde{\mu}, \tilde{K}\right) = h.$$

This concludes the proof. □

This lemma not only shows that every Gaussian hint is induced by a conditional Gaussian density; it also shows that *different* conditional Gaussian densities (with different head and tail) may induce the same Gaussian hint: Whenever the submatrix corresponding to variables $x \subseteq d(h)$ is regular, there is a conditional Gaussian density with head x inducing the Gaussian hint. The lemma also shows that, if the design matrix is regular, the conditional Gaussian potential constructed in the proof has empty tail and thus corresponds to a non-conditional Gaussian density. The following very simple example illustrates the construction given in the proof.

Example **7.2**. Let a Gaussian hint h on the variables X_1, X_2 be given by $A = \begin{pmatrix} 1 & 1 \end{pmatrix}$, $\mu = (1)$ and $K = (1)$. Let x_1 and x_2 be the singleton sets consisting of the corresponding variable only. Construct the Gaussian potential $\phi = (\bar{\mu}, \bar{K})$

$$\bar{\mu} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \bar{K} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Then, the conditional of ϕ for x_1 given x_2 induces h since $\mathcal{H}(\phi_{x_1|x_2}) = h$. Similarly, a Gaussian potential ϕ' could be constructed such that $\mathcal{H}(\phi'_{x_2|x_1}) = h$. ⊘

In summary, different conditional Gaussian densities (represented by a conditional Gaussian potential) may be related to the same Gaussian hint.

7.3 CGDs Related to the Same Gaussian Hint

Since a conditional Gaussian density $\phi_{x_1|z_1}$ induces the Gaussian hint $\mathcal{H}(\phi_{x_1|z_1})$, the sets of conditional Gaussian densities related to the same Gaussian hint form equivalence classes of the equivalence relation $\cong$ defined by

$$\phi_{x_1|z_1} \cong \psi_{x_2|z_2} \iff \mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}). \tag{7.6}$$

It is now shown how the equivalence relation $\cong$ in conditional Gaussian densities induced by Gaussian hints can be defined solely in terms of conditional Gaussian densities without reference to Gaussian hints.

Lemma **7.3**. *Let*

$$\phi = (\mu_1, K_1), \quad \psi = (\mu_2, K_2)$$

be Gaussian potentials such that

$$\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}).$$

Then, for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$

$$\phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) = c \cdot \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2})$$

for some constant $c > 0$ *not depending on* $\mathbf{u}$. ⊘

Proof. By definition,

$$\mathcal{H}(\phi_{x_1|z_1}) = \left(A_1, A_1\mu_1, K_1{}^{\downarrow x_1}\right)$$

and

$$\mathcal{H}(\psi_{x_2|z_2}) = \left(A_2, A_2\mu_2, K_2{}^{\downarrow x_2}\right)$$

where

$$A_1 = \left(I_{x_1}, {K_1^{\downarrow x_1}}^{-1} K_1^{\downarrow x_1,z_1} \right), \quad A_2 = \left(I_{x_2}, {K_2^{\downarrow x_2}}^{-1} K_2^{\downarrow x_2,z_2} \right).$$

Furthermore, $\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2})$ implies that there is a regular matrix $T \in \mathbb{R}(x_2, x_1)$ such that

$$A_2 = TA_1, \quad A_2\mu_2 = TA_1\mu_1$$

and

$$K_2^{\downarrow x_2} = T^{-1\prime} K_1^{\downarrow x_1} T^{-1}.$$

Then, for $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1}$,

$$\begin{aligned}
\phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) &= \phi_{0,K_1^{\downarrow x_1}}(A_1\mu_1 - A_1\mathbf{u}) \\
&= \phi_{0,K_1^{\downarrow x_1}}(T^{-1}T(A_1\mu_1 - A_1\mathbf{u})) \\
&= \sqrt{\frac{\det(K_1^{\downarrow x_1})}{(2\pi)^m}} e^{-\frac{1}{2}(TA_1(\mu_1 - \mathbf{u}))' T^{-1\prime} K_1^{\downarrow x_1} T^{-1} (TA_1(\mu_1 - \mathbf{u}))} \\
&= |\det(T)| \cdot \sqrt{\frac{\det(K_2^{\downarrow x_2})}{(2\pi)^m}} e^{-\frac{1}{2}(A_2\mu_2 - A_2\mathbf{u})' K_2^{\downarrow x_2} (A_2\mu_2 - A_2\mathbf{u})} \\
&= |\det(T)| \cdot \phi_{0,K_2^{\downarrow x_2}}(A_2\mu_2 - A_2\mathbf{u}) \\
&= |\det(T)| \cdot \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2})
\end{aligned}$$

where $m = |x_1 \cup z_1| = |x_2 \cup z_2|$ and since

$$\begin{aligned}
\sqrt{\det(K_1^{\downarrow x_1})} &= |\det(T)| \cdot |\det(T^{-1})| \cdot \sqrt{\det(K_1^{\downarrow x_1})} \\
&= |\det(T)| \cdot \sqrt{\det(T^{-1}) \det(K_1^{\downarrow x_1}) \det(T^{-1})} \\
&= |\det(T)| \cdot \sqrt{\det(T^{-1\prime}) \det(K_1^{\downarrow x_1}) \det(T^{-1})} \\
&= |\det(T)| \cdot \sqrt{\det(T^{-1\prime} K_1^{\downarrow x_1} T^{-1})} \\
&= |\det(T)| \cdot \sqrt{\det(K_2^{\downarrow x_2})}
\end{aligned}$$

in light of Theorem 13.3.7, Lemma 13.2.1 and Theorem 13.3.4 of (Harville, 1997; p.188;p.181;p.187). Then, $\det(T) \neq 0$ in light of Theorem 13.3.7 of (Harville, 1997; p.188), hence $c = |\det(T)| > 0$ proves the lemma. □

This shows that two conditional Gaussian densities which are related to the same Gaussian hint are (up to a constant factor) the same function of head and tail variables. The converse is also true, as shown by the following lemma.

Lemma **7.4**. *Let* $\phi, \psi \in \mathcal{G}$ *be Gaussian potentials on the same domain* $d(\phi) = x_1 \cup z_1 = x_2 \cup z_2 = d(\psi)$, $x_1 \cap z_1 = x_2 \cap z_2 = \emptyset$ *such that*

$$\phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) = c \cdot \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2}). \tag{7.7}$$

for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$ *and some positive constant* $c > 0$ *not depending on* $\mathbf{u}$. *Then,*

$$\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}).$$

⊘

Proof. Let $\phi = (\mu_1, K_1)$ and $\psi = (\mu_2, K_2)$ such that (7.7) holds, and define

$$A_1 = (I_{x_1}, {K_1^{\downarrow x_1}}^{-1} K_1^{\downarrow x_1, z_1}), \quad A_2 = (I_{x_2}, {K_2^{\downarrow x_2}}^{-1} K_2^{\downarrow x_2, z_2}).$$

Using this notation,

$$\mathcal{H}(\phi_{x_1|z_1}) = (A_1, A_1\mu_1, K_1^{\downarrow x_1}), \quad \mathcal{H}(\psi_{x_2|z_2}) = (A_2, A_2\mu_2, K_2^{\downarrow x_2}).$$

Clearly, $|x_1| = |x_2|$. Let $m = |x_1|$. In light of Lemma A.3, there is a regular matrix $T \in \mathbb{R}(x_2, x_1)$ such that

$$K_2^{\downarrow x_2} = {T^{-1}}' K_1^{\downarrow x_1} T^{-1}. \tag{7.8}$$

Then, using the definition (5.1) of conditional Gaussian densities, it holds for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$ that

$$\begin{aligned}
\phi_{0,K_1^{\downarrow x_1}}(A_1\mu_1 - A_1\mathbf{u}) = \phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) \quad &= \quad c \cdot \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2}) \\
&= c \cdot \phi_{0,K_2^{\downarrow x_2}}(A_2\mu_2 - A_2\mathbf{u}) \\
&= c \cdot c_2 \cdot e^{-\frac{1}{2}(A_2(\mu_2 - \mathbf{u}))'({T^{-1}}' K_1^{\downarrow x_1} T^{-1})(A_2(\mu_2 - \mathbf{u}))} \\
&= c \cdot \left|\det(T^{-1})\right| \cdot c_1 \cdot e^{-\frac{1}{2}(T^{-1}A_2(\mu_2 - \mathbf{u}))' K_1^{\downarrow x_1} (T^{-1}A_2(\mu_2 - \mathbf{u}))} \\
&= c \cdot \left|\det(T^{-1})\right| \cdot \phi_{0,K_1^{\downarrow x_1}}(T^{-1}A_2(\mu_2 - \mathbf{u})),
\end{aligned} \tag{7.9}$$

where

$$c_1 = \sqrt{\frac{\det(K_1^{\downarrow x_1})}{(2\pi)^m}}, \quad c_2 = \sqrt{\frac{\det(K_2^{\downarrow x_2})}{(2\pi)^m}} = \left|\det(T^{-1})\right| \cdot c_1$$

since

$$\begin{aligned}
\sqrt{\det(K_2^{\downarrow x_2})} = \sqrt{\det({T^{-1}}' K_1^{\downarrow x_1} T^{-1})} \quad &= \quad \sqrt{\det({T^{-1}}') \cdot \det(K_1^{\downarrow x_1}) \cdot \det(T^{-1})} \\
= \sqrt{\det(T^{-1}) \cdot \det(K_1^{\downarrow x_1}) \cdot \det(T^{-1})} \quad &= \quad \left|\det(T^{-1})\right| \sqrt{\det(K_1^{\downarrow x_1})}
\end{aligned}$$

in light of Theorem 13.3.4 and Lemma 13.2.1 of (Harville, 1997; p.187;p.181). In light of (7.9), there is a $\mathbf{u}$ at which the maximum is realised. More precisely, since A_1 and A_2 have full row rank, there is a $\mathbf{u}_0$ such that

$$A_1\mu_1 - A_1\mathbf{u}_0 = 0 = T^{-1}A_2(\mu_2 - \mathbf{u}_0).$$

Then,

$$\begin{aligned}
\sqrt{\frac{\det(K_1^{\downarrow x_1})}{(2\pi)^m}} &= \phi_{0,K_1^{\downarrow x_1}}(A_1\mu_1 - A_1\mathbf{u}_0) \\
&= c \cdot \left|\det(T^{-1})\right| \cdot \phi_{0,K_1^{\downarrow x_1}}(T^{-1}A_2(\mu_2 - \mathbf{u}_0)) \\
&= c \cdot \left|\det(T^{-1})\right| \cdot \sqrt{\frac{\det(K_1^{\downarrow x_1})}{(2\pi)^m}}.
\end{aligned}$$

Hence, $c \cdot \left|\det(T^{-1})\right| = 1$. Therefore,

$$\phi_{0,K_1^{\downarrow x_1}}(A_1(\mu_1 - \mathbf{u})) = \phi_{0,K_2^{\downarrow x_2}}(T^{-1}A_2(\mu_2 - \mathbf{u}))$$

for all $\mathbf{u}$, i.e.

$$c_1 \cdot e^{-\frac{1}{2}(A_1(\mu_1-\mathbf{u}))' K_1{}^{\downarrow x_1}(A_1(\mu_1-\mathbf{u}))} = c_1 \cdot e^{-\frac{1}{2}(T^{-1}A_2(\mu_1-\mathbf{u}))' K_1{}^{\downarrow x_1}(T^{-1}A_2(\mu_2-\mathbf{u}))}$$

for all $\mathbf{u}$. Hence, multiplying both sides by $\frac{1}{c_1}$, taking the logarithm to the natural basis [observing that the exponential function e is strictly monotone] and multiplying by -2 yields

$$(\mu_1 - \mathbf{u})' A_1' K_1{}^{\downarrow x_1} A_1(\mu_1 - \mathbf{u}) \quad = \quad (\mu_2 - \mathbf{u})' A_2' T^{-1'} K_1{}^{\downarrow x_1} T^{-1} A_2(\mu_2 - \mathbf{u}) \quad (7.10)$$

for all $\mathbf{u}$. Since $K_1{}^{\downarrow x_1}$ is symmetric and positive definite, both sides of this equation equal 0 if and only

$$A_1(\mu_1 - \mathbf{u}) \quad = \quad 0 \quad = \quad T^{-1} A_2(\mu_2 - \mathbf{u}),$$

i.e. the solution sets

$$\Gamma_1 = \{\mathbf{u} : \ A_1 \mathbf{u} = A_1 \mu_1\}$$

and

$$\Gamma_2 = \{\mathbf{u} : \ T^{-1} A_2 \mathbf{u} = T^{-1} A_2 \mu_2\}$$

are equal, $\Gamma_1 = \Gamma_2$. Notice that

$$\Gamma_2 = \{\mathbf{u} : \ A_2 \mathbf{u} = A_2 \mu_2\}$$

in light of Lemma A.2. Then, using the same Lemma A.2 for $\Gamma_1 = \Gamma_2$, there is a regular matrix $\tilde{T} \in \mathbb{R}(x_1, x_1)$ such that

$$A_2 = \tilde{T} A_1 \quad \text{and} \quad A_2 \mu_2 = \tilde{T} A_1 \mu_1. \quad (7.11)$$

Plugging (7.11) into (7.10) yields that

$$\begin{aligned}(\mu_1 - \mathbf{u})' A_1' K_1{}^{\downarrow x_1} A_1(\mu_1 - \mathbf{u}) &= (A_2\mu_2 - A_2\mathbf{u})' T^{-1'} K_1{}^{\downarrow x_1} T^{-1}(A_2\mu_2 - A_2\mathbf{u}) \\ &= (\tilde{T}A_1\mu_1 - \tilde{T}A_1\mathbf{u})' K_2{}^{\downarrow x_2} (\tilde{T}A_1\mu_1 - \tilde{T}A_1\mathbf{u})' \\ &= (\mu_1 - \mathbf{u})' A_1' \tilde{T}' K_2{}^{\downarrow x_2} \tilde{T} A_1(\mu_1 - \mathbf{u})\end{aligned}$$

for all $\mathbf{u}$. Hence, in light of Lemma A.4,

$$A_1' K_1{}^{\downarrow x_1} A_1 = A_1' \tilde{T}' K_2{}^{\downarrow x_2} \tilde{T} A_1.$$

Observing that A_1 has full row rank and that, equivalently, A_1' has full column rank, applying Lemma A.1 yields that

$$K_1{}^{\downarrow x_1} = \tilde{T}' K_2{}^{\downarrow x_2} \tilde{T}.$$

Thus,

$$K_2{}^{\downarrow x_2} = \tilde{T}^{-1'} \tilde{T}' K_2{}^{\downarrow x_2} \tilde{T} \tilde{T}^{-1} = \tilde{T}^{-1'} K_1{}^{\downarrow x_1} \tilde{T}. \quad (7.12)$$

In light of (7.11) and (7.12), $\tilde{T}$ establishes $\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2})$. □

The results from Lemmata 7.3 and 7.4 are summarised in the following theorem.

THEOREM **7.5**. *For $\phi, \psi \in \mathcal{G}$, $d(\phi) = x_1 \cup z_1 = x_2 \cup z_2 = d(\psi)$, $x_1 \cap z_1 = x_2 \cap z_2 = \emptyset$,*

$$\phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) = c \cdot \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2}) \iff \mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}) \tag{7.13}$$

for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$ and some positive constant $c > 0$ not depending on $\mathbf{u}$. ⊘

In light of Theorem 7.5, two conditional Gaussian densities $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ are related to the same Gaussian hint if and only if they represent the same function up to a constant factor. The following theorem gives another criterion for two conditional Gaussian densities to be related to the same Gaussian hint.

THEOREM **7.6**. *For conditional Gaussian densities $\phi_{x_1|z_1} = (\phi, \phi^{\downarrow z_1})$ and $\psi_{x_2|z_2} = (\psi, \psi^{\downarrow z_2})$, $\phi, \psi \in \mathcal{G}$ with $d(\phi) = x_1 \cup z_1 = x_2 \cup z_2 = d(\psi)$, $x_1 \cap z_1 = x_2 \cap z_2 = \emptyset$,*

$$\phi \otimes \psi^{\downarrow z_2} = \phi^{\downarrow z_1} \otimes \psi \quad \iff \quad \mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}). \tag{7.14}$$

⊘

PROOF. According to Theorem 7.5, different conditional Gaussian densities $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ induce the same Gaussian hint if and only if they represent the same function on $\mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$ up to a positive constant factor c, i.e.

$$\left.\begin{array}{l} c \cdot \phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) = \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2}) \\ \text{for all } \mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2} \end{array}\right\} \iff \mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2}).$$

The condition on the left-hand side is equivalent to

$$c \cdot \frac{\phi(\mathbf{u})}{\phi^{\downarrow z_1}(\mathbf{u}^{\downarrow z_1})} = c \cdot \phi_{x_1|z_1}(\mathbf{u}^{\downarrow x}|\mathbf{u}^{\downarrow z_2}) = \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2}) = \frac{\psi(\mathbf{u})}{\psi^{\downarrow z_2}(\mathbf{u}^{\downarrow z_2})}$$

for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$, whence to

$$c \cdot \phi(\mathbf{u}) \cdot \psi^{\downarrow z_2}(\mathbf{u}^{\downarrow z_2}) = \phi^{\downarrow z_1}(\mathbf{u}^{\downarrow z_1}) \cdot \psi(\mathbf{u})$$

for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$, and finally, in light of Theorem 3.3, to

$$c \cdot {k_1}^{-1} \cdot (\phi \otimes \psi^{\downarrow z_2})(\mathbf{u}) = c \cdot \phi(\mathbf{u}) \cdot \psi^{\downarrow z_2}(\mathbf{u}^{\downarrow z_2}) = \phi^{\downarrow z_1}(\mathbf{u}^{\downarrow z_1}) \cdot \psi(\mathbf{u}) = {k_2}^{-1} \cdot (\phi^{\downarrow z_1} \otimes \psi)(\mathbf{u})$$

for some positive constants $k_1, k_2 > 0$ not depending on $\mathbf{u}$. Here,

$$\begin{aligned} c \cdot {k_1}^{-1} &= c \cdot {k_1}^{-1} \int_{\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}} \phi^{\downarrow z_1} \otimes \psi(\mathbf{u}) \\ &= {k_2}^{-1} \cdot \int_{\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}} \phi \otimes \psi^{\downarrow z_2}(\mathbf{u}) \\ &= {k_2}^{-1}. \end{aligned}$$

Therefore, $\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2})$ if and only if $\phi \otimes \psi^{\downarrow z_2} = \phi^{\downarrow z_1} \otimes \psi$ and $d(\phi) = d(\psi)$ with $c = \frac{k_1}{k_2}$. □

The geometric interpretation of these results is very simple: A conditional Gaussian density represents a distribution over the parallel linear manifolds given by the regression equation (5.9). The scalar factor c depends on the head variables chosen for the axis of integration over these sets; more technically, c is the Jacobian determinant of the corresponding variable substitution, which is constant since the transformation is linear. In contrast, such a constant factor does not appear in the left-hand side of the equivalence (7.14) since the normalisation constants ${k_1}^{-1}$ and ${k_2}^{-1}$ of the combination already account for c.

Example **7.7**. Figure 7.2 shows parallel straight lines in the two-dimensional space. This situation corresponds to a Gaussian hint with a one-row design matrix. Here, the x_1- and the x_2 axis can be used as pointer to these straight lines. If neither coefficient in the design matrix is 0, both submatrices are regular, so both variables can be in the head (see Lemma 7.1). ⊘

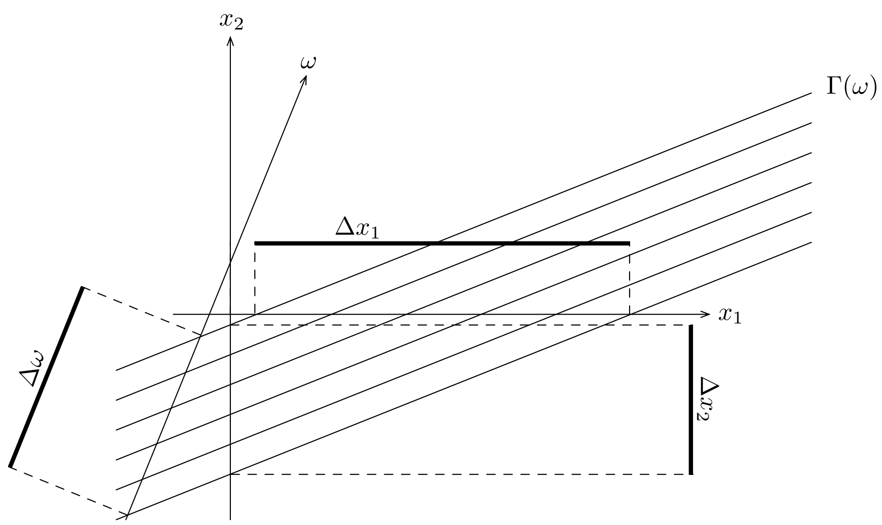


Figure 7.2: The same focal sets $\Gamma(\omega)$ can be indexed by the x_1, the x_2-, or the ω-axis. The variables x_1 and x_2 can both be chosen as head of a corresponding conditional Gaussian density. The constant factor c compensates for the ratio of Δx_1 and Δx_2.

7.4 Combination of Gaussian Hints and of CGDs

The following theorem shows that the combination of Gaussian hints can be carried over to conditional Gaussian densities.

Theorem **7.8**. *Let $\phi_{x_1|z_1}$ and $\psi_{x_2|z_2}$ be conditional Gaussian densities. Then, there is a Gaussian potential $\omega \in \mathcal{G}$ such that*

$$\mathcal{H}(\phi_{x_1|z_1}) \otimes \mathcal{H}(\psi_{x_2|z_2}) = \mathcal{H}\left(\omega_{x_1 \cup x_2 | z_1 \cup z_2 - (x_1 \cup x_2)}\right), \tag{7.15}$$

in which case it holds that

$$(\phi \otimes \psi) \otimes \omega^{\downarrow z_1 \cup z_2 - (x_1 \cup x_2)} = (\phi^{\downarrow z_1} \otimes \psi^{\downarrow z_2}) \otimes \omega. \tag{7.16}$$

⊘

PROOF. Consider the combined GLS (A, μ, K) of $\mathcal{H}(\phi_{x_1|z_1}) = ((I_{x_1}, B_1), z_1, K_1)$ and $\mathcal{H}(\psi_{x_2|z_2}) = ((I_{x_2}, B_2), z_2, K_2)$ where $A \in \mathbb{R}(m = |x_1| + |x_2|, x_1 \cup z_1 \cup x_2 \cup z_2)$,

$$A = \begin{pmatrix} I_{x_1-x_2} & 0_{x_1-x_2, x_1\cap x_2} & B_1{}^{\Rightarrow x_1, z_1\cap x_2} & B_1{}^{\Rightarrow x_1, (z_1\cup z_2)-(x_1\cup x_2)} \\ 0_{x_1\cap x_2 . x_1 - x_2} & I_{x_1\cap x_2} & & \\ B_2{}^{\Rightarrow x_2, z_2\cap x_1} & I_{x_1\cap x_2} & & B_2{}^{\Rightarrow x_2, (z_1\cup z_2)-(x_1\cup x_2)} \\ & & I_{x_2-x_1} & \end{pmatrix},$$

Here, the three submatrices corresponding to the variables $x_1 - x_2$, $x_1 \cap x_2$, and to $x_2 - x_1$ all have full row rank and since the remaining columns are linear combinations of the columns of these three submatrices, it follows that $\mathrm{r}(A) = |x_1 \cup x_2|$. Let $T \in \mathbb{R}(r, m)$ be an admissible matrix, which has full row rank $r = \mathrm{r}(T) = |x_1 \cup x_2|$, and let $A^{\downarrow x_1 \cup x_2}$ be the submatrix of A of the columns corresponding to the variables $x_1 \cup x_2$. Then, in light of Lemma 8.3.2 of (Harville, 1997; p.83),

$$\mathrm{r}(TA^{\downarrow x_1 \cup x_2}) = \mathrm{r}(T) = |x_1 \cup x_2|,$$

i.e. $TA^{\downarrow x_1 \cup x_2}$ is regular. Therefore, equation (7.15) follows from Lemma 7.3 and equation (7.16) then follows from Theorem 7.6. □

This theorem shows that the combination of Gaussian densities corresponds to combining the numerators and the denominators since it follows from (7.16) and Theorem 7.5 that

$$c \cdot \frac{\omega(\mathbf{u})}{\omega^{\downarrow z_1 \cup z_2 - (x_1 \cup x_2)}(\mathbf{u}^{\downarrow z_1 \cup z_2 - (x_1 \cup x_2)})} = \frac{(\phi_1 \otimes \psi_1)(\mathbf{u})}{(\phi^{\downarrow z_1} \otimes \psi^{\downarrow z_2})(\mathbf{u}^{\downarrow z_1 \cup z_2})}$$

for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1 \cup x_2 \cup z_2}$ and some constant $c > 0$ not depending on $\mathbf{u}$.

REMARK **7.9**. Of course, the pair $(\phi_1 \otimes \psi_1, \phi^{\downarrow z_1} \otimes \psi^{\downarrow z_2})$ is in general not a conditional Gaussian potential in the sense of Definition 5.2.
More precisely, it does not follow from equation (7.16) that $(\omega, \omega^{\downarrow z_1 \cup z_2 - (x_1 \cup x_2)})$ is the same pair as $(\phi_1 \otimes \phi_2, (\phi_1 \otimes \phi_2)^{\downarrow z_1 \cup z_2 - (x_1 \cup x_2)})$. In other words, it does not generally hold that $(\phi_1 \otimes \phi_2)_{x_1 \cup x_2 | z_1 \cup z_2 - (x_1 \cup x_2)}$ corresponds to the combination of $\mathcal{H}(\phi_{x_1|z_1})$ and $\mathcal{H}(\psi_{x_2|z_2})$. ⊘

Therefore, in the algebraic approach of Chapter 8, arbitrary pairs of Gaussian potentials will be considered: Equation (7.16) will be used as the definition of an equivalence relation in $\mathcal{G} \times \mathcal{G}$, and combination can be defined component-wise.

7.5 Variable Elimination in Gaussian Hints and CGDs

In order to analyse the marginalisation of Gaussian hints in terms of the related conditional Gaussian densities, the following definitions will be helpful.

Definition **7.10**. *Let $h = (A, z, K)$ be a Gaussian hint, where $A \in \mathbb{R}(m, x)$. Then, a variable $X \in x = d(h)$ is called* vacuous *in h if the column in A corresponding to X contains only zeros, i.e. if*

$$A^{\downarrow\{X\}} = 0_m;$$

else it is called non-vacuous. ⊘

This definition is sound: If (BA, Bz, BK) is any another representative of h for some regular matrix $B \in \mathbb{R}(m, m)$,

$$(BA)^{\downarrow\{X\}} = BA^{\downarrow\{X\}} = B0_m = 0_m$$

shows that X is vacuous in every representative of the hint. Hence, the definition does not depend on the representative of the hint. The following example gives a geometric interpretation of vacuous variables.

Example **7.11**. Let a Gaussian hint h on the variables X_1, X_2 be given by $A = \begin{pmatrix} 1 & 0 \end{pmatrix}$, $\mu = (1)$ and $K = (1)$. Let x_1 and x_2 be the singleton sets consisting of the corresponding variable only. Then, the focal sets are straight lines parallel to the x_2-axis as shown in Figure 7.3. Furthermore, since these focal sets contain points of the same conditional Gaussian density, this shows that the Gaussian density function does not depend on the vacuous variables x_2 or, in other words, that the vacuous variables are *irrelevant* for the conditional Gaussian density function. ⊘

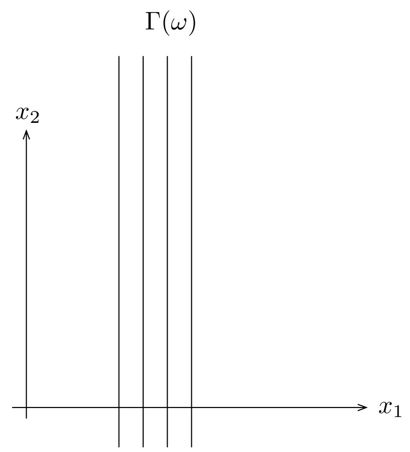


Figure 7.3: If the variable x_2 is vacuous, the focal sets are parallel to the x_2-axis.

The following two lemmata characterise vacuous and non-vacuous variables of a Gaussian hint in terms of the related conditional Gaussian densities. The first lemma shows that vacuous variables are always in the tail. Further, for a non-vacuous variable, a corresponding conditional Gaussian density can always be found with that variable in its head.

LEMMA **7.12**. *Let $h = (A, z, K)$ be a Gaussian hint where $A \in \mathbb{R}(m, s)$. Then, a variable $Z \in s$ is vacuous if and only if there is no conditional Gaussian density with Z in its head, i.e. there is no $\phi_{x|z}$ such that $\mathcal{H}(\phi_{x|z}) = h$ and $Z \in x$.* ⊘

PROOF. On the one hand, assume that Z is vacuous in h. Then, for every Gaussian potential $\phi = (\mu, K) \in \mathcal{G}$ such that $\mathcal{H}(\phi_{x|z}) = h$, it holds, in light of equation (7.1), that

$$0_{x,\{Z\}} = A^{\downarrow x,\{Z\}} = \left(K^{\downarrow x^{-1}} K^{\downarrow x,z}\right)^{\downarrow x,\{Z\}} = K^{\downarrow x^{-1}} K^{\downarrow x,\{Z\}}.$$

Thus, since every principal submatrix $K^{\downarrow x}$ of a symmetric positive definite matrix K is regular as well as its inverse $\left(K^{\downarrow x}\right)^{-1}$, it follows that $K^{\downarrow x,\{Z\}} = 0_x$. Therefore, if $Z \in x$, the diagonal element $K(Z, Z) = 0$. However, since K is positive definite, the diagonal elements are strictly positive in light of Corollary 14.2.13 of (Harville, 1997; p.214). Hence, $Z \in X$ leads to a contradiction. Thus, a vacuous variable must be in the tail z.

On the other hand, it will now be shown that, if Z is non-vacuous, there is always a Gaussian potential $\phi \in \mathcal{G}$ such that $\mathcal{H}(\phi_{x|z}) = h$ and $Z \in x$. If Z is non-vacuous, then by definition

$$A^{\downarrow x,\{Z\}} \neq 0_{x,\{Z\}}.$$

Since A has full row rank m and since $A^{\downarrow x,\{Z\}} \neq 0_{x,\{Z\}}$, there is thus a subset $x' \subseteq x \cup z$ such that $Z \in x'$ and $A^{\downarrow x,x'}$ is regular. Let $B = A^{\downarrow x,x'^{-1}} \in \mathbb{R}(x', x)$ and $z' = (x \cup z) - x'$. Then, the Gaussian hint

$$\left((I_{x'}, BA^{\downarrow x,z'}), Bz, B^{-1'} K B^{-1}\right)$$

equals h. Define

$$\phi = \left(\begin{pmatrix} Bz \\ 0_{z'} \end{pmatrix}, \begin{pmatrix} K & KBA^{\downarrow x,z'} \\ A^{\downarrow x,z''} B' K & I_z + A^{\downarrow x,z''} B' KBA^{\downarrow x,z'} \end{pmatrix}\right).$$

By Lemma A.7, the second element of ϕ is a symmetric and positive definite matrix, hence ϕ is a Gaussian potential. Then,

$$\begin{aligned} \mathcal{H}(\phi_{x|z}) &= \left((I_{x'}, K^{-1} KBA^{\downarrow x,z'}), Bz + K^{-1} KBA^{\downarrow x,z'} 0_{z'}, K\right) \\ &= \left((I_{x'}, BA^{\downarrow x,z'}), Bz, B^{-1'} K B^{-1}\right) = h, \end{aligned}$$

where $Z \in x'$. This shows by contraposition that the converse implication also holds. □

The following lemma shows that vacuous variables in a conditional Gaussian density can be separated out in a factor of their own.

Lemma **7.13**. *Let $\mathcal{H}(\phi_{x|z}) = h$, $s = x \cup z$, $\phi = (\mu, K) \in \mathcal{G}$, and $Z \in z$ be vacuous in h. Then, there are ϕ_1, ϕ_2 such that*

- $(\phi_1 \otimes \phi_2)_{x|z} = \phi_{x|z}$
- $Z \notin d(\phi_1)$ *and*
- $d(\phi_2) = \{Z\}$. ⊘

Proof. In light of equation (7.1),

$$\left(K^{\downarrow x^{-1}} K^{\downarrow x,z}\right)^{\downarrow x,\{Z\}} = K^{\downarrow x^{-1}} K^{\downarrow x,\{Z\}} = 0_x.$$

Since $K^{\downarrow x^{-1}}$ is regular, its columns are linearly independent, which then shows that

$$K^{\downarrow x,\{Z\}} = 0_{x,\{Z\}}. \tag{7.17}$$

Then,

$$\phi^{\downarrow z} = (\mu^{\downarrow z}, K_z)$$

where

$$K_z = K^{\downarrow z} - K^{\downarrow z,x} K^{\downarrow x^{-1}} K^{\downarrow x,z} = K^{\downarrow z} - \begin{pmatrix} K^{\downarrow \tilde{z},x} K^{\downarrow x^{-1}} K^{\downarrow x,\tilde{z}} & 0_{\tilde{z},\{Z\}} \\ 0_{\{Z\},\tilde{z}} & 0_{\{Z\},\{Z\}} \end{pmatrix}$$

for $\tilde{z} = z - \{Z\}$. Define the Gaussian potential $\phi_1 = (\mu_1, K_1)$ by

$$\mu_1 = \mu^{\downarrow x \cup \tilde{z}}, \qquad K_1 = K^{\downarrow x \cup \tilde{z}},$$

and let $\phi_2 = (\mu_2, K_2)$ be an arbitrary Gaussian potential with domain $d(\phi_2) = \{Z\}$. In light of Theorem 7.6 and the combination axiom, it has to be shown that

$$\phi \otimes {\phi_1}^{\downarrow \tilde{z}} \otimes \phi_2 = \phi \otimes (\phi_1 \otimes \phi_2)^{\downarrow z} = \phi^{\downarrow z} \otimes (\phi_1 \otimes \phi_2). \tag{7.18}$$

Let $(\bar{\mu}, \bar{K}) = \phi_1 \otimes \phi_2$ where

$$\bar{\mu} = \begin{pmatrix} \mu^{\downarrow x \cup \tilde{z}} \\ \mu_2 \end{pmatrix}, \qquad \bar{K} = \begin{pmatrix} K^{\downarrow x \cup \tilde{z}} & 0 \\ 0 & K_2 \end{pmatrix}.$$

Further,

$${\phi_1}^{\downarrow \tilde{z}} = (\mu^{\downarrow \tilde{z}}, K^{\downarrow \tilde{z}} - K^{\downarrow \tilde{z},x} K^{\downarrow x^{-1}} K^{\downarrow x,\tilde{z}}).$$

Let $(\bar{\mu}_z, \bar{K}_z) = {\phi_1}^{\downarrow \tilde{z}} \otimes \phi_2$ where

$$\bar{\mu}_z = \begin{pmatrix} \mu^{\downarrow \tilde{z}} \\ \mu_2 \end{pmatrix}, \qquad \bar{K}_z = \begin{pmatrix} K^{\downarrow \tilde{z}} - K^{\downarrow \tilde{z},x} K^{\downarrow x^{-1}} K^{\downarrow x,\tilde{z}} & 0 \\ 0 & K_2 \end{pmatrix}.$$

Hence, using (7.17),

$$\begin{aligned}
K+\bar{K}_z^{\uparrow s} =& K+\begin{pmatrix} K^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}} & 0\\ 0 & K_2\end{pmatrix}^{\uparrow s}\\
=&\begin{pmatrix} K^{\downarrow x} & K^{\downarrow x,\tilde{z}} & 0_{x,\{Z\}}\\ K_{\tilde{z},x} & 2K^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}} & K^{\downarrow\tilde{z},\{Z\}}\\ 0_{\{Z\},x} & K^{\downarrow\{Z\},\tilde{z}} & K^{\downarrow\{Z\}}+K_2\end{pmatrix}\\
=&\begin{pmatrix} K^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}} & K^{\downarrow\tilde{z},\{Z\}}\\ K^{\downarrow\{Z\},\tilde{z}} & K^{\downarrow\{Z\}}\end{pmatrix}^{\uparrow s}\\
&+\begin{pmatrix} K^{\downarrow x} & K^{\downarrow x,\tilde{z}} & 0\\ K^{\downarrow\tilde{z},x} & K^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}} & 0\\ 0 & 0 & K_2\end{pmatrix}\\
=& K_z{}^{\uparrow s}+\bar{K}
\end{aligned}$$

and

$$\begin{aligned}
K\mu+\left(\bar{K}_z\bar{\mu}_z\right)^{\uparrow s} =& K\mu+\begin{pmatrix}(K^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}})\mu^{\downarrow\tilde{z}}\\ K_2\mu_2\end{pmatrix}^{\uparrow s}\\
=&\begin{pmatrix} K^{\downarrow x}\mu^{\downarrow x}+K^{\downarrow x,\tilde{z}}\mu^{\downarrow\tilde{z}}\\ K^{\downarrow\tilde{z},x}\mu^{\downarrow x}+2K^{\downarrow\tilde{z}}\mu^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}}\mu^{\downarrow\tilde{z}}+K^{\downarrow\tilde{z},\{Z\}}\mu^{\downarrow\{Z\}}\\ K^{\downarrow\{Z\},z}\mu^{\downarrow z}+K_2\mu_2\end{pmatrix}\\
=&\begin{pmatrix} K^{\downarrow\tilde{z}}\mu^{\downarrow\tilde{z}}-K^{\downarrow\tilde{z},x}K^{\downarrow x^{-1}}K^{\downarrow x,\tilde{z}}\mu^{\downarrow\tilde{z}}+K^{\downarrow\tilde{z},\{Z\}}\mu^{\downarrow\{Z\}}\\ K^{\downarrow\{Z\},z}\mu^{\downarrow z}\end{pmatrix}^{\uparrow s}\\
&+\begin{pmatrix} K^{\downarrow x}\mu^{\downarrow x}+K^{\downarrow x,\tilde{z}}\mu^{\downarrow\tilde{z}}\\ K^{\downarrow\tilde{z},x}\mu^{\downarrow x}+K^{\downarrow\tilde{z}}\mu^{\downarrow\tilde{z}}\\ +K_2\mu_2\end{pmatrix}\\
=&(K_z\mu_z)^{\uparrow s}+\bar{K}\bar{\mu}.
\end{aligned}$$

This shows (7.18). □

The interpretation of the above lemma is as follows: The corresponding conditional Gaussian density does not depend on the vacuous variable Z since Z only appears in a constant factor, i.e.

$$\phi_{x|z}(\mathbf{x}|\mathbf{z})=\frac{\phi_1(\mathbf{x},\mathbf{z}^{\downarrow z'})}{\phi_1{}^{\downarrow z'}(\mathbf{z}^{\downarrow z'})}\cdot\frac{\phi_2(\mathbf{z}^{\downarrow\{Z\}})}{\phi_2(\mathbf{z}^{\downarrow\{Z\}})}. \tag{7.19}$$

The following theorem shows how the elimination of variables in Gaussian hints can be carried over to conditional Gaussian densities. Recall that

$$h^{-X}=h^{\downarrow u-\{X\}},$$

for $X\in d(h)=u$. Furthermore, every variable $X\in d(h)$ is either non-vacuous or vacuous. According to Lemma 7.1, there is then always a Gaussian potential $\phi\in\mathcal{G}$ such that $\mathcal{H}(\phi_{x|z})=h=(A,\mu_x,K_x)$ where $A\in\mathbb{R}(x,x\cup z)$.

THEOREM **7.14**. *Let $\phi_{x|z}$ be a conditional Gaussian density and let $h = \mathcal{H}(\phi_{x|z})$. Let $X \in x \cup z$.*

1. *If X is non-vacuous and $X \in x$, then*

$$\mathcal{H}({\phi^{-X}}_{x'|z}) = \mathcal{H}(\phi_{x|z})^{-X}. \tag{7.20}$$

for $x' = x - \{X\}$.

2. *If X is vacuous, then $X \in z$ and*

$$\mathcal{H}(\phi_{x|z})^{-X} = \mathcal{H}({\phi^{-X}}_{x|z'}) \tag{7.21}$$

for $z' = z - \{X\}$. ⊘

PROOF. In the first case, if X is non-vacuous, there is a subset $x' \subseteq x \cup z$ of cardinality $|x'| = |x|$ containing $X \in x'$ such that

$$B = A^{\downarrow x, x'}$$

is regular. Hence, transformation by $B^{-1} \in \mathbb{R}(x', x)$ yields the equivalent representative

$$(B^{-1}A, B^{-1}\mu_x, B'K_xB),$$

where $(B^{-1}A)^{\downarrow x'} = I_{x'}$. Therefore, it can be assumed without loss of generality that $X \in x$. On the one hand, the matrix

$$E = {I_x}^{\downarrow x - \{X\}, x}$$

is a projection matrix for $x' = x - \{X\}$, hence

$$\begin{aligned}\mathcal{H}(\phi_{x|z})^{-X} &= \left(EA, E\mu_x, (E{K_x}^{-1}E')^{-1}\right) \\ &= \left(A^{\downarrow x', x' \cup z}, {\mu_x}^{\downarrow x'}, \left(({K_x}^{-1})^{\downarrow x'}\right)^{-1}\right),\end{aligned}$$

where

$$A^{\downarrow x'} = I_{x'}. \tag{7.22}$$

On the other hand, define

$$\phi = (\mu, K), \qquad \mu \in \mathbb{R}^{x \cup z}, \quad K \in \mathbb{R}(x \cup z, x \cup z)$$

where

$$\mu = \begin{pmatrix} \mu_x \\ 0_z \end{pmatrix}, \quad K = \begin{pmatrix} K_x & K_x A^{\downarrow x,z} \\ (A^{\downarrow x,z})' K_x & I_z + (A^{\downarrow x,z})' K_x A^{\downarrow x,z} \end{pmatrix}, \tag{7.23}$$

which, according to the proof of Lemma 7.1, is a Gaussian potential such that

$$\mathcal{H}(\phi_{x|z}) = (A, \mu_x, K_x).$$

Furthermore, let $\phi^{-X} = (\bar{\mu}, \bar{K})$ where

$$\bar{\mu} = \begin{pmatrix} \mu_x{}^{\downarrow x'} \\ 0_{\mu_x} \end{pmatrix}, \qquad \bar{K} = \left(\left(K^{-1}\right)^{\downarrow x' \cup z} \right)^{-1}. \tag{7.24}$$

Then,

$$\bar{\mu}^{\downarrow x'} = \mu_x{}^{\downarrow x'}, \tag{7.25}$$

and

$$\begin{aligned}
\bar{K}^{\downarrow x'} &\overset{(1)}{=} \left(K^{\downarrow x' \cup z} - K^{\downarrow x' \cup z, \{X\}} (K^{\downarrow \{X\}})^{-1} K^{\downarrow \{X\}, x' \cup z} \right)^{\downarrow x'} \\
&\overset{(2)}{=} K^{\downarrow x'} - K^{\downarrow x', \{X\}} (K^{\downarrow \{X\}})^{-1} K^{\downarrow \{X\}, x'} \\
&\overset{(3)}{=} K_x{}^{\downarrow x'} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}, x'} \\
&\overset{(4)}{=} ((K_x{}^{-1})^{\downarrow x'})^{-1}
\end{aligned} \tag{7.26}$$

using Lemma A.6 in (1) and (4), Lemma 3.2 in (2), and using equation (7.23) in (3). Further,

$$\begin{aligned}
\bar{K}^{\downarrow x', z} &\overset{(1)}{=} \left(K_x{}^{\downarrow x' \cup z} - K_x{}^{\downarrow x' \cup z, \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}, x' \cup z} \right)^{\downarrow x', z} \\
&\overset{(2)}{=} K_x{}^{\downarrow x', z} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}, z} \\
&\overset{(3)}{=} \left(K_x A^{\downarrow x, z} \right)^{\downarrow x', z} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} (K_x A)^{\downarrow \{X\}, z} \\
&\overset{(4)}{=} K_x{}^{\downarrow x', x} A^{\downarrow x, z} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}, x} A^{\downarrow x, z} \\
&\overset{(5)}{=} \left(K_x{}^{\downarrow x', x} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}, x} \right) A^{\downarrow x, z} \\
&\overset{(6)}{=} \left(\left((K_x{}^{-1})^{\downarrow x'} \right)^{-1}, K_x{}^{\downarrow x', \{X\}} - K_x{}^{\downarrow x', \{X\}} (K_x{}^{\downarrow \{X\}})^{-1} K_x{}^{\downarrow \{X\}} \right) A^{\downarrow x, z} \\
&\overset{(7)}{=} \left(\bar{K}^{\downarrow x'}, 0_{x', \{X\}} \right) A^{\downarrow x, z} \\
&\overset{(8)}{=} \bar{K}^{\downarrow x'} A^{\downarrow x', z},
\end{aligned} \tag{7.27}$$

using (7.23) in (1) and (6), Lemma 3.2 in (2) and (4), using equation (7.23) in (3) and finally (7.26) in (7). Hence, using (7.22)–(7.27),

$$\begin{aligned}
\mathcal{H}(\phi^{-X}{}_{x'|z}) &= \left((I_{x'}, \left(\bar{K}^{\downarrow x'}\right)^{-1} \bar{K}^{\downarrow x', z}), \bar{\mu}^{\downarrow x'}, \bar{K}^{\downarrow x'} \right) \\
&= \left(A^{\downarrow x', x' \cup z}, \mu_x{}^{\downarrow x'}, ((K_x{}^{-1})^{\downarrow x'})^{-1} \right) \\
&= \mathcal{H}(\phi_{x|z})^{-X}.
\end{aligned}$$

In the second case, if X is vacuous, only the null column corresponding to the vacuous variable X has to be removed, i.e.

$$(A, \mu_x, K_x)^{-X} = (A^{\downarrow x, x' \cup z}, \mu_x, K_x).$$

Let $\phi = (\mu, K)$ such that $\mathcal{H}(\phi_{x|z}) = h$. As seen in the proof of Lemma 7.12,

$$K^{\downarrow x,\{X\}} = 0. \tag{7.28}$$

Further, let $\phi^{-X} = \left(\bar{\mu}, \bar{K}\right)$ where $\bar{\mu} = \mu^{\downarrow x \cup z'}$ and

$$\begin{aligned}
\bar{K} &\overset{(1)}{=} K^{\downarrow x \cup z'} - K^{\downarrow x \cup z',\{X\}} {K^{\downarrow \{X\}}}^{-1} K^{\downarrow \{X\}, x \cup z'} \\
&\overset{(2)}{=} \begin{pmatrix} K^{\downarrow x} & K^{\downarrow x,z'} \\ K^{\downarrow z',x} & K^{\downarrow z'} \end{pmatrix} - \begin{pmatrix} 0 \\ K^{\downarrow z',\{X\}} \end{pmatrix} {K^{\downarrow \{X\}}}^{-1} \begin{pmatrix} 0 & K^{\downarrow \{X\},z'} \end{pmatrix} \\
&= \begin{pmatrix} K^{\downarrow x} & K^{\downarrow x,z'} \\ K^{\downarrow z',x} & K^{\downarrow z'} - K^{\downarrow z'} {K^{\downarrow \{X\}}}^{-1} K^{\downarrow \{X\},z'} \end{pmatrix},
\end{aligned}$$

using (7.23) in (1) and equation (7.28) in (2). Then,

$$\begin{aligned}
A^{\downarrow x,(x \cup z')} &= \begin{pmatrix} I_x & {K^{\downarrow x}}^{-1} K^{\downarrow x,z} \end{pmatrix}^{\downarrow x,(x\cup z')} = \begin{pmatrix} I_x & {K^{\downarrow x}}^{-1} K^{\downarrow x,z'} \end{pmatrix} \\
&= \begin{pmatrix} I_x & {\bar{K}^{\downarrow x}}^{-1} \bar{K}^{\downarrow x,z'} \end{pmatrix}, \\
\mu_x &= \mu^{\downarrow x} + {K^{\downarrow x}}^{-1} K^{\downarrow x,z} \mu^{\downarrow z} = \mu^{\downarrow x} + {K^{\downarrow x}}^{-1} K^{\downarrow x,z'} \mu^{\downarrow z'} \\
&= \bar{\mu}^{\downarrow x} + {\bar{K}^{\downarrow x}}^{-1} \bar{K}^{\downarrow x,z'} \bar{\mu}^{\downarrow z'}, \qquad \text{and} \\
K_x &= K^{\downarrow x} = \bar{K}^{\downarrow x},
\end{aligned}$$

using (7.28) for μ_x. This shows that indeed $\mathcal{H}(\phi_{x|z})^{-X} = \mathcal{H}({\phi^{-X}}_{x|z'})$. □

In summary, every variable X is either vacuous or non-vacuous in a Gaussian hint:

- If it is non-vacuous, there is a conditional Gaussian density $\phi_{x|z}$ with the variable in the head, $X \in x$. Then, the elimination of X in h corresponds to *integration* over X in the numerator, i.e.

$$ {\phi^{-X}}_{x'|z}(\mathbf{x}'|\mathbf{z}) = \frac{\phi^{-X}(\mathbf{x}', \mathbf{z})}{\phi^{\downarrow z}(\mathbf{z})}. \tag{7.29}$$

- On the other hand, if X is vacuous, it is always in the tail of the related conditional Gaussian densities. Here, the elimination of X in h corresponds to the *reduction of an irrelevant constant factor* as seen in equation (7.19) and *not to integration* of the conditional Gaussian density.

Chapter Synopsis

Two conditional Gaussian densities $\phi_{x_1|z_1}, \psi_{x_2|z_2}$ are related to the same Gaussian hint $\mathcal{H}(\phi_{x_1|z_1}) = \mathcal{H}(\psi_{x_2|z_2})$ if and only if

- they represent the same function up to a constant factor $c > 0$, i.e.

  $$c \cdot \phi_{x_1|z_1}(\mathbf{u}^{\downarrow x_1}|\mathbf{u}^{\downarrow z_1}) = \psi_{x_2|z_2}(\mathbf{u}^{\downarrow x_2}|\mathbf{u}^{\downarrow z_2})$$

  for all $\mathbf{u} \in \mathbb{R}^{x_1 \cup z_1} = \mathbb{R}^{x_2 \cup z_2}$,

or, equivalently, if and only if

- they are related by the equation $\phi \otimes \psi^{\downarrow z_2} = \phi^{\downarrow z_1} \otimes \psi$ and $d(\phi) = d(\psi)$.

If a variable is vacuous in a Gaussian hint (i.e. if the corresponding column in the design matrix is 0), it is always in the tail of all corresponding conditional Gaussian densities and can be *reduced* in the numerator and the denominator. On the other hand, if a variable is non-vacuous in a Gaussian hint, there is a conditional Gaussian density that has the variable in its head, and the variable can be marginalised out in the numerator respecting the rules of *integration*. Therefore, every variable can be eliminated (either because its vacuous or non-vacuous). This leads to *full marginalisation* of conditional Gaussian densities.

Furthermore, since Gaussian hints are closed under combination, the same holds for conditional Gaussian densities since the operations are compatible. In conditional Gaussian densities, the union of any two heads becomes a head of the combination.

Discussion

In the Bayesian approach, head and tail of a conditional Gaussian potential are fixed. Nonetheless, Lauritzen and Jensen (2001) distinguish marginalising out head variables from the reduction of tail variables. This corresponds to the eliminiation of non-vacuous and vacuous variables, respectively. However, they do not identify equivalent conditional Gaussian densities, i.e. a non-vacuous variable in the tail cannot be eliminated in their approach. In contrast, by considering conditional Gaussian densities related to the same Gaussian hint by the regression equations, every variable can either be eliminated by integration in the numerator or reduced. By an interplay of elimination by integration in the numerator and reduction of variables in an equivalent conditional Gaussian densities, this then leads to full marginalisation

8
Separative Extension of Gaussian Potentials

As introduced in Section 5.1, conditional Gaussian densities $\phi_{x|z}$ can be represented algebraically by pairs or fractions $(\phi, \phi^{\downarrow z})$ of Gaussian potentials. It is well known from semigroup theory that a *cancellative* semigroup can be embedded in a group of quotients of equivalent fractions. The most famous example is the embedding of the multiplicative semigroup of natural numbers (without zero) in the rational numbers, represented by fractions of non-zero natural numbers.
The same idea can be generalised to valuation algebras: Combination in the extension is defined component-wise as the combination of the numerators and denominators, which corresponds to the laws of calculus of the product of rational numbers. Marginalisation can only be partially defined in the extension: If variables only appear in the numerator (but not in the denominator) of a fraction, they can be marginalised out (or eliminated) in the numerator without affecting the denominator. This complies with the laws of integration of quotient functions.

Chapter Outline

First, the theory of separative valuation algebras is developed for cancellative valuation algebras. In Section 8.1, a valuation algebra of pairs is constructed from a cancellative valuation algebra. Different fractions may be equivalent, as discussed in Section 8.2. For this equivalence relation to be complete under marginalisation, a further Property (M) is required. This *Property (M)* is a sort of converse of the combination axiom, going from a factorisation of a marginal to a factorisation before marginalisation. Such a separative valuation algebra can be embedded into a quotient valuation algebra with division. In Section 8.3, it is shown that this theory can be applied to Gaussian potentials.

Probability densities are not cancellative. However, the subsemigroups of densities of the same *support* (i.e. with the same zeros) are cancellative. In fact, support is an idempotent congruence which decomposes densities into cancellative semigroups. In Section 8.4, the theory of separative valuation algebras is generalised to cover this example. Finally, in Section 8.5, *construction sequences* (Shafer, 1996; Kohlas, 2003) are introduced to generalise the *Chain Rule* of *Bayesian Networks*: A construction sequence factorises an element of the underlying separative valuation algebra into *conditionals*, which are elements of the separative extension only.

The theory of separative valuation algebras was first set forth in (Kohlas, 2003), in a slightly different way. These differences are pointed in the Discussion at the end of this chapter.

8.1 Valuation Algebra of Fractions

Example **8.1**. Rational numbers (without zero) can be represented by pairs or fractions of integers. For instance, $(1, 2)$ represents the rational number 0.5. Furthermore, non-zero integer numbers $\mathbb{Z}^* = \{1, -1, 2, -2\}$ form a valuation algebra on the trivial lattice $D = \{\emptyset\}$ with labelling $d(p) = \emptyset$, combination $\cdot$ and trivial marginalisation $p^{\downarrow\emptyset} = p$. Then, the product of the two rational number represented by (p_1, q_1) and (p_2, q_2) can be represented by the fraction $(p_1 \cdot p_2, q_1 \cdot q_2)$. More formally, the multiplication of rational numbers can be carried over to the multiplication $*$ among these fractions by

$$(p_1, q_1) * (p_2, q_2) = (p_1 \cdot p_2, q_1 \cdot q_2).$$

⊘

Example **8.2**. Similarly, one can extend positive densities by pairs (f, g). Here, a positive density can be represented by the pair (f, e) where e is the constant function with empty domain $e(\diamond) = 1$. The combination of these pairs of densities can be defined in the same as way as for integers. On the contrary, marginalisation is more involved. If no variables in the denominator are integrated out,

$$\int_{\mathbf{t} \in \mathbb{R}^t} \frac{f(\mathbf{t}, \mathbf{u}^{\downarrow x - t})}{g(\mathbf{u}^{\downarrow y})}$$

for $x = d(f)$ and $y = d(g)$, $u = x \cup y$, $\mathbf{u} \in \mathbb{R}^u$ such that $t \cap y = \emptyset$. Therefore, the marginal of (f, g) to $s \supseteq d(\psi)$ can be defined by $(f^{\downarrow s \cap d(\phi)}, g)$. ⊘

These examples motivate the following definitions. Let $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ be a valuation algebra with *full marginalisation.* Let Φ^* be the set of pairs of valuations,

$$\Phi^* = \Phi \times \Phi = \{(\phi, \psi) : \phi, \psi \in \Phi\}. \tag{8.1}$$

Define

$$d^*(\phi, \psi) = d(\phi) \cup d(\psi). \tag{8.2}$$

$$(\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2) = (\phi_1 \otimes \phi_2, \psi_1 \otimes \psi_2) \tag{8.3}$$

and

$$(\phi, \psi)^{\downarrow^* s} = (\phi^{\downarrow s \cap d(\phi)}, \psi) \tag{8.4}$$

for

$$s \in \mathcal{M}^*(\phi, \psi) = \{s : d(\psi) \subseteq s \subseteq d(\phi) \cup d(\psi)\}. \tag{8.5}$$

Lemma **8.3**. *$\mathfrak{A}^* = (\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*)$ is a valuation algebra that extends $\mathfrak{A}$ by the embedding $\phi \mapsto (\phi, e)$.* ⊘

Proof. (A1) Let $(\phi_1, \psi_1), (\phi_2, \psi_2), (\phi_3, \psi_3) \in \Phi^*$. Then, using the semigroup axiom in $\mathfrak{A}$,

$$\begin{aligned}(\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2) &= (\phi_1 \otimes \phi_2, \psi_1 \otimes \psi_2)\\ &= (\phi_2 \otimes \phi_1, \psi_2 \otimes \psi_1)\\ &= (\phi_2, \psi_2) \otimes^* (\phi_1, \psi_1).\end{aligned}$$

Furthermore,

$$\begin{aligned}((\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2)) \otimes^* (\phi_2, \psi_2) &= (((\phi_1 \otimes \phi_2) \otimes \phi_3), ((\psi_1 \otimes \psi_2) \otimes \psi_3))\\ &= ((\phi_1 \otimes (\phi_2 \otimes \phi_3)), (\psi_1 \otimes (\psi_2 \otimes \psi_3)))\\ &= (\phi_1, \psi_1) \otimes^* ((\phi_2, \psi_2) \otimes^* (\phi_2, \psi_2)).\end{aligned}$$

(A2) Let $(\phi_1, \psi_1), (\phi_2, \psi_2) \in \Phi^*$ and $x_1 = d^*(\phi_1, \psi_1)$, $x_2 = d^*(\phi_2, \psi_2)$. Then, using the labelling axiom in $\mathfrak{A}$, $d^*((\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2)) = d^*(\phi_1 \otimes \phi_2, \psi_1 \otimes \psi_2) = d(\phi_1) \cup d(\psi_1) \cup d(\phi_2) \cup d(\psi_2) = x_1 \cup x_2$.

(A3) Let $s \in \mathcal{M}^*(\phi, \psi)$, i.e. $d(\psi) \subseteq s \subseteq d(\phi) \cup d(\psi)$. Then, using the marginalisation axiom in $\mathfrak{A}$, $d^*((\phi, \psi)^{\downarrow^* s}) = d^*(\phi^{\downarrow d(\phi) \cap s}, \psi) = (d(\phi) \cap s) \cup d(\psi) = (d(\phi) \cap s) \cup (d(\psi) \cap s) = (d(\phi) \cup d(\psi)) \cap s = s$.

(A4) Let $s \subseteq t \subseteq d(\phi) \cup d(\psi)$. Then, $s \in \mathcal{M}^*(\phi, \psi) \iff d(\psi) \subseteq s \iff t \in \mathcal{M}^*(\phi, \psi)$ and $s \in \mathcal{M}^*((\phi, \psi)^{\downarrow^* t}) = \mathcal{M}^*(\phi^{\downarrow t \cap d(\phi)}, \psi)$. In both cases, $(\phi, \psi)^{\downarrow^* s} = (\phi^{\downarrow s}, \psi) = (\phi^{\downarrow t \cap d(\phi)}, \psi)^{\downarrow^* s} = ((\phi, \psi)^{\downarrow^* t})^{\downarrow^* s}$.

(A5) Let $(\phi_1, \psi_1), (\phi_2, \psi_2) \in \Phi^*$ with domains y_1 and y_2 and let $z \cap y_2 \in \mathcal{M}^*(\phi_2, \psi_2)$ such that $y_1 \subseteq z \subseteq y_1 \cup y_2$. Hence, $d(\psi_1) \subseteq z$, $d(\psi_2) \subseteq z \cap y_2 \subseteq z$, and thus $d(\psi_1) \cup d(\psi_2) \subseteq z$. Therefore, it follows that $z \in \mathcal{M}^*(\phi_1 \otimes \phi_2, \psi_1 \otimes \psi_2) = \mathcal{M}^*((\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2))$. Finally, using the combination axiom in $\mathfrak{A}$,

$$\begin{aligned}\left((\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2)\right)^{\downarrow^* z} &= (\phi_1 \otimes \phi_2, \psi_1 \otimes \psi_2)^{\downarrow^* z}\\ &= ((\phi_1 \otimes \phi_2)^{\downarrow z \cap (d(\phi_1) \cup d(\phi_2))}, \psi_1 \otimes \psi_2)\\ &= (\phi_1 \otimes {\phi_2}^{\downarrow z \cap d(\phi_2)}, \psi_1 \otimes \psi_2)\\ &= (\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2)^{\downarrow^* z \cap y_2}.\end{aligned}$$

(A6) Let $(\phi, \psi) \in \Phi^*$ and let $x = d^*(\phi, \psi) = d(\phi) \cup d(\psi)$. Since $d(\psi) \subseteq x$, it follows that $x \in \mathcal{M}^*(\phi, \psi)$ and

$$(\phi, \psi)^{\downarrow^* x} = (\phi^{\downarrow x \cap d(\phi)}, \psi) = (\phi, \psi)$$

by the domain axiom in $\mathfrak{A}$.

(A7) The element $e^* = (e, e)$ is an identity element since $e^* \otimes^* (\phi, \psi) = (\phi, \psi) = (\phi, \psi) \otimes^* e^*$.

It remains to be verified that $\mathfrak{A}^*$ extends $\mathfrak{A}$. Indeed,

- $d^*(\phi, e) = d(\phi)$,
- $(\phi, e) \otimes^* (\psi, e) = (\phi \otimes \psi, e)$,
- $d(e) = \emptyset \subseteq s \subseteq d(\phi)$ implies $s \in \mathcal{M}^*(\phi, e)$ and $(\phi, e)^{\downarrow^* s} = (\phi^{\downarrow s}, e)$, and
- $e^* = (e \otimes e, e) = (e, e)$ is the identity element in $\mathfrak{A}^*$. □

The elements of Φ^* will be called separative fractions. It has to be remarked that $\mathfrak{A}^*$ is not an ordinary product of two algebras in the sense of universal algebra, where the Cartesian product of the underlying sets are taken and where the operations are defined coordinate-wise. In $\mathfrak{A}^*$, the lattice of domains is not a Cartesian product, and $\mathcal{M}^*$ is not symmetric, i.e. $\mathcal{M}^*(\phi, \psi)$ and $\mathcal{M}^*(\psi, \phi)$ are in general not the same.

8.2 Separative Valuation Algebras

EXAMPLE **8.4**. The same rational number can be represented by different, equivalent fractions. For instance, $(1, 2)$ and $(2, 4)$ represent the same rational number 0.5. Here, it holds that $1 \cdot 4 = 2 \cdot 2$. More generally, two fractions (p_1, q_1) and (p_2, q_2) represent the same rational number if and only if $p_1 \cdot q_2 = q_1 \cdot p_2$. This relation is clearly reflexive and symmetric. From $p_1 \cdot q_2 = q_1 \cdot p_2$ and $p_2 \cdot q_3 = q_2 \cdot p_3$, it follows that $p_1 \cdot q_2 \cdot q_3 = q_1 \cdot p_2 \cdot q_3 = q_1 \cdot q_2 \cdot p_3$. Hence, $p_1 \cdot q_3 = q_1 \cdot p_3$, which shows that the relation is also transitive. The last step requires more than a general semigroup. ⊘

DEFINITION **8.5**. *A semigroup* $(\Phi, \otimes)$ *is* cancellative *if*

$$\phi \otimes \psi = \phi \otimes \psi' \implies \psi = \psi'. \tag{8.6}$$

⊘

LEMMA **8.6**. *Let* $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ *be a valuation algebra with* full marginalisation *such that* $(\Phi, \otimes)$ *is* cancellative. *Then, the relation* $\equiv^*$ *in* Φ^* *defined by*

$$(\phi, \psi) \equiv^* (\phi', \psi') \qquad \iff \qquad \phi \otimes \psi' = \psi \otimes \phi'$$

is an equivalence relation. In particular, the relation $=^*$ *defined by*

$$\eta_1 =^* \eta_2 \iff \eta_1 \equiv^* \eta_2 \text{ and } d^*(\eta_1) = d^*(\eta_2) \tag{8.7}$$

is an equivalence relation in Φ^*. ⊘

PROOF. Reflexivity and symmetry follow from the commutativity of combination. In order to prove transitivity, assume $(\phi_1, \psi_1) \equiv^* (\phi_2, \psi_2)$ and $(\phi_2, \psi_2) \equiv^* (\phi_3, \psi_3)$, i.e.

$$\phi_1 \otimes \psi_2 = \psi_1 \otimes \phi_2, \tag{8.8}$$

and

$$\phi_2 \otimes \psi_3 = \psi_2 \otimes \phi_3. \tag{8.9}$$

Then, multiplying (8.8) by $\phi_3 \otimes \phi_3$ and substituting (8.9) into it yields

$$\phi_1 \otimes \psi_2 \otimes \psi_3 = \psi_1 \otimes \phi_2 \otimes \psi_3 = \psi_1 \otimes \psi_2 \otimes \phi_3.$$

Hence, by cancellativity, it follows that indeed $\phi_1 \otimes \psi_3 = \psi_1 \otimes \phi_3$, i.e. $(\phi_1, \psi_1) \equiv^* (\phi_3, \psi_3)$.
The relation $=^*$ is an equivalence relation since it refines the partition induced by $\equiv^*$. □

The equivalence classes

$$\Phi^*/=^* \tag{8.10}$$

are called separative quotients. In order for these quotients to form a valuation algebra, the following property of the underlying valuation algebra will be used to prove its completeness under marginalisation and thereby the transitivity of the marginalisation of quotients.

Definition **8.7**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ be a valuation algebra. Then, $\mathfrak{A}$ has the Property (M) if the following implication holds.*

- *Assume $\phi^{\downarrow t}$ factorises as*

$$\phi^{\downarrow t} = \phi_1 \otimes \chi \tag{8.11}$$

  for some ϕ_1 such that $d(\phi_1) = t$.

- *Then, there is a ϕ_2 with domain $x = d(\phi_2) = d(\phi)$ such that*

$$\phi = \phi_2 \otimes \chi \tag{8.12}$$

  and ${\phi_2}^{\downarrow t} = \phi_1$. ⊘

The Property (M) is a sort of converse of the combination axiom: Whereas the combination states that a factorisation $\phi = \phi_2 \otimes \chi$ has the marginal $\phi^{\downarrow t} = {\phi_2}^{\downarrow t} \otimes \chi$, the Property (M) goes in the opposite direction from $\phi^{\downarrow t} = \phi_1 \otimes \chi$ to the factorisation $\phi = \phi_2 \otimes \chi$.

Definition **8.8**. *A valuation algebra $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ is called separative if*

- *the semigroup $(\Phi, \otimes)$ is cancellative and*
- *$\mathfrak{A}$ has the Property (M).* ⊘

Lemma **8.9**. *Let $(\Phi, D, d, \otimes, \downarrow)$ be a separative valuation algebra. The relation $=^*$ is a domain-contained congruence in $(\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*)$.* ⊘

Proof. It has been shown in Lemma 8.6 that $=^*$ is an equivalence relation. Let $(\phi_1, \psi_1) =^* (\phi_1', \psi_1')$ and $(\phi_2, \psi_2) =^* (\phi_2', \psi_2')$, i.e.

$$\phi_1 \otimes \psi_1' = \psi_1 \otimes \phi_1', \quad \phi_2 \otimes \psi_2' = \psi_2 \otimes \phi_2'$$

and

$$d(\phi_1) \cup d(\psi_1) = d(\phi_1') \cup d(\psi_1') \quad \text{and} \quad d(\phi_2) \cup d(\psi_2) = d(\phi_2') \cup d(\psi_2').$$

Hence, the equivalence relation $=^*$ is domain-contained since $d^*(\phi_1, \psi_1) = d(\phi_1) \cup d(\psi_1) = d(\phi_1') \cup d(\psi_1') = d^*(\phi_1', \psi_1')$.
In order to show that $=^*$ is compatible with $\otimes^*$, let $(\phi_1, \psi_1) =^* (\phi_1', \psi_1')$ and $(\phi_2, \psi_2) =^* (\phi_2', \psi_2')$, i.e. $\phi_1 \otimes \psi_1' = \psi_1 \otimes \phi_1'$, $\phi_2 \otimes \psi_2' = \psi_2 \otimes \phi_2'$, $d(\phi_1) \cup d(\psi_1) = d(\phi_1') \cup d(\psi_1')$, and $d(\phi_2) \cup d(\psi_2) = d(\phi_2') \cup d(\psi_2')$. Then, on the one hand,

$$\begin{aligned}(\phi_1 \otimes \phi_2) \otimes (\psi_1' \otimes \psi_2') &= (\phi_1 \otimes \psi_1') \otimes (\phi_2 \otimes \psi_2') \\ &= (\psi_1 \otimes \phi_1') \otimes (\psi_2 \otimes \phi_2') \\ &= (\phi_1' \otimes \phi_2') \otimes (\psi_1' \otimes \psi_2')\end{aligned}$$

since $\otimes$ is associative and commutative. On the other hand,

$$\begin{aligned}d^*(\phi_1, \psi_1) \cup d^*(\phi_2, \psi_2) &= d(\phi_1) \cup d(\psi_1) \cup d(\phi_2) \cup d(\psi_2) \\ &= d(\phi_1') \cup d(\psi_1') \cup d(\phi_2') \cup d(\psi_2') \\ &= d^*(\phi_1', \psi_1') \cup d^*(\phi_2', \psi_2').\end{aligned}$$

This shows that $=^*$ is indeed compatible with $\otimes^*$.
In order to show that $=^*$ is compatible with marginalisation, let $(\phi_1, \psi_1) =^* (\phi_2, \psi_2)$ and $s \in \mathcal{M}^*(\phi_1, \psi_1), \mathcal{M}^*(\phi_2, \psi_2)$. This implies that $\phi_1 \otimes \psi_2 = \psi_1 \otimes \phi_2$ and

$$d(\psi_1), d(\psi_2) \subseteq s \subseteq d(\phi_1) \cup d(\psi_1) = d(\phi_2) \cup d(\psi_2).$$

Hence, by the combination axiom,

$$\phi_1{}^{\downarrow s \cap d(\phi_1)} \otimes \psi_2 = (\phi_1 \otimes \psi_2)^{\downarrow s} = (\phi_2 \otimes \psi_1)^{\downarrow s} = \phi_2{}^{\downarrow s \cap d(\phi_2)} \otimes \psi_1.$$

Since $(s \cap d(\phi_1)) \cup d(\psi_1) = (d(\phi_1) \cup d(\psi_1)) \cap s = s = (s \cap d(\phi_2)) \cup d(\psi_2)$,

$$(\phi_1, \psi_1)^{\downarrow^* s} = (\phi_1{}^{\downarrow s \cap d(\phi_1)}, \psi_1) =^* (\phi_2{}^{\downarrow s \cap d(\phi_2)}, \psi_2) = (\phi_2, \psi_2)^{\downarrow^* s}.$$

This shows that $=^*$ is indeed compatible with $\downarrow^*$.
In order to prove that marginalisation is complete under $=^*$, assume $t \in \mathcal{M}^*(\phi_1, \psi_1)$ and $(\phi_1{}^{\downarrow d(\phi_1) \cap t}, \psi_1) = (\phi_1, \psi_1)^{\downarrow^* t} =^* (\phi_2, \psi_2)$. Since $(\phi_1 \otimes \psi_1, \psi_1 \otimes \psi_1) =^* (\phi_1, \psi_1)$ and since $\mathcal{M}^*(\phi_1 \otimes \psi_1, \psi_1 \otimes \psi_1) = \mathcal{M}^*(\phi_1, \psi_1)$, assume without loss of generality that $d(\psi_1) \subseteq d(\phi_1)$. Using the same argument, assume $d(\psi_2) \subseteq d(\phi_2)$. Then, it holds that $d(\phi_1) \subseteq t = d(\phi_2)$. By the combination axiom,

$$(\phi_1 \otimes \psi_2)^{\downarrow t} = \phi_1{}^{\downarrow t \cap d(\phi_1)} \otimes \psi_2 = \psi_1 \otimes \phi_2.$$

Then, the Property (M) shows that there is a $\phi \in \Phi$ with $d(\phi) = d(\phi_1)$ such that $\phi_1 \otimes \psi_2 = \phi \otimes \psi_1$. Hence, $(\phi_1, \psi_1) =^* (\phi, \psi_2)$. Further, $s \in \mathcal{M}^*(\phi_2, \psi_2)$ implies $s \in \mathcal{M}^*(\phi, \psi_2)$. This shows that $=^*$ satisfies (2.34). □

Theorem **8.10**. $(\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*, =^*)$ *forms a quotient valuation algebra. Furthermore, the mapping* $\phi \mapsto (\phi, e)$ *is an embedding.* ⊘

PROOF. The first claim follows from the quotient valuation algebra Theorem 2.16 and Lemma 8.23.
The mapping is a homomorphism since, for $\phi, \psi \in \Phi$,

- $d(\phi) = d(\phi) \cup \emptyset = d^*(\phi, e)$,
- $(\phi, e) \otimes^* (\psi, e) = (\phi \otimes \psi, e)$,
- $\mathcal{M}(\phi, e) = \{s : \emptyset = d(e) \subseteq s \subseteq d(\phi)\} = \mathcal{M}(\phi)$,
- $(\phi, e)^{\downarrow^* s} = (\phi^{\downarrow s}, e)$, and
- $e^* = (e, e)$ is the identity element.

Finally, the mapping is injective since $(\phi, e) =^* (\phi', e)$ implies $\phi = \phi \otimes e = e \otimes \phi' = \phi'$. □

Recall the conventions of Table 2.1 for quotient valuation algebras:

- representatives are used for their equivalence class, i.e.
  - $(\phi, \psi) \in \Phi^*$ instead of $[(\phi, \psi)]_{=^*} \in \Phi^* / =^*$ and
  - $(\phi, \psi) = (\phi', \psi')$ instead of $(\phi, \psi) =^* (\phi', \psi')$;
- the operator symbols for separative fractions are used for separative quotients; for instance, $(\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2) = (\phi, \psi)$ stands for $(\phi_1, \psi_1) \otimes^* (\phi_2, \psi_2) =^* (\phi, \psi)$.

The valuation algebra $\mathfrak{A}^* = (\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*, =^*)$ is called the separative extension of $(\Phi, D, d, \otimes, \downarrow)$.

Every semigroup

$$\Phi_x^* = \{(\phi, \psi) \in \Phi^* : d^*(\phi, \psi) = x\}$$

of separative fractions of the same domain $x \in D$ is a group:

- the identity element is $e_x = (\chi, \chi)$ for any $\chi \in \Phi$ with domain $d(\chi) = x$ (since $(\phi, \psi) \otimes^* (\chi, \chi) = (\phi, \psi)$ for all $(\phi, \psi) \in \Phi_x^*$);
- the inverse of $(\phi, \psi) \in \Phi_x^*$ is $(\phi, \psi)^{-1} = (\psi, \phi)$ since $(\phi, \psi) \otimes^* (\psi, \phi) = e_x$.

These observations are captured in the following theorem.

THEOREM **8.11**. *A separative extension $\mathfrak{A}^* = (\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*, =^*)$ is a valuation algebra with division, where the groups are formed by the separative quotients of the same domain.* ⊘

8.3 Gaussian Quotients

In order to construct the separative extension of Gaussian potentials, it has to be shown that the semigroup of Gaussian potentials of the same domain are cancellative and that they satisfy the Property (M). As a preliminary step, it is first shown that the semigroups of Gaussian potentials of the same domain are cancellative.

Lemma **8.12**. *The sets $\mathcal{G}_x$ of Gaussian potentials of the same domain $x \in D$,*

$$\mathcal{G}_x = \{\phi \in \mathcal{G} : d(\phi) = x\},$$ ⊘

are cancellative semigroups.

Proof. That $\mathcal{G}_x$ $(x \in D)$ is a semigroup is a direct consequence of the labelling axiom in $\mathcal{G}$. Let $\phi, \psi, \psi' \in \mathcal{G}_x$ for some $x \in D$. Assume $\phi \otimes \psi = \phi \otimes \psi' = (\mu, K)$. If $d(\phi) = \emptyset$, it holds that $\psi = \psi' = e$ since $\mathcal{G}_\emptyset = \{e\}$. Assume $x \neq \emptyset$, $\phi = (\mu_1, K_1)$, $\psi = (\mu_2, K_2)$, and $\psi' = (\mu_3, K_3)$. Then,

$$(\mu_1, K_1) \otimes (\mu_2, K_2) = ((K_1 + K_2)^{-1}(K_1\mu_1 + K_2\mu_2), K_1 + K_2),$$

and

$$(\mu_1, K_1) \otimes (\mu_3, K_3) = ((K_1 + K_3)^{-1}(K_1\mu_1 + K_3\mu_3), K_1 + K_3).$$

The assumption $\phi \otimes \psi = \phi \otimes \psi' = (\mu, K)$ implies $K_1 + K_2 = K_1 + K_3$. Hence, since $\mathbb{R}(x, x)$ is an additive group, it follows that $K_2 = K_3$. Further,

$$(K_1 + K_2)^{-1}(K_1\mu_1 + K_2\mu_2) = (K_1 + K_2)^{-1}(K_1\mu_1 + K_2\mu_3),$$

so, since $\mathbb{R}^x$ is an additive group,

$$(K_1 + K_2)^{-1}K_2\mu_2 = (K_1 + K_2)^{-1}K_2\mu_3.$$

Finally, multiplication by ${K_2}^{-1}(K_1+K_2)$ yields $\mu_2 = \mu_3$. This concludes the proof.□

In order to use Lemma 8.6, it is now shown that the whole semigroup of Gaussian potentials is cancellative.

Lemma **8.13**. *The semigroup of Gaussian potentials is cancellative.* ⊘

Proof. Let $\phi_1 \otimes \phi_2 = \phi_1 \otimes \phi_3$ for any $\phi_1, \phi_2, \phi_3 \in \mathcal{G}$. Let $\phi_1 = (\mu_1, K_1), \phi_2 = (\mu_2, K_2), \phi_3 = (\mu_3, K_3)$, and assume

$$\phi_1 \otimes \phi_2 = \phi_1 \otimes \phi_3 = (\mu, K) = \phi.$$

Let $x = d(\phi)$. Then,

$$K = {K_1}^{\uparrow x} + {K_2}^{\uparrow x} = {K_1}^{\uparrow x} + {K_3}^{\uparrow x}.$$

Hence, since $\mathbb{R}(x, x)$ is an additive group, it follows that

$${K_2}^{\uparrow x} = {K_3}^{\uparrow x}.$$

Since the diagonal elements of the positive definite matrix K are positive (Corollary 14.2.13 of (Harville, 1997; p.214)), and since K_2 and K_3 are positive definite, it follows from ${K_2}^{\uparrow x} = {K_3}^{\uparrow x}$ that $d(\phi_2) = d(\phi_3)$, so, using Lemma 8.12, $K_2 = K_3$. Further,

$$K^{-1}((K_1\mu_1)^{\uparrow x} + (K_3\mu_3)^{\uparrow x}) = \mu = K^{-1}((K_1\mu_1)^{\uparrow x} + (K_2\mu_2)^{\uparrow x}).$$

Then, multiplying by K and subtracting $(K_1\mu_1)^{\uparrow x}$ on both sides yields

$$(K_2\mu_2)^{\uparrow x} = (K_3\mu_3)^{\uparrow x}.$$

Since $d(\phi_2) = d(\phi_3)$ and $K_2 = K_3$, it follows that

$$K_2\mu_2 = K_2\mu_3.$$

Premultiplication by ${K_2}^{-1}$ yields $\mu_2 = \mu_3$. Hence, indeed $(\mu_2, K_2) = (\mu_3, K_3)$. □

As a consequence of Lemma 8.13, the following corollary is obtained.

Corollary **8.14**. *The relation $=^*$ defined by*

$$\eta_1 =^* \eta_2 \iff \eta_1 \equiv^* \eta_2 \text{ and } d^*(\eta_1) = d^*(\eta_2) \tag{8.13}$$

is an equivalence relation in the set $\mathcal{G}^* = \mathcal{G} \times \mathcal{G}$ of fractions of Gaussian potentials. ⊘

The second requirement for Gaussian potentials to be separative is the Property (M).

Lemma **8.15**. *Gaussian potentials have the propery (M)* ⊘

Proof. Let $\phi = (\mu, K)$, $\phi_1 = (\mu_1, K_1)$ and $\chi = (\mu_\chi, K_\chi)$ such that $\phi^{\downarrow t} = \phi_1 \otimes \chi$ and $d(\phi_1) = t$. Define $s = d(\phi) - t$. Then,

$$\phi^{\downarrow t} = (\mu^{\downarrow t}, K^{\downarrow t} - K^{\downarrow t,s}{K^{\downarrow s}}^{-1}K^{\downarrow s,t}).$$

It follows from equation (8.11) that

$$K^{\downarrow t} - K^{\downarrow t,s}{K^{\downarrow s}}^{-1}K^{\downarrow s,t} = K_1 + {K_\chi}^{\uparrow t} \tag{8.14}$$

and

$$\mu^{\downarrow t} = (K_1 + {K_\chi}^{\uparrow t})^{-1}(K_1\mu_1 + {K_\chi}^{\uparrow t}{\mu_\chi}^{\uparrow t}).$$

Define

$$K_2 = \begin{pmatrix} K^{\downarrow s} & K^{\downarrow s,t} \\ K^{\downarrow t,s} & K_1 + K^{\downarrow t,s}{K^{\downarrow s}}^{-1}K^{\downarrow s,t} \end{pmatrix}$$

and

$$\mu_2 = {K_2}^{-1}(K\mu - (K_\chi)^{\uparrow x}(\mu_\chi)^{\uparrow x}).$$

It follows from Lemma A.7 that K_2 is symmetric and positive definite. Thus, $\phi_2 = (\mu_2, K_2)$ is a Gaussian potential.
On the one hand, equation (8.14) implies that

$$K^{\downarrow t} - {K_\chi}^{\uparrow t} = K_1 + K^{\downarrow t,s}{K^{\downarrow s}}^{-1}K^{\downarrow s,t},$$

hence

$$K = (K - {K_\chi}^{\uparrow x}) + {K_\chi}^{\uparrow x} = \begin{pmatrix} K^{\downarrow s} & K^{\downarrow s,t} \\ K^{\downarrow t,s} & K^{\downarrow t} - {K_\chi}^{\uparrow t} \end{pmatrix} + {K_\chi}^{\uparrow x} = K_2 + {K_\chi}^{\uparrow x}.$$

On the other hand, it follows from the definition of μ_2 that

$$\mu = K^{-1}(K_2\mu_2 + (K_\chi)^{\uparrow x}(\mu_\chi)^{\uparrow x}) = (K_2 + {K_\chi}^{\uparrow x})^{-1}(K_2\mu_2 + (K_\chi)^{\uparrow x}(\mu_\chi)^{\uparrow x}).$$

This shows that $\phi = \phi_2 \otimes \chi$. □

Since Gaussian potentials are cancellative and have the Property (M), they form a separative valuation algebra.

Theorem **8.16**. *The valuation algebra of Gaussian potentials is separative.* ⊘

Pairs of Gaussian potentials are called Gaussian fractions and their equivalence classes Gaussian quotients. The set of all Gaussian fractions is denoted $\mathcal{G}^*$.

8.4 Generalisation of Separative Valuation Algebras

In contrast to Gaussian potentials, the semigroup of probability densities is not cancellative. However, the semigroups of probability densities having the same *support*, i.e. the same zeros, are cancellative and have the Property (M). Based on this example of probability densities, the concept of separative valuation algebra will now be generalised

Example **8.17**. Consider probability densities (see Example 2.40). Here, $f \otimes g = f \otimes g'$ does not imply $g = g'$ since g and g' may differ whenever f is zero. Therefore, only densities which have the same support should be considered, i.e.

$$\text{supp}(f) = \{\mathbf{x} \in \mathbb{R}^x : x = d(f), f(\mathbf{x}) > 0\}. \tag{8.15}$$

Define

$$\text{supp}(f)^{\uparrow y} = \text{supp}(f) \times \mathbb{R}^{y-x} \tag{8.16}$$

for $d(f) \subseteq y$. Then, it is easily verified that for densities f, g with domains x and y

$$\text{supp}(f \otimes g) = \text{supp}(f)^{\uparrow x \cup y} \cap \text{supp}(g)^{\uparrow x \cup y}. \tag{8.17}$$

In order to simplify notation,

$$\text{supp}(f) = \text{supp}(g) \tag{8.18}$$

will be used for $\text{supp}(f)^{\uparrow x \cup y} = \text{supp}(g)^{\uparrow x \cup y}$ where $x = d(f)$ and $y = d(g)$. Furthermore, it is easily verified that

$$\text{supp}(f^{\downarrow s}) \supseteq \text{supp}(f) \tag{8.19}$$

for $s \subseteq d(f)$. Using (8.17), it is easily seen that $\text{supp}(f) = \text{supp}(f')$ and $\text{supp}(g) = \text{supp}(g')$ imply that

$$\text{supp}(f \otimes g) = \text{supp}(f' \otimes g') \text{ and } \text{supp}(f^{\downarrow s}) = \text{supp}(f'^{\downarrow s}) \tag{8.20}$$

for $s \subseteq d(f), d(f')$. Using equations (8.16) and (8.19),

$$\text{supp}(f \otimes f^{\downarrow s}) = \text{supp}(f). \tag{8.21}$$

Hence, supp is an idempotent congruence.

It is now shown that densities of the same support (modulo label) are cancellative. Let f, g_1, g_2 be densities with domains $s = d(f)$, $t_1 = d(g_1)$, $t_2 = d(g_2)$ such that $\text{supp}(f)^{\uparrow u} = \text{supp}(g_1)^{\uparrow u} = \text{supp}(g_2)^{\uparrow u}$ and $u = s \cup t_1 = s \cup t_2$. Assume $f \otimes g_1 = f \otimes g_2$. Then,

$$f(\mathbf{u}^{\downarrow s}) \cdot g_1(\mathbf{u}^{\downarrow t_1}) = f(\mathbf{u}^{\downarrow s}) \cdot g_2(\mathbf{u}^{\downarrow t_2})$$

for all $\mathbf{u} \in \mathbb{R}^u$. Since they have the same support, this implies

$$g_1(\mathbf{u}^{\downarrow t_1}) = g_2(\mathbf{u}^{\downarrow t_2})$$

for all $\mathbf{u} \in \mathbb{R}^u$, Hence, it has to be shown that $t_1 = t_2$. For fixed $\mathbf{t}_{12} \in \mathbb{R}^{t_1 \cap t_2}$, it holds that

$$g_1(\mathbf{t}_{12}, \mathbf{t}_1) \quad = \quad g_2(\mathbf{t}_{12}, \mathbf{t}_2),$$

for all $\mathbf{t}_1 \in \mathbb{R}^{t_1 - t_2}$, $\mathbf{t}_2 \in \mathbb{R}^{t_2 - t_1}$. This defines a function $g : \mathbb{R}^{t_1 \cap t_2} \to \mathbb{R}$ by

$$g(\mathbf{t}_{12}) \quad = \quad g_1(\mathbf{t}_{12}, \mathbf{t}_1) \quad = \quad g_2(\mathbf{t}_{12}, \mathbf{t}_2),$$

irrespective of the choice of $\mathbf{t}_1 \in \mathbb{R}^{t_1 - t_2}$, $\mathbf{t}_2 \in \mathbb{R}^{t_2 - t_1}$. Since g_1 and g_2 are densities,

$$\int_{\mathbf{z_1} \in \mathbb{R}^{t_1 - t_2}} g_1(\mathbf{x}, \mathbf{z_1}) d\mathbf{z_1} = g(\mathbf{x}) \int_{\mathbf{z_1} \in \mathbb{R}^{t_1 - t_2}} 1 d\mathbf{z_1} \quad < \quad \infty$$

and

$$\int_{\mathbf{z_2} \in \mathbb{R}^{t_2 - t_1}} g_2(\mathbf{x}, \mathbf{z_2}) d\mathbf{z_2} = g(\mathbf{x}) \int_{\mathbf{z_2} \in \mathbb{R}^{t_2 - t_1}} 1 d\mathbf{z_2} \quad < \quad \infty,$$

so $t_1 - t_2 = t_2 - t_1 = \emptyset$, thus $t_1 = t_2$. This shows that $g_1 = g_2$. Hence, the semigroups densities of the same support are indeed cancellative. ⊘

It is now shown how quotients of such generalised fractions can be built.

Lemma **8.18**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ be a valuation algebra with* full marginalisation *and let γ be an idempotent congruence in it such that the semigroups $\gamma(\phi)$ are* cancellative. *Then, the relation $\equiv^*$ in Φ^* defined by*

$$(\phi, \psi) \equiv^* (\phi', \psi') \quad \iff \quad \chi \otimes \phi \otimes \psi' = \chi \otimes \psi \otimes \phi' \text{ and } \gamma(\phi \otimes \psi) = \gamma(\phi' \otimes \psi') \tag{8.22}$$

for $\chi = \phi \otimes \psi \otimes \phi' \otimes \psi'$ is an equivalence relation. ⊘

PROOF. Let $(\phi_1, \psi_1), (\phi_2, \psi_2), (\phi_3, \psi_3) \in \Phi^*$.

1. *Reflexivity:* Commutativity of $\otimes$ implies that $\phi_1 \otimes \psi_1 = \psi_1 \otimes \phi_1$. Hence, $(\phi_1, \psi_1) \equiv^* (\phi_1, \psi_1)$.

2. *Symmetry:* Assume $(\phi_1, \psi_1) \equiv^* (\phi_2, \psi_2)$. Then,

$$\psi_2 \otimes \phi_1 = \phi_1 \otimes \psi_2 = \psi_1 \otimes \phi_2 = \phi_2 \otimes \psi_1$$

and $\gamma(\phi_1 \otimes \psi_1) = \gamma(\phi_2 \otimes \psi_2)$. Hence, $(\phi_2, \psi_2) \equiv^* (\phi_1, \psi_1)$.

3. *Transitivity:* Assume $(\phi_1, \psi_1) \equiv^* (\phi_2, \psi_2)$ and $(\phi_2, \psi_2) \equiv^* (\phi_3, \psi_3)$. Then,

$$\chi_1 \otimes \phi_1 \otimes \psi_2 = \chi_1 \otimes \psi_1 \otimes \phi_2, \tag{8.23}$$

$$\chi_2 \otimes \phi_2 \otimes \psi_3 = \chi_2 \otimes \psi_2 \otimes \phi_3 \tag{8.24}$$

where $\chi_1 = \phi_1 \otimes \psi_1 \otimes \phi_2 \otimes \psi_2$, $\chi_2 = \phi_2 \otimes \psi_2 \otimes \phi_3 \otimes \psi_3$ and $\gamma(\phi_1 \otimes \psi_1) = \gamma(\phi_2 \otimes \psi_2) = \gamma(\phi_3 \otimes \psi_3)$. Therefore, multiplying (8.23) by $\phi_3 \otimes \psi_3 \otimes \psi_3$ and applying (8.24) yields

$$\begin{aligned} \chi_1 \otimes \phi_1 \otimes \psi_2 \otimes \phi_3 \otimes \psi_3 \otimes \psi_3 &= \chi_1 \otimes \psi_1 \otimes \phi_2 \otimes \phi_3 \otimes \psi_3 \otimes \psi_3 \\ &= \chi_1 \otimes \psi_1 \otimes \psi_2 \otimes \phi_3 \otimes \phi_3 \otimes \psi_3. \end{aligned}$$

Hence, using cancellativity, the term $\phi_2 \otimes \psi_2 \otimes \psi_2$ can be erased,

$$(\phi_1 \otimes \psi_1 \otimes \phi_3 \otimes \psi_3) \otimes \phi_1 \otimes \psi_3 = (\phi_1 \otimes \psi_1 \otimes \phi_3 \otimes \psi_3) \otimes \psi_1 \otimes \phi_3.$$

Together with $\gamma(\phi_1 \otimes \psi_1) = \gamma(\phi_3 \otimes \psi_3)$, this shows that $\equiv^*$ is indeed transitive. □

COROLLARY **8.19**. *Let $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ be a valuation algebra with* full marginalisation *and let γ be an idempotent congruence in it such that the semigroups $\gamma(\phi)$ are* cancellative. *Then, the relation $=^*$ defined by*

$$\eta_1 =^* \eta_2 \iff \eta_1 \equiv^* \eta_2 \text{ and } d^*(\eta_1) = d^*(\eta_2) \tag{8.25}$$

is an equivalence relation in Φ^*. ⊘

PROOF. The elements of the same domain of the equivalence classes modulo $\equiv^*$ form a finer partition. □

REMARK **8.20**. The term χ in the definition of $\equiv^*$ in equation (8.22) is necessary for the transitivity of $=^*$. Take the following example: Let $(\phi_1, \psi_1) \equiv^* (\phi_2, \psi_2)$ and $(\phi_2, \psi_2) \equiv^* (\phi_3, \psi_3)$. It is then possible that $\gamma(\phi_1) < \gamma(\psi_1) = \gamma(\phi_1 \otimes \psi_1)$ and $\gamma(\psi_3) < \gamma(\phi_3) = \gamma(\phi_3 \otimes \psi_3)$. Here, $<$ stands for "$\leq$ and $\neq$." However, it then holds that $\gamma(\phi_1 \otimes \psi_3) < \gamma(\phi_3 \otimes \psi_1)$. Hence, the term χ carries $\phi_1 \otimes \psi_3$ and $\psi_1 \otimes \phi_3$ to the same equivalence class $\gamma(\phi_1 \otimes \psi_1) = \gamma(\phi_3 \otimes \psi_3)$. ⊘

It is now shown that probability densities have the Property (M).

EXAMPLE **8.21**. Let f, f_1, g be probability densities and let $t = d(f_1)$ and $x = d(f)$. Assume $f^{\downarrow t} = f_1 \otimes g$ such that $d(g) \subseteq t$. Define the function $f_2 : \mathbb{R}^x \rightarrow \mathbb{R}$ by

$$f_2(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x}^{\downarrow d(g)})}$$

whenever $g(\mathbf{x}^{\downarrow d(g)}) > 0$ and to be 0 else. This convention will also be followed in the following divisions. Then,

$$\int_{\mathbf{s} \in \mathbb{R}^{x-t}} f_2(\mathbf{s}, \mathbf{t}) d\mathbf{s} = \frac{\int_{\mathbf{s} \in \mathbb{R}^{x-t}} f(\mathbf{s}, \mathbf{t}) d\mathbf{s}}{g(\mathbf{t}^{\downarrow d(g)})} = \frac{f^{\downarrow t}(\mathbf{t})}{g(\mathbf{t}^{\downarrow d(g)})} = \frac{f_1 \otimes g(\mathbf{t})}{g(\mathbf{t}^{\downarrow d(g)})} = f_1(\mathbf{t})$$

for $\mathbf{t} \in \mathbb{R}^t$. Hence, $\int_{\mathbf{x} \in \mathbb{R}^x} f_2(\mathbf{x}) d\mathbf{x} = \int_{\mathbf{t} \in \mathbb{R}^t} f_1(\mathbf{t}) d\mathbf{t}$, which shows that f_2 is indeed a probability density. Clearly, it holds that

$$f(\mathbf{x}) = \frac{f(\mathbf{x}) \cdot g(\mathbf{x}^{\downarrow d(g)})}{g(\mathbf{x}^{\downarrow d(g)})} = f_2(\mathbf{x}) \cdot g(\mathbf{x}^{\downarrow d(g)}).$$

Since $f^{\downarrow t} = f_1 \otimes g$ implies $g(\mathbf{x}^{\downarrow d(g)}) = 0$ implies $f(\mathbf{x}) = 0$, it follows that $f = f_2 \otimes g$. Hence,

$$f^{\downarrow t} = (f_2 \otimes g)^{\downarrow t} = {f_2}^{\downarrow t} \otimes g$$

by the combination axiom. This shows that probability densities have the Property (M). ⊘

Based on the example of probability densities, the concept of valuation algebras can be generalised as follows.

DEFINITION **8.22**. *A valuation algebra* $\mathfrak{A} = (\Phi, D, d, \otimes, \downarrow)$ *is called* **separative** *if there is an* idempotent *congruence* γ *in it such that*

- *the semigroups* $\gamma(\phi)$ *are cancellative and if*
- $\mathfrak{A}$ *has the Property (M).* ⊘

These generalised separative valuation algebras can be embedded into a valuation algebra of quotients modulo $=^*$ in essentially the same way as above.

LEMMA **8.23**. *Let* $(\Phi, D, d, \otimes, \downarrow)$ *be a separative valuation algebra. The relation* $=^*$ *is a domain-contained congruence in* $(\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*)$. ⊘

PROOF. The claim can be proved in the same way as Lemma 8.9. □

THEOREM **8.24**. $(\Phi^*, D, d^*, \otimes^*, \mathcal{M}^*, \downarrow^*, =^*)$ *forms a quotient valuation algebra with division in the groups*

$$\gamma_x(\phi) = \{(\phi, \psi) \in \gamma(\phi) : d^*(\phi, \psi) = x\}.$$

Furthermore, the mapping $\phi \mapsto (\phi, e)$ *is an embedding.* ⊘

PROOF. The claim can be proved in the same way as Theorems 8.10 and 8.11. □

In this more general setting, Gaussian quotients form a subalgebra of the separative extension of *positive densities.* More precisely, they are included in the group $\gamma(\phi)$ of separative quotients of positive densities, which are at the bottom of the partial order induced by supp.

8.5 Conditionals in a Separative Extension

The notion of conditional Gaussian potentials from Chapter 5 can be generalised for any separative extension.

Definition **8.25**. *An element $\eta = (\phi, \phi^{\downarrow t}) \in \Phi^*$ is called a* conditional *in a separative extension $\mathfrak{A}^* = (\Phi^*, D, d, \otimes, \mathcal{M}, \downarrow)$. The variables $h = d(\phi) - t$ are called* head *of the conditional and the variables t are called* tail. *The set of all conditionals is denoted Φ_c^*.* ⊘

The following lemma gives a sufficient condition for the combination of two conditionals to be a conditional. The situation is shown in Figure 8.1: An oval stands for a conditional's domain and its grey-shaded part stands for the head.

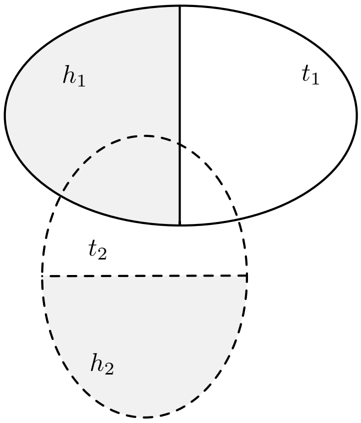


Figure 8.1: The combination of conditionals yields a conditional if head h_2 does not overlap with the other factor's domain $h_1 \cup t_1$.

Lemma **8.26**. *Let $\eta_1 = (\phi_1, {\phi_1}^{\downarrow t_1}) \in \Phi^*$ and $\eta_2 = (\phi_2, {\phi_2}^{\downarrow t_2}) \in \Phi^*$ be conditionals with heads $h_1 = d(\phi_1) - t_1$ and $h_2 = d(\phi_2) - t_2$ such that*

- *the head $h_2 = d(\phi_2) - t_2$ and $d(\phi_1)$ are disjoint, i.e. $h_2 \cap d(\phi_1) = \emptyset$, and*
- $\gamma({\phi_2}^{\downarrow t_2}) \leq \gamma({\phi_1}^{\downarrow t_1})$.

Then, $\eta_1 \otimes \eta_2$ is a conditional with head $h_1 \cup h_2$ and tail $t_1 \cup (t_2 - h_1)$, i.e. there is a $\phi \in \Phi$ with domain $d(\phi) = d(\phi_1) \cup d(\phi_2)$ such that $\eta_1 \otimes \eta_2 = (\phi, \phi^{\downarrow t_1 \cup (t_2 - h_1)})$. ⊘

Proof. Since $\eta_1 \otimes \eta_2 = (\phi_1 \otimes \phi_2, {\phi_1}^{\downarrow t_1} \otimes {\phi_2}^{\downarrow t_2})$, it follows that $t_1 \cup t_2 \subseteq h_1 \cup t_1 \cup t_2 \in \mathcal{M}(\eta_1 \otimes \eta_2)$. Then, in light of the combination axiom in $\mathfrak{A}$ [since $d(\phi_1) \cap d(\phi_2) \subseteq t_2$ and since $d(\phi_2) \cap (h_1 \cup t_1 \cup t_2) = t_2$],

$$(\eta_1 \otimes \eta_2)^{\downarrow h_1 \cup t_1 \cup t_2} = (\phi_1 \otimes {\phi_2}^{\downarrow t_2}, {\phi_1}^{\downarrow t_1} \otimes {\phi_2}^{\downarrow t_2}).$$

Then,

$$(\phi_1 \otimes {\phi_2}^{\downarrow t_2}) \otimes ({\phi_1}^{\downarrow t_1} \otimes {\phi_2}^{\downarrow t_2 - h_1}) = ({\phi_1}^{\downarrow t_1} \otimes {\phi_2}^{\downarrow t_2}) \otimes (\phi_1 \otimes {\phi_2}^{\downarrow t_2 - h_1})$$

and

$$\gamma(\phi_1 \otimes \phi_2{}^{\downarrow t_2} \otimes \phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1}) = \gamma(\phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1} \otimes \phi_1 \otimes \phi_2{}^{\downarrow t_2 - h_1})$$

since $\gamma(\phi_2{}^{\downarrow t_2 - h_1}) \leq \gamma(\phi_2{}^{\downarrow t_2} \leq \gamma(\phi_1{}^{\downarrow t_1}) \leq \gamma(\phi_1)$. Therefore,

$$(\eta_1 \otimes \eta_2)^{\downarrow h_1 \cup t_1 \cup t_2} = (\phi_1 \otimes \phi_2{}^{\downarrow t_2 - h_1}, \phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1})$$

Hence, $t_1 \cup (t_2 - h_1) \in \mathcal{M}((\eta_1 \otimes \eta_2)^{\downarrow h_1 \cup t_1 \cup t_2})$, and

$$\Big((\eta_1 \otimes \eta_2)^{\downarrow h_1 \cup t_1 \cup t_2}\Big)^{\downarrow t_1 \cup (t_2 - h_1)} = (\phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1}, \phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1}).$$

Then, by the transitivity axiom in $\mathfrak{A}^*$, $t_1 \cup (t_2 - h_1) \in \mathcal{M}(\eta_1 \otimes \eta_2)$, i.e. $\eta_1 \otimes \eta_2 = (\phi, \psi)$ such that $d(\psi) \subseteq t_1 \cup (t_2 - h_1)$ and

$$(\phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1}, \phi_1{}^{\downarrow t_1} \otimes \phi_2{}^{\downarrow t_2 - h_1}) = (\eta_1 \otimes \eta_2)^{\downarrow t_1 \cup (t_2 - h_1)} = (\phi^{\downarrow d(\phi) \cap (t_1 \cup (t_2 - h_1))}, \psi).$$

By cancellativity, $\psi = \phi^{\downarrow d(\phi) \cap (t_1 \cup (t_2 - h_1))}$. Since $d(\phi) \cap (t_1 \cup (t_2 - h_1)) = d(\psi) \subseteq t_1 \cup (t_2 - h_1)$, it also follows that $t_1 \cup (t_2 - h_1) \subseteq d(\phi)$. Hence,

$$\eta_1 \otimes \eta_2 = (\phi, \psi) = (\phi, \phi^{\downarrow t_1 \cup (t_2 - h_1)})$$

is a conditional. □

Hence, conditional Gaussian potentials

$$\mathcal{G}_c^* = \{(\phi, \phi^{\downarrow t}) \in \mathcal{G}^*\}. \tag{8.26}$$

are the conditionals in the separative extension of Gaussian potentials. It has to be remarked that not all representations of a conditional have the same form $(\phi, \phi^{\downarrow t})$, for instance $(\phi \otimes \phi, \phi \otimes \phi^{\downarrow t})$.

REMARK **8.27**. Conditional Gaussian potentials are closed under combination (see Theorem 7.8), whereas only Lemma 8.26 holds for conditionals in general. Furthermore, in contrast to Gaussian hints, conditional Gaussian potentials are not fully marginalisable. On the one hand, a non-vacuous variable in a Gaussian hint corresponds to a variable in the head of the corresponding conditional Gaussian potential. On the other hand, by using Lemma 7.13, a vacuous variable Z corresponds to a conditional Gaussian potential

$$(\phi_1, \phi_1{}^{\downarrow z - \{Z\}}) \otimes^* e_{\{Z\}} \tag{8.27}$$

such that $Z \notin d(\phi_1)$. Therefore, *in a separative extension, vacuous variables* in the sense of (8.27) *cannot be eliminated*. Would it be possible to extend marginalisation so that vacuous-variables can be eliminated? The answer is negative: In general, it is not possible to eliminate the tail variables t of a conditional η with head h. However, when η is marginalised to t, the variables t become vacuous. *In this general situation*, it is impossible to extend marginalisation to cover the elimination of vacuous variables the transitivity axiom would require that $\emptyset \in \mathcal{M}^*(\eta)$ implies that all marginals are defined, i.e. that $s \in \mathcal{M}^*(\eta)$ for all $s \subseteq h \cup t$. However, if every variable is either vacuous or non-vacuous, there is no contradiction with the transitivity axiom. It remains an open question whether this property is sufficient for extending marginalisation in Φ^* to cover vacuous variables. ⊘

Conditionals are *kernels* in a valuation algebra with division in the sense of Section 4.6. The elements $(\phi, e) \in \Phi^*$ are *densities* in that terminology. If a sequence of conditionals forms a *construction sequence*, it is factorisation of a density. This is a reformulation of the *Chain Rule* of *Bayesian Networks* where a network of conditional probability distributions is used to construct a full probability distribution. Further, it has been shown in Section 4.6 that there is always a scheduling of the collect algorithm if the factors of the join tree form a construction sequence.

Chapter Synopsis & Discussion

Separative valuation algebras offer an algebraic motivation for relating different conditional Gaussian densities. Furthermore, the derived combination and marginalisation operations correspond to combination and elimination of non-vacuous variables of conditional Gaussian densities and Gaussian hints as discussed in Chapter 7. However, two aspects of the general algebraic approach must be pointed out:

- Conditional Gaussian potentials are closed under combination (see Theorem 7.8), whereas conditionals in a separative extension are not generally closed under combination;
- vacuous variables cannot be eliminated in the separative extension, whereas Gaussian hints are fully marginalisable.

The notion of separative valuation algebras originates from (Kohlas, 2003), whose presentation differs in some minor aspects.

1. The Property (M) is not required in the definition of a separative valuation algebra, since the weaker transitivity axiom (A4)$''$ is used (see also the discussion at the end of Chapter 2).
2. A further minor difference is that he only considers pairs of valuations of the same domain and the same equivalence class, i.e. the set

   $$\Phi' = \{(\phi, \psi) : \ d(\phi) = d(\psi), \phi \equiv \psi \pmod{\gamma}\}.$$

   Obviously,

   $$\Phi' \subseteq \Phi^*.$$

   However, every element of $(\phi, \psi) \in \Phi^*$ can be represented by $(\phi \otimes \phi \otimes \psi, \phi \otimes \psi \otimes \psi) \in \Phi'$. Indeed, by the commutativity of combination,

   $$(\phi \otimes \phi \otimes \psi) \otimes \psi = (\phi \otimes \psi \otimes \psi) \otimes \phi$$

   and $\phi \otimes \phi \otimes \psi \equiv \phi \otimes \psi \otimes \psi \pmod{\gamma}$. This shows that the quotients defined in this chapter have more representatives. However, the quotient algebra is not essentially larger since quotients are in one-to-one correspondence. In particular, an element $\phi \in \Phi$ is represented by the pair $(\phi \otimes \phi, \phi) \in \Phi'$.

3. The construction presented in this chapter is based on the quotient valuation algebra theorem, whereas (Kohlas, 2003) gives a direct proof that the quotients form a valuation algebra.

The theory of separative valuation algebras is rooted in semigroup theory. Hewitt and Zuckerman (1956) have shown that a commutative semigroup $(\Phi, \otimes)$ can be embedded into a semigroup that is the union of disjoint groups if and only if the semigroup satisfies

$$\phi \otimes \psi = \phi \otimes \phi = \psi \otimes \psi \quad \Longrightarrow \quad \phi = \psi.$$

Commutative semigroups satisfying this property have been called *separative* (Clifford and Preston, 1967). On the one hand, Hewitt and Zuckerman (1956) have shown that a separative semigroup always decomposes into disjoint cancellative sub-semigroups. They then used the same construction as presented in this chapter to construct a semigroup that is the union of disjoint groups. On the other hand, if a semigroup Φ is embedded in the union of disjoint commutative groups G_i,

$$\Phi = \bigcup_i G_i,$$

the separativity condition $\phi \otimes \psi = \phi \otimes \phi = \psi \otimes \psi$ implies that $\phi, \phi \otimes \phi, \psi$, and $\psi \otimes \psi$ are in the same group G_i (since it is closed under $\otimes$) and hence

$$\phi = \phi^{-1} \otimes \phi \otimes \phi = \phi^{-1} \otimes \phi \otimes \psi = \psi.$$

In the domain of local computation, Lauritzen and Jensen (1997) claim that they investigated which extra assumptions are needed to introduce division in a valuation algebra. However, they do not construct the extension from a cancellative semigroup. In particular, they do not define marginalisation in the extension. They essentially describe the Lauritzen-Spiegelhalter architecture for local computation in valuation algebras which have been called *domain-free* in (Kohlas, 2003). However, they presuppose a valuation algebra which already decomposes into disjoint groups and whose unit elements form a semi-lattice.

9

Symmetric Gaussian Potentials

As introduced in Chapter 5, a conditional Gaussian density $\phi_{x|z}$ can be represented by a symmetric Gaussian potential: a vector and a symmetric matrix. The matrix is a "pseudo-concentration matrix" since it is the difference of the two concentration matrices of ϕ and $\phi^{\downarrow z}$.

Chapter Outline

This chapter is devoted to the relation of symmetric Gaussian potentials to the separative extension of Gaussian potentials and to Gaussian hints: In Section 9.1, it is shown that there is a bijection between Gaussian Quotients and symmetric Gaussian potentials. In Section 9.2, it is shown that there is an injection from Gaussian hints into symmetric Gaussian potentials. Hence, equivalent Gaussian fractions and equivalent Gaussian hints can be represented in a unique, canonical way by a symmetric Gaussian potential.

In Sections 9.3 to 9.4, it is shown how the operations of marginalisation and combination can be carried over from the separative extension of Gaussian potentials and from Gaussian hints to symmetric Gaussian potentials. In Section 9.5, it is then shown that symmetric Gaussian potentials form a valuation algebra extending both Gaussian hints and the separative extension of Gaussian potentials.

Finally, it is shown in Section 9.6 how generalised moment matrices and symmetric Gaussian potentials corresponding to a Gaussian linear system are related.

9.1 Relating Gaussian Quotients to Symmetric Gaussian Potentials

The following theorem shows that equivalent Gaussian fractions in $\mathcal{G}^*$ are related to the same symmetric Gaussian potential in Δ.

LEMMA **9.1**. *Let $(\phi_{11}, \phi_{12}) = (\phi_{21}, \phi_{22}) \in \mathcal{G}^*$ be equivalent Gaussian fractions,*

$$\phi_{11} = (\mu_{11}, K_{11}), \quad \phi_{12} = (\mu_{12}, K_{12}), \quad \phi_{21} = (\mu_{21}, K_{21}), \quad \phi_{22} = (\mu_{22}, K_{22}),$$

and let $x = d(\phi_{11}) \cup d(\phi_{12}) = d(\phi_{21}) \cup d(\phi_{22})$. *Then,*

$$K_{11}{}^{\uparrow x} - K_{12}{}^{\uparrow x} = K_{21}{}^{\uparrow x} - K_{22}{}^{\uparrow x}$$

and

$$K_{11}{}^{\uparrow x}\mu_{11}{}^{\uparrow x} - K_{12}{}^{\uparrow x}\mu_{12}{}^{\uparrow x} = K_{21}{}^{\uparrow x}\mu_{21}{}^{\uparrow x} - K_{22}{}^{\uparrow x}\mu_{22}{}^{\uparrow x}.$$

⊘

PROOF. It follows from $[\phi_{11}, \phi_{12}] = [\phi_{21}, \phi_{22}]$ that $\phi_{11} \otimes \phi_{22} = \phi_{12} \otimes \phi_{21}$, i.e.

$$K_{11}{}^{\uparrow x} + K_{22}{}^{\uparrow x} = K_{12}{}^{\uparrow x} + K_{21}{}^{\uparrow x}$$

and

$$K^{-1}\left(K_{11}{}^{\uparrow x}\mu_{11}{}^{\uparrow x} + K_{22}{}^{\uparrow x}\mu_{22}{}^{\uparrow x}\right) = K^{-1}\left(K_{21}{}^{\uparrow x}\mu_{21}{}^{\uparrow x} + K_{12}{}^{\uparrow x}\mu_{12}{}^{\uparrow x}\right)$$

for $K = K_{11}{}^{\uparrow x} + K_{22}{}^{\uparrow x} = K_{12}{}^{\uparrow x} + K_{21}{}^{\uparrow x}$. Then,

$$K_{11}{}^{\uparrow x} - K_{12}{}^{\uparrow x} = K_{21}{}^{\uparrow x} - K_{22}{}^{\uparrow x}$$

and

$$K_{11}{}^{\uparrow x}\mu_{11}{}^{\uparrow x} + K_{22}{}^{\uparrow x}\mu_{22}{}^{\uparrow x} = K_{21}{}^{\uparrow x}\mu_{21}{}^{\uparrow x} + K_{12}{}^{\uparrow x}\mu_{12}{}^{\uparrow x},$$

hence also

$$K_{11}{}^{\uparrow x}\mu_{11}{}^{\uparrow x} - K_{12}{}^{\uparrow x}\mu_{12}{}^{\uparrow x} = K_{21}{}^{\uparrow x}\mu_{21}{}^{\uparrow x} - K_{22}{}^{\uparrow x}\mu_{22}{}^{\uparrow x},$$

which proves the claim. □

In light of this lemma, equivalent Gaussian fractions are mapped to the same symmetric Gaussian potential. Therefore, a Gaussian quotient $\eta = (\phi_1, \phi_2)$ with $\phi_1 = (\mu_1, K_1) \in \mathcal{G}$, $\phi_2 = (\mu_2, K_2) \in \mathcal{G}$, $x = d(\phi_1)$, and $y = d(\phi_2)$ induces the symmetric Gaussian potential $i^*(\eta)$ by the mapping $i^* : \mathcal{G}^* \to \mathit{\Delta}$ defined by

$$i^*(\eta) = (\mu, K), \tag{9.1}$$

where

$$\mu = (K_1\mu_1)^{\uparrow x\cup y} - (K_2\mu_2)^{\uparrow x\cup y}, \quad K = K_1{}^{\uparrow x\cup y} - K_2{}^{\uparrow x\cup y}. \tag{9.2}$$

Notice that K is symmetric since both K_1, K_2 and thus $K_1{}^{\uparrow x\cup y}$ and $K_2{}^{\uparrow x\cup y}$ are symmetric.

REMARK **9.2**. In particular, a Gaussian potential (μ, K) is represented by the pair

$$(K\mu, K). \tag{9.3}$$

In this representation, combination only requires the addition of the pseudo-mean vectors and of the pseudo-covariance matrices. ⊘

Surjectivity

It will now be proved that i^* is surjective, i.e. that each symmetric Gaussian potential is the image under i^* of a Gaussian quotient. As a preliminary step, the following lemma shows that any symmetric matrix can be written as the difference of two symmetric positive definite matrices vacuously extended to the union of their domains. The proof is constructive: Two symmetric positive definite matrices are constructed whose difference is the given symmetric matrix.

LEMMA **9.3**. *Any symmetric matrix Σ can be written as the difference of two symmetric positive definite matrices Σ_1, Σ_2,*

$$d(\Sigma_1) \cup d(\Sigma_2) = d(\Sigma),$$

such that

$$\Sigma = {\Sigma_1}^{\uparrow d(\Sigma)} - {\Sigma_2}^{\uparrow d(\Sigma)}. \quad \oslash$$

PROOF. If $d(\Sigma) = \emptyset$, then $\Sigma_1 = \Sigma_2 = 0_{\emptyset,\emptyset}$ satisfy the claim since $0_{\emptyset,\emptyset} = 0_{\emptyset,\emptyset} - 0_{\emptyset,\emptyset}$. If $d(\Sigma) \neq \emptyset$, let $d(\Sigma) = \{X_1, \ldots, X_n\}$, and define

$$\Sigma^{(k)} = \Sigma^{\downarrow\{X_1,\ldots,X_k\}}$$

for $1 \leq k \leq n$. The proof then goes by induction over $k = 1, \ldots, n$, i.e. it will be proved that

(a) there are symmetric positive definite matrices $\Sigma_1^{(1)}, \Sigma_2^{(1)}$ with

$$d(\Sigma_1^{(1)}) = d(\Sigma_2^{(1)}) = \{X_1\}$$

such that

$$\Sigma^{(1)} = \Sigma_1^{(1)} - \Sigma_2^{(1)},$$

and

(b) if, for $1 \leq k < n$, there are symmetric positive definite matrices $\Sigma_1^{(k)}, \Sigma_2^{(k)}$ with

$$d(\Sigma_1^{(k)}) = d(\Sigma_2^{(k)}) = \{X_1, \ldots, X_k\}$$

such that

$$\Sigma^{(k)} = \Sigma_1^{(k)} - \Sigma_2^{(k)},$$

then there are symmetric positive definite matrices $\Sigma_1^{(k+1)}, \Sigma_2^{(k+1)}$ with

$$d(\Sigma_1^{(k+1)}) = d(\Sigma_2^{(k+1)}) = \{X_1, \ldots, X_{k+1}\}$$

such that

$$\Sigma^{(k+1)} = \Sigma_1^{(k+1)} - \Sigma_2^{(k+1)}.$$

Let $\Sigma^{(1)} = \big(\sigma\big)$, then define

$$\Sigma_1^{(1)} = \big(|\sigma| + \sigma + 1\big)\,,$$
$$\Sigma_2^{(1)} = \big(|\sigma| + 1\big)\,.$$

Since $\Sigma_1^{(1)}$ and $\Sigma_2^{(1)}$ are both positive definite, the claim (a) is proved.
Assume that, for some k, $1 \leq k < n$, there are positive definite matrices $\Sigma_1^{(k)}$, $\Sigma_2^{(k)}$ with $d(\Sigma_1^{(k)}), d(\Sigma_2^{(k)}) \subseteq \{X_1, \ldots, X_k\}$ such that

$$\Sigma^{(k)} = \Sigma_1^{(k)} - \Sigma_2^{(k)}.$$

The induction step (b) is now to construct from $\Sigma_1^{(k)}$, $\Sigma_2^{(k)}$ the matrices $\Sigma_1^{(k+1)}$, $\Sigma_2^{(k+1)}$. Let $\sigma_{ij} = \Sigma(i,j)$ be the elements of Σ. Then, define the diagonal matrix $\Sigma_1^{k,(k+1)}$ as

$$\begin{pmatrix} \sigma_{11}^2 + k + 2 & & & \\ & \ddots & & \\ & & \sigma_{kk}^2 + k + 2 & \\ & & & \sum_{j=1}^k \sigma_{j(k+1)}^2 + k + k \cdot |\sigma_{(k+1)(k+1)}| + \sigma_{(k+1)(k+1)} \end{pmatrix}$$

and

$$\Sigma_2^{k,(k+1)} = \sum_{j=1}^{k} \Sigma_2^{k,k+1,j}$$

where, for $j = 1, \ldots, k$,

$$\Sigma_2^{k,k+1,j} = \begin{pmatrix} I_{j-1} & 0_{j-1,1} & 0_{j-1,k-j} & 0_{j-1,1} \\ 0_{1,j-1} & \sigma_{jj}^2 + 2 & 0_{1,k-j} & -\sigma_{j(k+1)} \\ 0_{k-j,j-1} & 0_{k-j,1} & I_{k-j} & 0_{k-j,1} \\ 0 & -\sigma_{(k+1)j} & 0_{1,k-j} & \sigma_{j(k+1)}^2 + 1 + \big|\sigma_{(k+1)(k+1)}\big| \end{pmatrix}.$$

Define

$$\Sigma_1^{(k+1)} = \Sigma_1^{k\uparrow\{X_1,\ldots,X_{k+1}\}} + \Sigma_1^{k,(k+1)} \quad \text{and}$$
$$\Sigma_2^{(k+1)} = \Sigma_2^{k\uparrow\{X_1,\ldots,X_{k+1}\}} + \Sigma_2^{k,(k+1)}.$$

Then, it is readily verified that

$$\begin{aligned} \Sigma^{(k+1)} &= \Sigma^{(k)\uparrow\{X_1,\ldots,X_{k+1}\}} + \Sigma_1^{k,k+1} - \Sigma_2^{k,k+1} \\ &= \Sigma_1^{(k)\uparrow\{X_1,\ldots,X_{k+1}\}} + \Sigma_1^{k,k+1} - (\Sigma_2^{k\uparrow\{X_1,\ldots,X_{k+1}\}} + \Sigma_2^{k,k+1}) \\ &= \Sigma_1^{(k+1)} - \Sigma_2^{(k+1)}. \end{aligned}$$

It remains to be proved that $\Sigma_1^{(k+1)}, \Sigma_2^{(k+1)}$ are symmetric and positive definite. By the induction hypothesis, $\Sigma_1^{(k)}$ and $\Sigma_2^{(k)}$ are positive definite. Hence, according to

Lemma A.5, it is sufficient to prove that $\Sigma_1^{k,(k+1)}$ and $\Sigma_2^{k,(k+1)}$ are symmetric and positive definite. Even more, in order to prove that $\Sigma_2^{k,(k+1)}$ is symmetric and positive definite, it is sufficient to prove that the matrices $\Sigma_2^{k,(k+1),j}$ are symmetric and positive definite. First, it is shown that the matrix $\Sigma_1^{k,(k+1)}$ is symmetric and positive definite. Since $\Sigma_1^{k,(k+1)}$ is diagonal, it is symmetric, and in light of Lemma 14.2.1 of (Harville, 1997; p.211), $\Sigma_1^{k,(k+1)}$ is positive definite since all diagonal elements are positive (note that $k \geq 1$ by assumption). Therefore, the matrix $\Sigma_1^{k,(k+1)}$ is positive definite. It is now shown that the matrices $\Sigma_2^{k,(k+1),j}$ are symmetric and positive definite. Symmetry follows by definition. For $\mathbf{x} \in \mathbb{R}^{k+1}$,

$$\mathbf{x}'\Sigma_2^{k,(k+1),j}\mathbf{x} = (\sum_{i=1,i\neq j}^{k+1} \underbrace{\mathbf{x}_i^2}) + \underbrace{\mathbf{x}_j^2 - 2\mathbf{x}_j\sigma_{j(k+1)}\mathbf{x}_{k+1} + \mathbf{x}_{k+1}^2\sigma_{j(k+1)}^2}_{=(\mathbf{x}_j - \sigma_{j(k+1)}\mathbf{x}_{k+1})^2}$$
$$+ \underbrace{\mathbf{x}_j^2(\sigma_{jj}^2 + 1)} + \underbrace{\mathbf{x}_{k+1}^2 \left|\sigma_{(k+1)(k+1)}\right|}.$$

For any values $\mathbf{x}_i, \sigma_{j(k+1)}, \sigma_{(k+1)(k+1)}$ $(i \in \{1, \ldots, k+1\})$, all underbraced summands are non-negative. Furthermore, assuming $\mathbf{x}_i \neq 0$ for some $i \in \{1, \ldots, k\}$, then the summand $\mathbf{x}_i^2$ is strictly positive and thus the sum as well. Therefore, the matrices $\Sigma_2^{k,(k+1),j}$ are positive definite for $j \in \{1, \ldots, k\}$, and thus their sum $\Sigma_2^{k,(k+1)}$ is positive definite as well. This concludes the proof of the assertion (b), too. □

Notice that the decomposition of a symmetric matrix into the difference of two symmetric positive definite matrices need not be unique; for instance, the null matrix $0_x = K_x - K_x$ for any domain x and symmetric and positive definite matrix $K_x \in \mathbb{R}(x, x)$.
Using this decomposition, it can now be proved that i^* is surjective.

Lemma **9.4**. *The mapping $i^* : \mathcal{G}^* \to \Delta$ is* surjective. ⊘

Proof. It has to be shown that, for $\phi \in \Delta$, there is a $(\phi_1, \phi_2) \in \Phi^*$ such that

$$i^*(\phi_1, \phi_2) = \phi.$$

Let $\phi = (\mu, K) \in \Delta$ and $s = d(\phi)$. By Lemma 9.3, there are symmetric positive definite matrices K_1 and K_2 with $d(K_1) \cup d(K_2) = s$ such that

$$K = {K_1}^{\uparrow s} - {K_2}^{\uparrow s}.$$

Note that they are invertible by Corollary 14.2.11 (Harville, 1997; p.214). Let $x = d(K_1)$ and $y = d(K_2)$. Define

$$\mu_1 = {K_1}^{-1}\mu^{\downarrow x}$$

and

$$\mu_2 = {K_2}^{-1}\left(\mu^{\downarrow s-x}\right)^{\uparrow y}.$$

Then, since $x \cup y = s$,

$$\mu = \begin{pmatrix} \mu^{\downarrow x} \\ \mu^{\downarrow s-x} \end{pmatrix} = \left(\mu^{\downarrow x}\right)^{\uparrow s} + \left(\left(\mu^{\downarrow s-x}\right)^{\uparrow y}\right)^{\uparrow s} = (K_1\mu_1)^{\uparrow s} + (K_2\mu_2)^{\uparrow s}.$$

Define $\phi_1 = (\mu_1, K_1)$ and $\phi_2 = (\mu_2, K_2)$, which are both Gaussian potentials since K_1 and K_2 are both symmetric and positive definite. It has been shown that

$$i^*(\phi_1, \phi_2) = \phi,$$

hence i^* is indeed surjective. □

Injectivity

It will now be proved that i^* is injective, i.e. that no more than one Gaussian quotient is mapped to the same symmetric Gaussian potential. As a preliminary step, the following lemma is needed.

LEMMA **9.5**. *Let*

$$\Sigma = {\Lambda_1}^{\uparrow d(\Sigma)} - {\Lambda_2}^{\uparrow d(\Sigma)}, \qquad \Sigma = {N_1}^{\uparrow d(\Sigma)} - {N_2}^{\uparrow d(\Sigma)}$$

where Λ_1, Λ_2, N_1 and N_2 are symmetric positive definite matrices with $d(\Lambda_1), d(\Lambda_2) \subseteq d(\Sigma)$ and $d(N_1), d(N_2) \subseteq d(\Sigma)$. Then,

$$d(\Lambda_1) \cup d(N_2) = d(\Lambda_2) \cup d(N_1) \supseteq d(\Lambda_1) \cup d(\Lambda_2), d(N_1) \cup d(N_2).$$

If $d(\Lambda_1) \cup d(\Lambda_2) = d(N_1) \cup d(N_2)$, then even

$$d(\Lambda_1) \cup d(N_2) = d(\Lambda_2) \cup d(N_1) = d(\Lambda_1) \cup d(\Lambda_2) = d(N_1) \cup d(N_2).$$ ⊘

PROOF. Let $\Lambda_1, \Lambda_2, N_1, N_2$ as in the statement of the theorem, and let

$$S_1 = d(\Lambda_1) \cup d(N_2), \qquad S_2 = d(\Lambda_2) \cup d(N_1).$$

By Corollary 14.2.13 of (Harville, 1997; p.214), the diagonal elements of Λ_1 and Λ_2 are positive. By definition of matrix difference, for $X \in d(\Sigma)$,

$$\Sigma(X,X) = ({\Lambda_1}^{\uparrow d(\Sigma)})(X,X) - ({\Lambda_2}^{\uparrow d(\Sigma)})(X,X) = ({N_1}^{\uparrow d(\Sigma)})(X,X) - ({N_2}^{\uparrow d(\Sigma)})(X,X).$$

First, $\Sigma(X,X) > 0$ implies that

$$X \in d(\Lambda_1), d(N_1), \quad \text{thus} \quad X \in S_1, S_2.$$

Second, $\Sigma(X,X) < 0$ implies that

$$X \in d(\Lambda_2), d(N_2), \quad \text{thus} \quad X \in S_1, S_2.$$

Third, $\Sigma(X,X) = 0$ implies that

$$X \in d(\Lambda_1), d(\Lambda_2)$$

or

$$X \in d(N_1), d(N_2)$$

or

$$X \in d(\Lambda_1), d(\Lambda_2), d(N_1), d(N_2),$$

or

$$X \notin d(\Lambda_1), d(\Lambda_2), d(N_1), d(N_2),$$

thus X is either in both S_1, S_2 or in none of them. Since

$$S_1, S_2 \subseteq d(\Lambda_1) \cup d(\Lambda_2) \cup d(N_1) \cup d(N_2) \subseteq d(\Sigma),$$

it follows that $S_1 = S_2$. Since

$$d(N_2) \subseteq S_1 = S_2 = d(\Lambda_2) \cup d(N_1)$$

implies

$$d(N_1) \cup d(N_2) \subseteq S_2,$$

and, similarly,

$$d(\Lambda_2) \subseteq S_2 = S_1 = d(\Lambda_1) \cup d(N_2)$$

implies

$$d(\Lambda_1) \cup d(\Lambda_2) \subseteq S_1,$$

it follows that

$$d(N_1) \cup d(N_2), d(\Lambda_1) \cup d(\Lambda_2) \subseteq d(\Lambda_1) \cup d(N_2) = d(\Lambda_2) \cup d(N_1).$$

Furthermore,

$$d(N_1) \cup d(N_2) \cup d(\Lambda_1) \cup d(\Lambda_2) = S_1 = S_2.$$

If

$$d(N_1) \cup d(N_2) = d(\Lambda_1) \cup d(\Lambda_2),$$

then

$$d(N_1) \cup d(N_2) = d(\Lambda_1) \cup d(\Lambda_2) = d(N_1) \cup d(N_2) \cup d(\Lambda_1) \cup d(\Lambda_2) = S_1 = S_2.$$ □

Lemma **9.6**. *The mapping $i^* : \mathcal{G}^* \to \Delta$ is* injective. ⊘

Proof. It has to be shown that

$$i^*([\phi_{11}, \phi_{12}]) = \eta = i^*([\phi_{21}, \phi_{22}])$$

implies

$$[\phi_{11}, \phi_{12}] = [\phi_{21}, \phi_{22}].$$

Let

$$\phi_{11} = (\mu_{11}, K_{11}), \quad \phi_{12} = (\mu_{12}, K_{12}), \quad \phi_{21} = (\mu_{21}, K_{21}), \quad \phi_{22} = (\mu_{22}, K_{22}) \in \mathcal{G},$$

such that

$$i^*([\phi_{11}, \phi_{12}]) = \eta = i^*([\phi_{21}, \phi_{22}]).$$

Then,

$$K_{11}{}^{\uparrow s} - K_{12}{}^{\uparrow s} = K_{21}{}^{\uparrow s} - K_{22}{}^{\uparrow s} \tag{9.4}$$

and

$$K_{11}{}^{\uparrow s} \mu_{11}{}^{\uparrow s} - K_{12}{}^{\uparrow s} \mu_{12}{}^{\uparrow s} = K_{21}{}^{\uparrow s} \mu_{21}{}^{\uparrow s} - K_{22}{}^{\uparrow s} \mu_{22}{}^{\uparrow s} \tag{9.5}$$

for

$$s = d(\phi_{11}) \cup d(\phi_{21}) = d^*(\eta_1) = d^*(\eta_2) = d(\phi_{21}) \cup d(\phi_{22}).$$

On the one hand, in light of Lemma 9.5,

$$d(\phi_{11}) \cup d(\phi_{22}) = d(\phi_{12}) \cup d(\phi_{21}).$$

On the other hand, it follows from (9.4) that

$$K_{11}{}^{\uparrow s} + K_{22}{}^{\uparrow s} = K_{12}{}^{\uparrow s} + K_{21}{}^{\uparrow s}$$

and from (9.5) that

$$K_{11}{}^{\uparrow s} \mu_{11}{}^{\uparrow s} + K_{22}{}^{\uparrow s} \mu_{22}{}^{\uparrow s} = K_{21}{}^{\uparrow s} \mu_{21}{}^{\uparrow s} + K_{12}{}^{\uparrow s} \mu_{12}{}^{\uparrow s}.$$

Therefore,

$$\begin{aligned}
\phi_{11} \otimes \phi_{22} &= ((K_{11}{}^{\uparrow s} + K_{22}{}^{\uparrow s})^{-1}((K_{11}\mu_{11})^{\uparrow s} + (K_{22}\mu_{22})^{\uparrow s}), K_{11}{}^{\uparrow s} + K_{22}{}^{\uparrow s}) \\
&= ((K_{21}{}^{\uparrow s} + K_{12}{}^{\uparrow s})^{-1}((K_{21}\mu_{21})^{\uparrow s} + (K_{12}\mu_{12})^{\uparrow s}), K_{21}{}^{\uparrow s} + K_{12}{}^{\uparrow s}) \\
&= \phi_{12} \otimes \phi_{21}.
\end{aligned}$$

It has been proved that $[\phi_{11}, \phi_{12}] = [\phi_{21}, \phi_{22}]$. It follows that i^* is indeed injective.□

In summary, Lemma 9.4 and Lemma 9.6 show that i^* is a bijection.

Theorem **9.7**. *The mapping* $i^* : \mathcal{G}^* \to \Delta$ *is a* bijection. ⊘

Relating CGPs to Symmetric Gaussian Potentials

The following theorem shows that a conditional Gaussian potential induces a symmetric Gaussian potential whose pseudo-concentration matrix is non-negative definite.

Theorem **9.8**. *Let* $(\nu, C) \in \Delta$ *be a symmetric Gaussian potential,* $C \in \mathbb{R}(x \cup z, x \cup z)$*,* $\mu \in \mathbb{R}^{x \cup z}$*,* $x \cap z = \emptyset$*. Then, there is a Gaussian potential* $\phi = (\mu, K) \in \mathcal{G}$*,* $d(\phi) = x \cup z$*, such that*

$$i^*(\phi, \phi^{\downarrow z}) = (\nu, C)$$

if and only if

(1) *C is non-negative definite of rank* $\mathrm{r}(C) = |x|$,

(2) $C^{\downarrow x}$ *is symmetric and positive definite, and*

(3) $\nu \in \mathcal{C}(C)$*, i.e.* ν *is in the column space* $\mathcal{C}(C)$ *of* C. ⊘

Proof. On the one hand, assume $i^*(\phi, \phi^{\downarrow z}) = (\nu, C)$ for some $\phi = (\mu, K)$, i.e.

$$C = K - \left(K^{\downarrow z}\right)^{\uparrow x \cup z} + \left(K^{\downarrow z,x}(K^{\downarrow x})^{-1}K^{\downarrow x,z}\right)^{\uparrow x \cup z}$$

and

$$\nu = K\mu - \left((K^{\downarrow z} + K^{\downarrow z,x}(K^{\downarrow x})^{-1}K^{\downarrow x,z})\mu^{\downarrow z}\right)^{\uparrow x \cup z}.$$

Then, since $K^{\downarrow x}$ is symmetric and positive definite since it is a principal submatrix of the symmetric and positive definite matrix K and since

$$C = \begin{pmatrix} K^{\downarrow x} & K^{\downarrow x,z} \\ K^{\downarrow z,x} & K^{\downarrow z,x}{K^{\downarrow x}}^{-1}K^{\downarrow x,z} \end{pmatrix},$$

it follows from Lemma A.9 that C is symmetric and non-negative definite of rank $\mathrm{r}(K^{\downarrow x}) = |x|$. Finally,

$$\nu = K\mu - \left(K^{\downarrow z} + K^{\downarrow z,x}(K^{\downarrow x})^{-1}K^{\downarrow x,z})\right)^{\uparrow x \cup z}\mu = C\mu$$

shows that $\nu \in \mathcal{C}(C)$.
On the other hand, assume that (ν, C) satisfies (1)–(3) of the claim. Since every principal submatrix $C^{\downarrow z}$ of a symmetric and non-negative definite matrix C is symmetric and non-negative definite in light of Corollary 14.2.12 of (Harville, 1997; p.214), it follows from Lemma A.10 that

$$C^{\downarrow z} = C^{\downarrow z,x}{C^{\downarrow x}}^{-1}C^{\downarrow x,z}.$$

Define

$$K = \begin{pmatrix} C^{\downarrow x} & C^{\downarrow x,z} \\ C^{\downarrow z,x} & I_z + C^{\downarrow z} \end{pmatrix},$$

which is symmetric and positive definite in light of Lemma A.7 since $C^{\downarrow x}$ and

$$K^{\downarrow z} - K^{\downarrow z,x}{K^{\downarrow x}}^{-1}K^{\downarrow x,z} = I_z + C^{\downarrow z} - C^{\downarrow z,x}{C^{\downarrow x}}^{-1}C^{\downarrow x,z} = I_z$$

are both symmetric and positive definite. On the one hand, it then holds that

$$K - \left(K^{\downarrow z}\right)^{\uparrow x \cup z} + \left(K^{\downarrow z,x}{K^{\downarrow x}}^{-1}K^{\downarrow x,z}\right)^{\uparrow x \cup z} = \begin{pmatrix} C^{\downarrow x} & C^{\downarrow x,z} \\ C^{\downarrow z,x} & C^{\downarrow z,x}{C^{\downarrow x}}^{-1}C^{\downarrow x,z} \end{pmatrix} = C.$$

On the other hand, since $\nu \in \mathcal{C}(C)$, there is a μ such that

$$C\mu = \nu.$$

Then,

$$K\mu - \left(K^{\downarrow z} + K^{\downarrow z,x}(K^{\downarrow x})^{-1}K^{\downarrow x,z})\right)^{\uparrow x \cup z}\mu = C\mu = \nu.$$

This shows that, for $\phi = (\mu, K) \in \mathcal{G}$, indeed $i^*([\phi, \phi^{\downarrow z}]) = (\nu, C)$ where . □

9.2 Relating Gaussian Hints to Symmetric Gaussian Potentials

Gaussian linear systems can be related to symmetric Gaussian potentials by the mapping $e_{\mathfrak{L}} : \mathfrak{L} \to \Delta$, $\ell \mapsto e_{\mathfrak{L}}(\ell)$, $(A, z, K) \mapsto e_{\mathfrak{L}}(A, z, K)$, defined by

$$e_{\mathfrak{L}}(A, z, K) = (A'Kz, A'KA). \tag{9.6}$$

In a first step, the following lemma shows that the mapping $e_{\mathfrak{L}}$ maps equivalent Gaussian hints to the same symmetric Gaussian potential. More generally, it will be proved below that $e_{\mathfrak{L}}$ maps equivalent Gaussian linear systems to the same symmetric Gaussian potential.

Lemma **9.9**. *For $h_1, h_2 \in \mathcal{H}$,*

$$h_1 =_{\mathcal{H}} h_2 \quad \Longleftrightarrow \quad e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2). \qquad \oslash$$

Proof. Let $h_1, h_2 \in \mathcal{H}$. Then, both conditions, $h_1 =_{\mathcal{H}} h_2$ and $e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2)$, imply that $d(h_1) = d(h_2)$. So let $x = d(h_1) = d(h_2)$.
On the one hand, assume $h_1 =_{\mathcal{H}} h_2$. It has to be shown that $e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2)$. Let

$$h_1 = (A_1, z_1, K_1), \qquad A_1 \in \mathbb{R}(m, x), \quad z \in \mathbb{R}^m, \quad K \in \mathbb{R}(x, x),$$

for some $m \in \mathbb{N}$. Then, $h_1 =_{\mathcal{H}} h_2$ implies that there is a regular matrix $T \in \mathbb{R}(m, m)$ such that

$$h_2 = (TA_1, Tz_1, T^{-1'}K_1T^{-1}) = (A_2, z_2, K_2).$$

Then, in light of result (8.2.8) of (Harville, 1997; p.82), indeed

$$\begin{aligned} e_{\mathfrak{L}}(h_2) &= (A_1'T'T^{-1'}K_1T^{-1}Tz, A_1'T'T^{-1'}K_1T^{-1}TA_1) \\ &= (A_1'Kz, A'KA_1) \\ &= e_{\mathfrak{L}}(h_1). \end{aligned}$$

This shows the "only if" part of the lemma.
Conversely, assume $e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2)$. It has to be shown that $h_1 =_{\mathcal{H}} h_2$. Let $(\mu, K) = e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2)$,

$$h_1 = (A_1, z_1, K_1), \qquad A_1 \in \mathbb{R}(m_1, p), \quad z \in \mathbb{R}^{m_1}, \quad K_1 \in \mathbb{R}(m_1, m_1),$$

and

$$h_2 = (A_2, z_2, K_2), \qquad A_2 \in \mathbb{R}(m_2, p), \quad z \in \mathbb{R}^{m_2}, \quad K_2 \in \mathbb{R}(m_2, m_2), ,$$

for some $m_1, m_2 \in \mathbb{N}$ and let

$$r = \mathrm{r}(K).$$

Since A_1 and A_2 have full row rank m_1 and m_2, respectively, and since K_1 and K_2 are regular matrices of rank m_1 and m_2, respectively, and since A_1' and A_2' have full column rank m_1 and m_2, respectively, it follows in light of Lemma 8.3.2. of (Harville, 1997; p.83) that

$$\begin{aligned} m_1 = \mathrm{r}(A_1) \quad &= \quad \mathrm{r}(KA_1) \quad = \quad \mathrm{r}(A_1' K_1 A_1) \quad = \quad \mathrm{r}(K) \\ = \mathrm{r}(A_2' K_2 A_2) \quad &= \quad \mathrm{r}(K_2 A_2) \quad = \quad \mathrm{r}(A_2) \\ &= m_2, \end{aligned}$$

thus $r = m_1 = m_2$ and

$$\mathcal{R}(A_1) = \mathcal{R}(A_1' K_1 A_1) = \mathcal{R}(A_2' K_2 A_2) = \mathcal{R}(A_2),$$

i.e. the row spaces of A_1 and A_2 are equal. This means that every row of A_2 is a unique linear combination of the rows of A_1, so there is a regular matrix $T \in \mathbb{R}(r, r)$ such that

$$A_2 = TA_1.$$

In light of Lemma A.3, there is a regular matrix $B \in \mathbb{R}(r, r,)$ such that

$$K_2 = B' K_1 B.$$

Then, applying Lemma A.1 twice to

$$A_1' I_r K_1 I_r A_1 = A_1' K_1 A_1 = K = A_2' K_2 A_2 = A_1' T' B' K_1 B T A_1$$

yields $I_r = BT$, hence

$$K_2 = B' K_1 B = T^{-1'} K_1 T^{-1}.$$

It remains to be proved that $z_2 = Tz_1$. Since $A_1' K_1$ has full column rank r, Lemma A.1 can be applied to

$$A_1' K_1 z_1 = \mu = A_2' K_2 z_2 = A_1' T' T'^{-1} K_1 T^{-1} z_2 = A_1' K_1 T^{-1} z_2,$$

yielding that $z_1 = T^{-1} z_2$, thus $z_2 = Tz_1$. In summary, T establishes that indeed $h_1 =_{\mathcal{H}} h_2$. This also shows the "if" part of the lemma. □

The following lemma shows that one can pass from a GLS to a symmetric Gaussian potential either directly or via its associated hint.

LEMMA **9.10**. *Let $\ell = (A, z, K) \in \mathfrak{L}$ be a Gaussian linear system of $m \in \mathbb{N}$ equations on variables $x \in D$, where $A \in \mathbb{R}(m, x)$ of rank $r \leq m, |x|$, and $z \in \mathbb{R}^m$, and $K \in \mathbb{R}(m, m)$ symmetric and positive definite. Let $h = (\tilde{A}, \tilde{z}, \tilde{K})$ be the inferred Gaussian hint using the admissible basis $B = \begin{pmatrix} B_1 & B_2 \end{pmatrix}$, $B_1 \in \mathbb{R}(m, r)$, $B_2 \in \mathbb{R}(m, m - r)$. Then,*

$$e_{\mathfrak{L}}(\ell) = e_{\mathfrak{L}}(h).$$

⊘

PROOF. Define $T = B^{-1}$ and partition

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}, \qquad T_1 \in \mathbb{R}(r,m), \quad T_2 \in \mathbb{R}(m-r,m).$$

In light of Theorem 6.12,

$$\begin{aligned} \tilde{A} &= T_1 A, \\ \tilde{z} &= T_1 z + \big(B_1' K B_1\big)^{-1} (B_1' K B_2) T_2 z, \qquad \text{and} \\ \tilde{K} &= B_1' K B_1. \end{aligned}$$

Then, since $T_2 A = 0_{m-r,x}$,

$$\begin{aligned} A'KA &= (TA)'(B'KB)(TA) \\ &= \big((T_1A)' \quad 0_{x,m-r}\big) \begin{pmatrix} B_1'KB_1 & B_1'KB_2 \\ B_2'KB_1 & B_2'KB_2 \end{pmatrix} \begin{pmatrix} T_1 A \\ 0_{m-r,x} \end{pmatrix} \\ &= (T_1A)'(B_1'KB_1)T_1A \\ &= \tilde{A}'\tilde{K}\tilde{A} \end{aligned}$$

and

$$\begin{aligned} A'Kz &= (TA)'(B'KB)Tz \\ &= \big((T_1A)' \quad 0_{x,m-r}\big) \begin{pmatrix} B_1'KB_1 & B_1'KB_2 \\ B_2'KB_1 & B_2'KB_2 \end{pmatrix} \begin{pmatrix} T_1 z \\ T_2 z \end{pmatrix} \\ &= (T_1A)' \big((B_1'KB_1)T_1z + B_1'KB_2T_2z\big) \\ &= (T_1A)'(B_1'KB_1)\Big(T_1z + (B_1'KB_1)^{-1}B_1'KB_2T_2z\Big) \\ &= \tilde{A}'\tilde{K}\tilde{z}. \end{aligned}$$

This shows that indeed $e_{\mathfrak{L}}(\ell) = e_{\mathfrak{L}}(h)$. □

Gaussian linear systems are equivalent if and only if they induce the same symmetric Gaussian potential, as shown by the following theorem.

THEOREM **9.11**. *For $\ell_1, \ell_2 \in \mathcal{G}$*

$$e_{\mathfrak{L}}(\ell_1) = e_{\mathfrak{L}}(\ell_2) \iff \ell_1 =_{\mathfrak{L}} \ell_2,$$

and, in particular, for $h_1, h_2 \in \mathcal{H}$

$$e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2) \iff h_1 =_{\mathcal{H}} h_2.$$ ⊘

PROOF. Let $\ell_1, \ell_2 \in \mathfrak{L}$ and let h_1, h_2 be associated Gaussian hints. Then, by Lemma 9.10, $e_{\mathfrak{L}}(\ell_1) = e_{\mathfrak{L}}(h_1)$ and $e_{\mathfrak{L}}(\ell_2) = e_{\mathfrak{L}}(h_2)$. By Lemma 9.9, $e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2)$ if and only if $h_1 =_{\mathcal{H}} h_2$. Therefore, $e_{\mathfrak{L}}(\ell_1) = e_{\mathfrak{L}}(h_1) = e_{\mathfrak{L}}(h_2) = e_{\mathfrak{L}}(\ell_2)$ if and only if $h_1 =_{\mathcal{H}} h_2$ if and only if $\ell_1 =_{\mathfrak{L}} \ell_2$ by the definition of $=_{\mathfrak{L}}$. The second assertion is that of Lemma 9.9. □

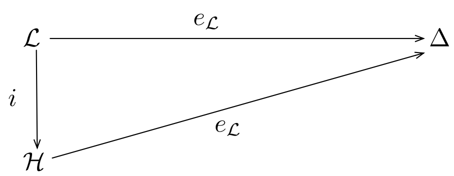


FIGURE 9.1: A symmetric Gaussian potentials can be either derived directly from a Gaussian linear system or by passing via a Gaussian hint.

As a consequence of this lemma, the diagram of Figure 9.1 is commutative.

The following theorem shows that Gaussian hints correspond to symmetric Gaussian potentials with non-negative definite pseudo-concentration matrix.

THEOREM **9.12**. *Let* $(\mu, C) \in \Delta$, $\mu \in \mathbb{R}^x$, $K \in \mathbb{R}(x,x)$. *Then, there is a Gaussian hint* $h = (A, z, K)$, $A \in \mathbb{R}(r,x)$, $z \in \mathbb{R}^r$, $K \in \mathbb{R}(r,r)$ *symmetric and positive definite such that*

$$e_{\mathfrak{L}}(h) = (\mu, C)$$

if and only if

- C *is non-negative definite of rank* r, *and*
- $\mu \in \mathcal{C}(C)$, *that is* μ *is in the column space of* C. ⊘

PROOF. First, the "only if" part is proved. Assume h is a Gaussian hint on $x = d(h)$ as stated above, and let $(\mu, C) = e_{\mathfrak{L}}(h)$, i.e.

$$C = A'KA, \quad \mu = A'Kz.$$

Since A' has full column rank and since K is regular, Lemma A.1 shows that $\mathrm{r}(A'KA) = \mathrm{r}(A'K) = \mathrm{r}(A') = \mathrm{r}(A) = r$, and hence $\mathcal{C}(A'KA) = \mathcal{C}(A'K) = \mathcal{C}(A')$. Therefore, C has rank r and $\mu \in \mathcal{C}(A'K) = \mathcal{C}(A'KA) = \mathcal{C}(C)$. Since K is positive definite,

$$\mathbf{x}'A'KA\mathbf{x} \quad = \quad (A\mathbf{x})'K(A\mathbf{x}) \quad \geq \quad 0$$

for every vector $\mathbf{x} \in \mathbb{R}^x$, thus $C = A'KA$ is symmetric non-negative definite. This shows the "only if" part.

Now the "if" part is proved. Assume $(\mu, C) \in \Delta$, $C \in \mathbb{R}(x,x)$ symmetric and positive definite of rank r, $\mu \in \mathcal{C}(C)$. By Theorem 14.3.7 of (Harville, 1997; p.218), a necessary and sufficient condition for a matrix $K \in \mathbb{R}(x,x)$ to be symmetric non-negative definite is that there is a matrix $A \in \mathbb{R}(r,x)$ of rank r such that

$$K = A'A.$$

Since $\mathrm{r}(C) = r = \mathrm{r}(A) = \mathrm{r}(A')$, it follows that $\mathcal{C}(C) = \mathcal{C}(A')$, hence, if $\mu \in \mathcal{C}(C) = \mathcal{C}(A')$, then there is a $z \in \mathbb{R}^r$ such that $A'z = \mu$. Therefore, $h = (A, z, I_r)$ is a Gaussian hint such that

$$e_{\mathfrak{L}}(h) = (A'I_r z, A'I_r A) = (A'z, A'A) = (\mu, C).$$

This also shows the "if" part of the lemma. □

In light of these considerations, conditional symmetric Gaussian potentials are those which correspond to a Gaussian hint or, equivalently, to a conditional Gaussian quotient. The set of conditional symmetric Gaussian potentials is denoted Δ_c,

$$\Delta_c = \{(\nu, C) \in \Delta : C \text{ non-negative definite, } \nu \in \mathcal{C}(C)\}. \tag{9.7}$$

This situation is shown in Figure 9.2.

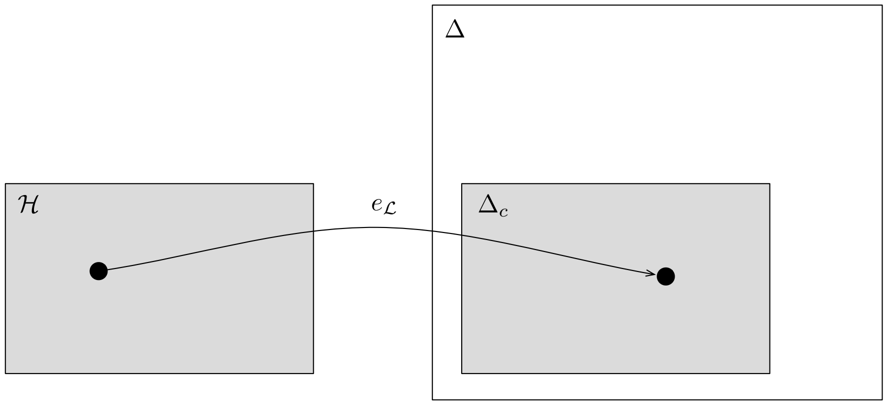


FIGURE 9.2: Gaussian hints and conditional symmetric Gaussian potentials are in one-to-one correspondence.

The following lemma shows that the focal sets of the Gaussian hints corresponding to a symmetric Gaussian potential can be directly given in terms of the latter.

LEMMA **9.13**. *Let $(\nu, C) \in \Delta_c$ be a conditional symmetric Gaussian potential on x. Then, it a represents a Gaussian hint with assumptions*

$$\Xi = \{\xi \in \mathbb{R}^x : \xi - \nu \in \mathcal{C}(C)\}$$

and focal sets

$$\Gamma(\xi) = \{\mathbf{x} \in \mathbb{R}^x : C\mathbf{x} + \xi = \nu\} \tag{9.8}$$

for $\xi \in \Xi$. ⊘

PROOF. Let $h = (A, z, K)$, $A \in \mathbb{R}(m, x)$ be a Gaussian hint inducing (ν, C), i.e. $(\nu, C) = e_{\mathfrak{L}}(h) = (A'KA, A'Kz)$. Then, the focal sets of h are

$$\Gamma_h(\omega) = \{\mathbf{x} : A\mathbf{x} + \omega = z\}$$

for $\omega \in \mathbb{R}^m$. Let $\omega \in \mathbb{R}^m$. Then, for $\xi(\omega) = A'K\omega$, it holds that

$$\Gamma(\xi(\omega)) = \{\mathbf{x} : C\mathbf{x} + \xi(\omega) = \nu\} = \{\mathbf{x} : A'KA\mathbf{x} + \xi(\omega) = A'Kz\} = \Gamma_h(\omega).$$

It remains to be proved that every $\xi \in \Xi$ can be written as $\xi = \xi(\omega)$ for some $\omega \in \mathbb{R}^m$. Indeed, $\xi \in \Xi$ implies $\xi = A'K(z - A\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^x$, hence $\xi \in \mathcal{C}(A'K)$ and thus there is some $\omega \in \mathbb{R}^m$ such that $\xi = A'K\omega = \xi(\omega)$. □

Example **9.14**. The situation of Lemma 9.13 is depicted in Figure 9.3 for $A = (1, -1)$, $K = (1)$ and $z = (1)$. Then, $C = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and $\nu = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. For instance, the assumption $\omega = -1$ yields $\xi(-1) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, thus $\Gamma(\xi(-1)) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 - \mathbf{x}_2 - 1 = 1, -\mathbf{x}_1 + \mathbf{x}_2 + 1 = -1\} = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_2 = \mathbf{x}_1 - 2\} = \Gamma_h(-1)$. The second equation in $C\mathbf{x} + \xi = \nu$ is obtained from the first one by changing the sign. Therefore, the admissible assumptions are on the straight line $\Xi = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_2 = -\mathbf{x}_1\}$. ⊘

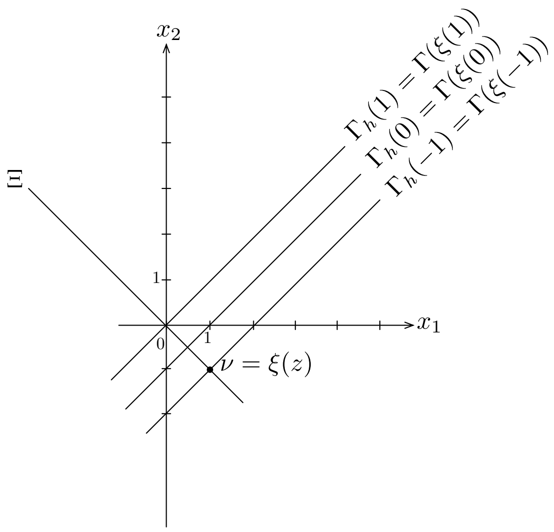


Figure 9.3: Focal sets in terms of the associated symmetric Gaussian potential

9.3 Combination of Symmetric Gaussian Potentials

Symm. Gaussian Potentials and the Combination of Gaussian Quotients

The following theorem shows that the combination of Gaussian quotients corresponds to the addition of the vector and the pseudo-concentration matrix of the corresponding symmetric Gaussian potentials.

Theorem **9.15**. *Let (ν_1, C_1) and (ν_2, C_2) be symmetric Gaussian potentials on domains x and y and let $\eta_1 = [\phi_{11}, \phi_{12}] \in \mathcal{G}^*$ and $\eta_2 = [\phi_{21}, \phi_{22}] \in \mathcal{G}^*$ be Gaussian quotients such that*

$$i^*(\eta_1) = (\nu_1, C_1), \qquad i^*(\eta_2) = (\nu_2, C_2).$$

Then,

$$i^*(\eta_1 \otimes^* \eta_2) = (\nu_1{}^{\uparrow u} + \nu_2{}^{\uparrow u}, C_1{}^{\uparrow u} + C_2{}^{\uparrow u}) \tag{9.9}$$

for $u = x \cup y$. ⊘

PROOF. Let

$$\phi_{11} = (\mu_{11}, K_{11}), \quad \phi_{12} = (\mu_{12}, K_{12}), \quad \phi_{21} = (\mu_{21}, K_{21}), \quad \phi_{22} = (\mu_{22}, K_{22}) \in \Phi.$$

Then,

$$\nu_1 = (K_{11}\mu_{11})^{\uparrow x} - (K_{12}\mu_{12})^{\uparrow x}, \qquad \nu_2 = (K_{21}\mu_{21})^{\uparrow y} - (K_{22}\mu_{22})^{\uparrow y},$$
$$C_1 = {K_{11}}^{\uparrow x} - {K_{12}}^{\uparrow x}, \, C_2 = {K_{21}}^{\uparrow y} - {K_{22}}^{\uparrow y}.$$

On the other hand,

$$\eta_1 \otimes^* \eta_2 = [\phi_{11} \otimes \phi_{12}, \phi_{21} \otimes \phi_{22}];$$

let

$$(\mu_1, K_1) = \phi_{11} \otimes \phi_{12}, \, (\mu_2, K_2) = \phi_{21} \otimes \phi_{22},$$

where

$$\mu_1 = {K_1}^{-1}\left((K_{11}\mu_{11})^{\uparrow x_1} + (K_{21}\mu_{21})^{\uparrow x_1}\right), \, \mu_2 = {K_2}^{-1}\left((K_{21}\mu_{21})^{\uparrow x_2} + (K_{22}\mu_{22})^{\uparrow x_2}\right)$$

and

$$K_1 = {K_{11}}^{\uparrow x_1} + {K_{21}}^{\uparrow x_1}, \, K_2 = {K_{12}}^{\uparrow x_2} + {K_{22}}^{\uparrow x_2}$$

for $x_1 = d(\phi_{11}) \cup d(\phi_{12})$ and $x_2 = d(\phi_{21}) \cup d(\phi_{22})$. Then, using the transitivity of vacuous extension,

$$\begin{aligned}
&(K_1\mu_1)^{\uparrow x \cup y} - (K_2\mu_2)^{\uparrow x \cup y} \\
=&(K_{11}\mu_{11})^{\uparrow x \cup y} + (K_{21}\mu_{21})^{\uparrow x \cup y} - (K_{21}\mu_{21})^{\uparrow x \cup y} - (K_{22}\mu_{22})^{\uparrow x \cup y} \\
=&(K_{11}\mu_{11})^{\uparrow x \cup y} - (K_{21}\mu_{21})^{\uparrow x \cup y} + (K_{21}\mu_{21})^{\uparrow x \cup y} - (K_{22}\mu_{22})^{\uparrow x \cup y} \\
=&{\nu_1}^{\uparrow x \cup y} + {\nu_2}^{\uparrow x \cup y}
\end{aligned}$$

and

$$\begin{aligned}
{K_1}^{\uparrow x \cup y} - {K_2}^{\uparrow x \cup y} &= {K_{11}}^{\uparrow x \cup y} + {K_{21}}^{\uparrow x \cup y} - {K_{21}}^{\uparrow x \cup y} - {K_{22}}^{\uparrow x \cup y} \\
&= {K_{11}}^{\uparrow x \cup y} - {K_{21}}^{\uparrow x \cup y} + {K_{21}}^{\uparrow x \cup y} - {K_{22}}^{\uparrow x \cup y} \\
&= {C_1}^{\uparrow x \cup y} + {C_2}^{\uparrow x \cup y}.
\end{aligned}$$

This shows that indeed $i^*(\eta_1 \otimes^* \eta_2) = ({\nu_1}^{\uparrow u} + {\nu_2}^{\uparrow u}, {C_1}^{\uparrow u} + {C_2}^{\uparrow u})$. □

The claim holds in particular for the combination of two conditional Gaussian potentials.

Symmetric Gaussian Potentials and the Combination of Gaussian Hints

The following theorem shows that combination of two Gaussian linear systems or their associated Gaussian hints induces the addition of the pseudo-mean and the pseudo-concentration of the associated symmetric Gaussian potentials. Since the combination of Gaussian hints and Gaussian quotients is compatible (see Theorem 7.8), this yields the same combination rule for symmetric Gaussian potentials as that carried over from Gaussian quotients in Theorem 9.15 above.

THEOREM **9.16**. *Let* $\ell_1 = (A_1, \mu_1, K_1) \in \mathfrak{L}$ *and* $\ell_2 = (A_2, \mu_2, K_2) \in \mathfrak{L}$ *be Gaussian linear systems on domains* x *and* y *and let*

$$e_{\mathfrak{L}}(\ell_1) = (\nu_1, C_1), \qquad e_{\mathfrak{L}}(\ell_2) = (\nu_2, C_2).$$

Then,

$$e_{\mathfrak{L}}(\ell_1 \oplus \ell_2) = ({\nu_1}^{\uparrow u} + {\nu_2}^{\uparrow u}, {C_1}^{\uparrow u} + {C_2}^{\uparrow u}) \tag{9.10}$$

where $u = x \cup y$*. In particular, if* $\ell_1, \ell_2 \in \mathcal{H}$*, then*

$$e_{\mathfrak{L}}(\ell_1 \otimes \ell_2) = ({\nu_1}^{\uparrow u} + {\nu_2}^{\uparrow u}, {C_1}^{\uparrow u} + {C_2}^{\uparrow u}). \tag{9.11}$$

⊘

PROOF. Let $A_1 \in \mathbb{R}(m_1, x)$, $A_2 \in \mathbb{R}(m_2, y)$. Define $m = m_1 + m_2$ and $A \in \mathbb{R}(m, x)$, $z \in \mathbb{R}^m$, and $K \in \mathbb{R}(m, m)$ by

$$A = \begin{pmatrix} {A_1}^{\uparrow x\cup y} \\ {A_2}^{\uparrow x\cup y} \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad K = \begin{pmatrix} K_1 & 0_{m_1,m_2} \\ 0_{m_2,m_1} & K_2 \end{pmatrix}.$$

Then,

$$\begin{aligned}
&A'KA \\
&= \begin{pmatrix} ({A_1}^{\downarrow x-y})'K_1{A_1}^{\downarrow x-y} & ({A_1}^{\downarrow x-y})'K_1{A_1}^{\downarrow x\cap y} & 0_{x-y,y-x} \\ ({A_1}^{\downarrow x\cap y})'K_1{A_1}^{\downarrow x-y} & \begin{bmatrix} ({A_1}^{\downarrow x\cap y})'K_1{A_1}^{\downarrow x\cap y} \\ +({A_2}^{\downarrow x\cap y})'K_2{A_2}^{\downarrow x\cap y} \end{bmatrix} & ({A_2}^{\downarrow x\cap y})'K_2{A_2}^{\downarrow y-x} \\ 0_{y-x,x-y} & ({A_2}^{\downarrow y-x})'K_2{A_2}^{\downarrow x\cap y} & ({A_2}^{\downarrow y-x})'K_2{A_2}^{\downarrow m_1,y-x} \end{pmatrix} \\
&= (A_1'K_1A_1)^{\uparrow u} + (A_2'K_2A_2)^{\uparrow u} \quad = \quad {C_1}^{\uparrow u} + {C_2}^{\uparrow u}
\end{aligned}$$

and

$$\begin{aligned}
A'Kz &= \begin{pmatrix} ({A_1}^{\downarrow x-y})'K_1z_1 \\ ({A_1}^{\downarrow x\cap y})'K_1z_1 + ({A_2}^{\downarrow x\cap y})'K_2z_2 \\ ({A_2}^{\downarrow y-x})'K_2z_2 \end{pmatrix} \\
&= (A_1'K_1z_1)^{\uparrow u} + (A_2'K_2z_2)^{\uparrow u} = {\nu_1}^{\uparrow u} + {\nu_2}^{\uparrow u}.
\end{aligned}$$

Finally, equation (9.11) follows since $(A, z, K) = \ell_1 \oplus \ell_2$ and $[\ell_1 \oplus \ell_2]_{\mathcal{H}} = \ell_1 \otimes \ell_2$. □

Notice that this theorem could have been derived directly from the results of Chapter 7.4 and Theorem 9.15.

Combination of Symmetric Gaussian Potentials

Therefore, define combination $\otimes : \Delta \times \Delta$, $(\delta_1, \delta_2) \mapsto \delta_1 \otimes \delta_2$, of symmetric Gaussian potentials (ν_1, C_1) and (ν_2, C_2) on domains x and y by

$$(\nu_1, K_1) \otimes (\nu_2, K_2) = \left({\nu_1}^{\uparrow x\cup y} + {\nu_2}^{\uparrow x\cup y}, {C_1}^{\uparrow x\cup y} + {C_2}^{\uparrow x\cup y}\right). \tag{9.12}$$

Combination of symmetric Gaussian potentials defined in this way is compatible with the combination of Gaussian hints and Gaussian quotients. This generalises the results from Chapter 7 for Gaussian quotients and Gaussian hints via the intermediate step of symmetric Gaussian potentials.

9.4 Marginalisation of Symmetric Gaussian Potentials

Symmetric Gaussian Potentials and Marginalisation of Gaussian Quotients

THEOREM **9.17**. *Let $\eta \in \mathcal{G}^*$ and $(\nu, C) = i^*(\eta)$. If $x \in \mathcal{M}^*(\eta)$, then*

$$i^*(\eta^{\downarrow^* x}) = (\nu^{\downarrow x} - C^{\downarrow x,y} {C^{\downarrow y}}^{-1} \nu^{\downarrow y}, C^{\downarrow x} - C^{\downarrow x,y} {C^{\downarrow y}}^{-1} C^{\downarrow y,x}). \tag{9.13}$$

where $y = d^*(\eta) - x$. ⊘

PROOF. If $x \in \mathcal{M}^*(\eta)$, then there is a $(\phi_1, \phi_2) \in \eta$ such that $d(\phi_2) \subseteq x \subseteq d(\phi_1) \cup d(\phi_2)$. Let $(\mu_1, K_1) = \phi_1$, and $(\mu_2, K_2) = \phi_2$.

By Corollary 14.2.12 of (Harville, 1997; p.214), every principal submatrix of a symmetric positive definite matrix is symmetric positive definite, thus ${K_1}^{\downarrow y}$ is symmetric positive definite and hence also invertible. On the one hand,

$$C = \begin{pmatrix} {K_1}^{\downarrow y} & {K_1}^{\rightarrow x,y} \\ {K_1}^{\rightarrow y,x} & {K_1}^{\rightarrow x} - {K_2}^{\uparrow x} \end{pmatrix},$$

hence, for $x' = x \cap d(\phi)$,

$$\begin{aligned} C^{\downarrow x} - C^{\downarrow x,y} {(C^{\downarrow y})}^{-1} C^{\downarrow y,x} &= {K_1}^{\rightarrow x} - {K_2}^{\uparrow x} - {({K_1}^{\rightarrow x,y} {({K_1}^{\downarrow y})}^{-1} {K_1}^{\rightarrow y,x})}^{\uparrow x} \\ &= {({K_1}^{\downarrow x'} - {K_1}^{\downarrow x',y} {({K_1}^{\downarrow y})}^{-1} {K_1}^{\downarrow y,x'})}^{\uparrow x} - {K_2}^{\uparrow x}. \end{aligned} \tag{9.14}$$

On the other hand,

$$(K_1\mu_1)^{\rightarrow x \cup y} = \begin{pmatrix} {K_1}^{\rightarrow y,x} {\mu_1}^{\rightarrow x} + {K_1}^{\downarrow y} {\mu_1}^{\downarrow y} \\ {K_1}^{\rightarrow x} {\mu_1}^{\rightarrow x} + {K_1}^{\rightarrow x,y} {\mu_1}^{\downarrow y} \end{pmatrix},$$

and

$$\nu = \begin{pmatrix} (K_1\mu_1)^{\rightarrow y} \\ (K_1\mu_1)^{\rightarrow x} - (K_2\mu_2)^{\uparrow x} \end{pmatrix},$$

it holds that

$$\begin{aligned} \nu^{\downarrow x} - C^{\downarrow x,y} {(C^{\downarrow y})}^{-1} \nu^{\downarrow y} =& (K_1\mu_1)^{\rightarrow x} - (K_2\mu_2)^{\uparrow x} - {K_1}^{\rightarrow x,y} {({K_1}^{\downarrow y})}^{-1} (K_1\mu_1)^{\rightarrow y,x} \\ =& {K_1}^{\rightarrow x} {\mu_1}^{\rightarrow x} + {K_1}^{\rightarrow x,y} {\mu_1}^{\downarrow y} - (K_2\mu_2)^{\uparrow x} \\ & - {K_1}^{\rightarrow x,y} {({K_1}^{\downarrow y})}^{-1} ({K_1}^{\rightarrow y,x} {\mu_1}^{\rightarrow x} + {K_1}^{\downarrow y} {\mu_1}^{\downarrow y}) \\ =& {(({K_1}^{\downarrow x'} - {K_1}^{\downarrow x',y} {({K_1}^{\downarrow y})}^{-1} {K_1}^{\downarrow y,x'}) {\mu_1}^{\downarrow x'})}^{\uparrow x} - (K_2\mu_2)^{\uparrow x}. \end{aligned} \tag{9.15}$$

Equations (86) and (9.15) show that indeed

$$i^*(\eta^{\downarrow^* x}) = (\nu^{\downarrow x} - C^{\downarrow x,y} {C^{\downarrow y}}^{-1} \nu^{\downarrow y}, C^{\downarrow x} - C^{\downarrow x,y} {C^{\downarrow y}}^{-1} C^{\downarrow y,x}).$$ □

Symmetric Gaussian Potentials and Marginalisation of Gaussian Hints

Since the columns in the design matrix of a GLS may be linearly dependent, eliminating a set of variables also eliminates the variables whose columns are linear combinations of the eliminated variables; these variables become vacuous. The following theorem shows that this corresponds to finding a regular submatrix of the pseudo-concentration matrix which makes the entries of the other variables to be eliminated zero, both in the mean vector and in the pseudo-concentration matrix.

THEOREM **9.18**. *Let $h = (A, z, K)$ be a Gaussian hint and let $(\nu, C) = e_{\mathfrak{L}}(h)$. For $x \subseteq d(h)$ there is a subset $y_2 \subseteq y = d(h) - x$ such that $C^{\downarrow y_2}$ is regular of rank $\mathrm{r}(C^{\downarrow y_2}) = \mathrm{r}(C^{\downarrow y})$. For each such set,*

$$e_{\mathfrak{L}}(h^{\downarrow x}) = (\nu^{\downarrow x} - C^{\downarrow x,y_2}{C^{\downarrow y_2}}^{-1}\nu^{\downarrow y_2}, C^{\downarrow x} - C^{\downarrow x,y}{C^{\downarrow y_2}}^{-1}C^{\downarrow y,x}). \tag{9.16}$$

⊘

PROOF. Notice that C is symmetric and non-negative definite in light of Theorem 9.12. Then, by Corollary 14.2.12 of (Harville, 1997; p.214), the principal submatrix $C^{\downarrow y}$ is symmetric non-negative definite, too. Let r be the rank of $C^{\downarrow y}$. Then, in light of Lemma A.11, there is a subset $y_2 \subseteq y$ of cardinality $|y_2| = r$ such that $C^{\downarrow y_2}$ is symmetric positive definite. This proves the first assertion of the theorem. In order to prove the second assertion, let $C^{\downarrow y_2}$ be a symmetric and positive definite submatrix of $C^{\downarrow y}$ of rank $r = \mathrm{r}(C^{\downarrow y_2}) = \mathrm{r}(C^{\downarrow y})$. Furthermore, since K is symmetric positive definite, in light of Corollary 14.3.13 of (Harville, 1997; p.219), there is a regular matrix $P \in \mathbb{R}(m, m)$ such that $K = P'P$. Define $\tilde{A} = PA$ and $\tilde{z} = Pz$. Then, since

$$P^{-1\prime}(P'P)P^{-1} = (P'^{-1}P')(PP^{-1}) = I_m,$$

it holds that $h =_{\mathcal{H}} (\tilde{A}, \tilde{z}, I_m) = \tilde{h}$ and, in light of Theorem 9.11,

$$C = \tilde{A}'\tilde{A}, \quad \nu = \tilde{A}'\tilde{z}.$$

Partition

$$\tilde{A} = (A_1, A_{21}, A_{22})$$

such that $A_1 \in \mathbb{R}(m, x), A_{21} \in \mathbb{R}(m, y_1), A_{22} \in \mathbb{R}(m, y_2)$ and let

$$A_2 = (A_{21}, A_{22}).$$

Then, since $C^{\downarrow y} = A_2'A_2$, $\mathrm{r}(A_2) \leq \mathrm{r}(C^{\downarrow y}) = r$. However, since the principal submatrix $C^{\downarrow y_2} = A_{22}'A_{22}$ has rank r, it follows that $r \leq \mathrm{r}(A_{22}) \leq \mathrm{r}(A_2)$, hence $\mathrm{r}(A_2) = r$. Let $B \in \mathbb{R}(m, m)$ be a regular matrix and partition

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$$

such that $B_1 \in \mathbb{R}(r, m)$ is a projection matrix for the variables y in $\tilde{h}$, i.e. $B_1 A_2 = 0_{r,y}$. In light of Gram-Schmidt orthogonalisation (Theorem 6.4.1. and Corollary 6.4.2

of (Harville, 1997; pp.63–65)), assume without loss of generality that B is orthonormal. Then,

$$\tilde{h}^{\downarrow x} = (B_1\tilde{z}, (B_1A_1), I_r).$$

and

$$e_{\mathfrak{L}}(h^{\downarrow x}) = e_{\mathfrak{L}}(\tilde{h}^{\downarrow x}) = ((B_1A_1)'B_1\tilde{z}, (B_1A_1)'(B_1A_1)).$$

Then,

$$B\tilde{A} = \begin{pmatrix} B_1A_1 & 0_{r,y_1} & 0_{r,y_2} \\ B_2A_1 & B_2A_{21} & B_2A_{22} \end{pmatrix}, \qquad (B\tilde{A})' = (\begin{pmatrix} (B_1A_1)' & (B_2A_1)' \\ 0_{y_1,r} & (B_2A_{21})' \\ 0_{y_2,r} & (B_2A_{22})' \end{pmatrix},$$

and, since $B'B = I_m$,

$$\begin{aligned} C &= \tilde{A}'\tilde{A} = \tilde{A}'B'B\tilde{A} = (B\tilde{A})'B\tilde{A} \\ &= \begin{pmatrix} (B_1A_1)'B_1A_1 + (B_2A_1)'B_2A_1 & (B_2A_1)'B_2A_{21} & (B_2A_1)'B_2A_{22} \\ (B_2A_{21})'B_2A_1 & (B_2A_{21})'B_2A_{21} & (B_2A_{21})'B_2A_{22} \\ (B_2A_{22})'B_2A_1 & (B_2A_{22})'B_2A_{21} & (B_2A_{22})'B_2A_{22} \end{pmatrix} \end{aligned}$$

and

$$\nu = \tilde{A}'\tilde{z} = \tilde{A}'B'B\tilde{z} = (B\tilde{A})'B\tilde{z} = (B\tilde{A})' \begin{pmatrix} B_1\tilde{z} \\ B_2\tilde{z} \end{pmatrix}.$$

Then, since B_2 has full row rank r and since A_{22} has full column rank r, in light of Lemma 8.3.2 (Harville, 1997; p.83), B_2A_{22} has full rank r, thus is regular. Therefore, in light of results (8.2.8) and (8.2.4) of (Harville, 1997; p.82),

$$\begin{aligned} & C^{\downarrow x} - C^{\downarrow x,y_2}(C^{\downarrow y_2})^{-1}C^{\downarrow y_2,x} \\ =& ((B_1A_1)'B_1A_1 + (B_2A_1)'B_2A_1) - (B_2A_1)'B_2A_{22}((B_2A_{22})'B_2A_{22})^{-1}(B_2A_{22})'B_2A_1 \\ =& (B_1A_1)'B_1A_1 + (B_2A_1)'B_2A_1 - (B_2A_1)'B_2A_{22}(B_2A_{22})^{-1}(B_2A_{22})'^{-1}(B_2A_{22})'B_2A_1 \\ =& (B_1A_1)'B_1A_1, \end{aligned}$$

and, for $z_1 = B_1\tilde{z}$ and $z_2 = B_2\tilde{z}$,

$$\begin{aligned} & \nu^{\downarrow x} - C^{\downarrow x,y_2}(C^{\downarrow y_2})^{-1}\nu^{\downarrow y_2} \\ =& ((B_1A_1)'z_1 + (B_2A_1)'z_2) - (B_2A_1)'B_2A_{22}((B_2A_{22})'B_2A_{22})^{-1}(B_2A_{22})'z_2 \\ =& (B_1A_1)'B_1\tilde{z}. \qquad \square \end{aligned}$$

By eliminating the variables y_2 in h, the variables $y_1 = y - y_2$ become *vacuous*, i.e.

$$h^{\downarrow x \cup y_1} = (\left(\nu^{\downarrow x} - C^{\downarrow x,y}C^{\downarrow y_2}{}^{-1}\nu^{\downarrow y_2}\right)^{\uparrow x \cup y_1}, \left(C^{\downarrow x} - C^{\downarrow x,y}C^{\downarrow y_2}{}^{-1}C^{\downarrow y,x}\right)^{\uparrow x \cup y_1}).$$

Marginalisation of Symmetric Gaussian Potentials

In light of these considerations, marginalisation of a symmetric Gaussian potential $(\nu, C) \in \Delta$ to x is defined, denoted

$$x \in \mathcal{M}(\nu, C), \tag{9.17}$$

if and only if there are $y_1 \cup y_2 = d(\nu, C) - x$, $y_1 \cap y_2 = \emptyset$ such that

1. $C^{\downarrow y_2}$ is positive definite and
2. $\nu^{\downarrow x \cup y_1} - C^{\downarrow x \cup y_1, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2} = \begin{pmatrix} 0_{y_1} \\ \nu^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2} \end{pmatrix}$ and
3. $C^{\downarrow x \cup y_1} - C^{\downarrow x \cup y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x \cup y_1} = \begin{pmatrix} 0_{y_1} & 0_{y_1, x} \\ 0_{x, y_1} & C^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} \end{pmatrix}$.

and then define

$$(\nu, C)^{\downarrow x} = (\nu^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2}, C^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x}). \tag{9.18}$$

It has to be verified that marginalisation of symmetric Gaussian potentials is well defined, i.e. that it does not depend on the particular choice of y_2.

LEMMA **9.19**. *Let $\phi = (\nu, C) \in \Delta$ be a symmetric Gaussian potential with domain $d(\phi) = x \cup y$, $x \cap y = \emptyset$. Assume that the marginal of ϕ to x is defined,*

$$(\nu, C)^{\downarrow x} = (\nu^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2}, C^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x}).$$

Then, for any $\tilde{y}_2 \subseteq y$ such that $C^{\downarrow \tilde{y}_2}$ is positive definite and $\mathrm{r}(C^{\downarrow \tilde{y}_2}) = \mathrm{r}(C^{\downarrow y}) = \mathrm{r}(C^{\downarrow y_2})$, it holds that

$$(\nu, C)^{\downarrow x} = (\nu^{\downarrow x} - C^{\downarrow x, \tilde{y}_2} {C^{\downarrow \tilde{y}_2}}^{-1} \nu^{\downarrow \tilde{y}_2}, C^{\downarrow x} - C^{\downarrow x, \tilde{y}_2} {C^{\downarrow \tilde{y}_2}}^{-1} C^{\downarrow \tilde{y}_2, x}).$$ ⊘

PROOF. Notice that, in light of Lemma A.10, $C^{\downarrow y_2}$ being symmetric and positive definite and

$$C^{\downarrow y_1} - C^{\downarrow y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, y_1} = 0_{y_1}$$

imply that $C^{\downarrow y} = C^{\downarrow y_1 \cup y_2}$ is symmetric and non-negative definite and that $\mathrm{r}(C^{\downarrow y_2}) = \mathrm{r}(C^{\downarrow y})$. Define

$$\tilde{C} = \begin{pmatrix} C^{\downarrow y} & C^{\downarrow y, x} \\ C^{\downarrow x, y} & C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} \end{pmatrix} \tag{9.19}$$

Then,

$$\begin{aligned}
&\tilde{C}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x \cup y_1} \\
&= \begin{pmatrix} \tilde{C}^{\downarrow y_1} - \tilde{C}^{\downarrow y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, y_1} & \tilde{C}^{\downarrow y_1, x} - \tilde{C}^{\downarrow y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x} \\ \tilde{C}^{\downarrow x, y_1} - \tilde{C}^{\downarrow y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, y_1} & \tilde{C}^{\downarrow x} - \tilde{C}^{\downarrow x, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x} \end{pmatrix} \\
&= \begin{pmatrix} C^{\downarrow y_1} - C^{\downarrow y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, y_1} & C^{\downarrow y_1, x} - C^{\downarrow y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} \\ C^{\downarrow x, y_1} - C^{\downarrow y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, y_1} & C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} \end{pmatrix} \\
&= 0_{x \cup y_1, x \cup y_1}.
\end{aligned}$$

Hence, in light of Lemma A.10, $\tilde{C}$ is symmetric and non-negative definite. Define

$$\tilde{\nu} = \begin{pmatrix} \nu^{\downarrow y} \\ 0_x \end{pmatrix}. \tag{9.20}$$

Then, $(\tilde{\nu}, \tilde{C})$ is a symmetric Gaussian potential, which corresponds to a Gaussian hint in light of Theorem 9.12. Therefore, in light of Theorem 9.18, choosing any $\tilde{y}_2 \subseteq y$ such that $\mathrm{r}(C^{\downarrow \tilde{y}_2}) = \mathrm{r}(C^{\downarrow y})$ leads to the same result $(0_x, 0_{x,x})$, i.e.

$$\tilde{C}^{\downarrow x} - \tilde{C}^{\downarrow x,\tilde{y}_2} \tilde{C}^{\downarrow \tilde{y}_2}{}^{-1} \nu^{\downarrow \tilde{y}_2} = 0_{x,x}$$

and

$$\tilde{\nu}^{\downarrow x} - \tilde{C}^{\downarrow x,\tilde{y}_2} \tilde{C}^{\downarrow \tilde{y}_2}{}^{-1} \tilde{C}^{\downarrow \tilde{y}_2,x} = 0_x,$$

hence

$$\begin{aligned} C^{\downarrow x,\tilde{y}_2} C^{\downarrow \tilde{y}_2}{}^{-1} \nu^{\downarrow \tilde{y}_2} &= \tilde{C}^{\downarrow x,\tilde{y}_2} \tilde{C}^{\downarrow \tilde{y}_2}{}^{-1} \nu^{\downarrow \tilde{y}_2} = 0_x = \tilde{C}^{\downarrow x,y_2} \tilde{C}^{\downarrow y_2}{}^{-1} \nu^{\downarrow y_2} \\ &= C^{\downarrow x,y_2} C^{\downarrow y_2}{}^{-1} \nu^{\downarrow y_2} \end{aligned}$$

and

$$\begin{aligned} C^{\downarrow x,\tilde{y}_2} C^{\downarrow \tilde{y}_2}{}^{-1} C^{\downarrow \tilde{y}_2,x} &= \tilde{C}^{\downarrow x,\tilde{y}_2} \tilde{C}^{\downarrow \tilde{y}_2}{}^{-1} \tilde{C}^{\downarrow \tilde{y}_2,x} = \tilde{C}^{\downarrow x,y_2} \tilde{C}^{\downarrow y_2}{}^{-1} \tilde{C}^{\downarrow y_2,x} \\ &= C^{\downarrow x,y_2} C^{\downarrow y_2}{}^{-1} C^{\downarrow y_2,x}. \end{aligned}$$

This shows that either y_2 or $\tilde{y}_2$ can be chosen for the marginalisation of ϕ to x. □

The following Lemma shows that marginalisation of symmetric Gaussian potentials and of Gaussian hints correspond, i.e. that $e_{\mathfrak{L}}$ is compatible with marginalisation.

LEMMA **9.20**. *Let $\phi = (\nu, C)$ be a symmetric Gaussian potential on $d(\phi) = x \cup y$, $x \cap y = \emptyset$. Then, $x \in \mathcal{M}(\phi)$ if and only if*

$$(\nu, C) = (\tilde{\nu}, \tilde{C}) \otimes (\nu_x, C_x)$$

such that $(\tilde{\nu}, \tilde{C}) = e_{\mathfrak{L}}(h)$ for some Gaussian hint on domain $d(h) = x \cup y$ and a symmetric Gaussian potential $\phi_x = (\nu_x, C_x) \in \Delta$ on domain $d(\phi_x) = x$. Furthermore,

$$(\nu, C)^{\downarrow x} = e_{\mathfrak{L}}(h^{\downarrow x}) \otimes (\nu_x, C_x).$$

⊘

PROOF. On the one hand, assume $x \in \mathcal{M}(\phi)$. Let $(\tilde{\nu}, \tilde{C})$ as defined in (9.20) and (9.19), which corresponds to a Gaussian hint. Define

$$\nu_x = \nu^{\downarrow x}, \qquad C_x = C^{\downarrow x} - C^{\downarrow x,y_2} C^{\downarrow y_2}{}^{-1} C^{\downarrow y_2,x}.$$

Then,

$$(\nu, C) = (\tilde{\nu}, \tilde{C}) \otimes (\nu_x, C_x).$$

Conversely, assume $(\nu, C) = (\tilde{\nu}, \tilde{C}) \otimes (\nu_x, C_x)$ such that $(\tilde{\nu}, \tilde{C}) = e_{\mathfrak{L}}(h)$ for some Gaussian hint on domain $d(h) = x \cup y$ and a symmetric Gaussian potential $\phi_x = (\nu_x, C_x) \in \Delta$ on domain $d(\phi_x) = x$. Then, in light of Theorem 9.12, $\tilde{C}$ is symmetric and non-negative definite and $\tilde{\nu} \in \mathcal{C}(\tilde{C})$. Choosing any subset $y_2 \subseteq y$ such that $\mathrm{r}(C^{\downarrow y}) = \mathrm{r}(C^{\downarrow y_2})$, then, in light of Theorem 9.18,

$$e_{\mathfrak{L}}(h^{\downarrow x_1 \cup y_1}) = (\tilde{\nu}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2}, \tilde{C}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x \cup y_1})$$

for $y_1 = y - y_2$. Furthermore, since $C^{\downarrow y}$, being a principal submatrix of a symmetric and non-negative definite matrix, is symmetric non-negative definite as well, Lemma A.9 shows that $C^{\downarrow y_2}$ is symmetric and positive definite. However, since the variables y_1 become vacuous in $h^{\downarrow x \cup y_1}$ by the definition of the marginalisation of Gaussian hints, the corresponding entries in the associated symmetric Gaussian potential are all 0, i.e.

$$\tilde{\nu}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2} = \begin{pmatrix} 0_{y_1} \\ \tilde{\nu}^{\downarrow x} - \tilde{C}^{\downarrow x, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2} \end{pmatrix}$$

and

$$\tilde{C}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x \cup y_1} = \begin{pmatrix} 0_{y_1} & 0_{y_1, x} \\ 0_{x, y_1} & \tilde{C}^{\downarrow x} - \tilde{C}^{\downarrow x, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x} \end{pmatrix}.$$

Therefore, marginalising $(\nu, C) = (\tilde{\nu}, \tilde{C} \otimes (\nu_x, C_x)$ to $x \cup y_1$,

$$\begin{aligned} \nu^{\downarrow x \cup y_1} - C^{\downarrow x \cup y_1, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2} &= (\tilde{\nu}^{\downarrow x \cup y_1} + \left(\nu^{\downarrow x}\right)^{\uparrow x \cup y_1}) - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2} \\ &= \begin{pmatrix} 0_{y_1} \\ \nu^{\downarrow x} + \tilde{\nu}^{\downarrow x} - \tilde{C}^{\downarrow x, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2} \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} & C^{\downarrow x \cup y_1} - C^{\downarrow x \cup y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x \cup y_1} \\ =& (\tilde{C}^{\downarrow x \cup y_1} + \left(C^{\downarrow x}\right)^{\uparrow x \cup y_1}) - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x \cup y_1} \\ =& \begin{pmatrix} 0_{y_1} & 0_{y_1, x} \\ 0_{x, y_1} & C^{\downarrow x} + \tilde{C}^{\downarrow x} - \tilde{C}^{\downarrow x, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{C}^{\downarrow y_2, x} \end{pmatrix}. \end{aligned}$$

Hence, indeed

$$\tilde{\nu}^{\downarrow x \cup y_1} - \tilde{C}^{\downarrow x \cup y_1, y_2} {\tilde{C}^{\downarrow y_2}}^{-1} \tilde{\nu}^{\downarrow y_2} = \begin{pmatrix} 0_{y_1} \\ \nu^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} \nu^{\downarrow y_2} \end{pmatrix}$$

and

$$\begin{pmatrix} & 0_{y_1} & 0_{y_1, x} \\ C^{\downarrow x \cup y_1} - C^{\downarrow x \cup y_1, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x \cup y_1} = & 0_{x, y_1} & C^{\downarrow x} - C^{\downarrow x, y_2} {C^{\downarrow y_2}}^{-1} C^{\downarrow y_2, x} \end{pmatrix}.$$

This shows that $x \in \mathcal{M}(\nu, C)$.
The second claim then follows from Theorem 9.18. □

Marginalisation of symmetric Gaussian potentials defined in this way extends that of both

- Gaussian quotients (since vacuous variables can be eliminated) and
- Gaussian hints (since Gaussian hints correspond only to $\Delta_c \subseteq \Delta$).

Furthermore, the algebra of symmetric Gaussian potentials has inverses in contrast to Gaussian hints; the only exception are vacuous Gaussian hints, which form a one-element group since they are idempotent.

9.5 Valuation Algebra of Symmetric Gaussian Potentials

In order to prove that marginalisation of symmetric Gaussian potentials is transitive, the following lemma will be needed.

Lemma **9.21**. *Let (ν, C) be a symmetric Gaussian potential with domain u such that $C^{\downarrow y}$ is positive definite. Let y be partitioned into $y = y_1 \cup y_2$, $y_1 \cap y_2 = \emptyset$. Let $x_1 = u - y_1$ and let $x = u - y = x_1 - y_2$. Let*

$$\nu_1 = \nu^{\downarrow x_1} - C^{\downarrow x_1} {C^{\downarrow y_1}}^{-1} \nu^{\downarrow y_1}$$

and

$$C_1 = C^{\downarrow x_1} - C^{\downarrow x_1, y_1} {C^{\downarrow y_1}}^{-1} C^{\downarrow y_1, x_1}.$$

Then,

$$\nu^{\downarrow x} - C^{\downarrow x} {C^{\downarrow y}}^{-1} \nu^{\downarrow y} = {\nu_1}^{\downarrow x} - {C_1}^{\downarrow x} {{C_1}^{\downarrow y_2}}^{-1} {\nu_1}^{\downarrow y_2}$$

and

$$C^{\downarrow x} - C^{\downarrow x,y} {C^{\downarrow y}}^{-1} C^{\downarrow y,x} = {C_1}^{\downarrow x} - {C_1}^{\downarrow x,y} {{C_1}^{\downarrow y}}^{-1} {C_1}^{\downarrow y,x}.$$

⊘

Proof. Define

$$\tilde{\nu} = \begin{pmatrix} \nu^{\downarrow y} \\ 0_x \end{pmatrix}, \qquad \tilde{C} = \begin{pmatrix} C^{\downarrow y} & C^{\downarrow y,x} \\ C^{\downarrow x,y} & I_x + C^{\downarrow x,y} {C^{\downarrow y}}^{-1} C^{\downarrow y,x} \end{pmatrix}, \tag{9.21}$$

where $\tilde{C}$ is a symmetric and positive definite matrix according to Lemma A.7. Hence, $(\tilde{\nu}, \tilde{C})$ is a Gaussian potential. Define

$$\tilde{\nu}_1 = \tilde{\nu}^{\downarrow x_1} - \tilde{C}^{\downarrow x_1, y_1} {\tilde{C}^{\downarrow y_1}}^{-1} \tilde{\nu}^{\downarrow y_1}$$

and

$$\tilde{C}_1 = \tilde{C}^{\downarrow x_1} - \tilde{C}^{\downarrow x_1, y_1} {\tilde{C}^{\downarrow y_1}}^{-1} \tilde{C}^{\downarrow y_1, x_1}.$$

By definition, it holds that

$$\nu = \tilde{\nu} + \nu^{\downarrow x \uparrow u}$$

and

$$C = \tilde{C} + \left(C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x}\right)^{\uparrow u},$$

and thus

$$\nu_1 = \tilde{\nu}^{\downarrow x_1} + \left(\nu^{\downarrow x}\right)^{\uparrow x_1} - \tilde{C}^{\downarrow x_1,y_1}{\tilde{C}^{\downarrow y_1}}^{-1}\tilde{\nu}^{\downarrow y_1} = \tilde{\nu}_1 + \left(\nu^{\downarrow x}\right)^{\uparrow x_1} \tag{9.22}$$

and

$$C_1 = \tilde{C}^{\downarrow x_1} + \left(C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x}\right)^{\uparrow x_1} - \tilde{C}^{\downarrow x_1,y_1}{\tilde{C}^{\downarrow y_1}}^{-1}\tilde{C}^{\downarrow y_1,x_1} \tag{9.23}$$

$$= \tilde{C}_1 + \left(C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x}\right)^{\uparrow x_1}. \tag{9.24}$$

Then, by the transitivity axiom holding in the valuation algebra of Gaussian potentials, it follows that

$$\tilde{\nu}^{\downarrow x} - \tilde{C}^{\downarrow x,y}{\tilde{C}^{\downarrow y}}^{-1}\tilde{\nu}^{\downarrow y} = \tilde{\nu}_1^{\downarrow x} - \tilde{C}_1^{\downarrow x,y_2}{\tilde{C}_1^{\downarrow y_2}}^{-1}\tilde{\nu}_1^{\downarrow y_2} \tag{9.25}$$

and

$$\tilde{C}^{\downarrow x} - \tilde{C}^{\downarrow x,y}{\tilde{C}^{\downarrow y}}^{-1}\tilde{C}^{\downarrow y,x} = \tilde{C}_1^{\downarrow x} - \tilde{C}_1^{\downarrow x,y_2}{\tilde{C}_1^{\downarrow y_2}}^{-1}\tilde{C}_1^{\downarrow y_2,x}. \tag{9.26}$$

Therefore, using (9.21), (9.25) and (9.22),

$$\begin{aligned}
\nu^{\downarrow x} - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}\nu^{\downarrow y} &= \tilde{\nu}^{\downarrow x} + \nu^{\downarrow x} - \tilde{C}^{\downarrow x,y}{\tilde{C}^{\downarrow y}}^{-1}\tilde{\nu}^{\downarrow y} \\
&= \nu^{\downarrow x} + \tilde{\nu}_1^{\downarrow x} - \tilde{C}_1^{\downarrow x,y_2}{\tilde{C}_1^{\downarrow y_2}}^{-1}\tilde{\nu}_1^{\downarrow y_2} \\
&= {\nu_1}^{\downarrow x} - {C_1}^{\downarrow x,y_2}{{C_1}^{\downarrow y_2}}^{-1}({\nu_1}^{\downarrow y_2} - (\left(\nu^{\downarrow x}\right)^{\uparrow x_1})^{\downarrow y_2}) \\
&= {\nu_1}^{\downarrow x} - {C_1}^{\downarrow x,y_2}{{C_1}^{\downarrow y_2}}^{-1}{\nu_1}^{\downarrow y_2}
\end{aligned}$$

and, using (9.21), (9.26) and (9.23),

$$\begin{aligned}
&C^{\downarrow x} - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x} \\
=&\tilde{C}^{\downarrow x} + C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x} - \tilde{C}^{\downarrow x,y}{\tilde{C}^{\downarrow y}}^{-1}\tilde{C}^{\downarrow y,x} \\
=&\tilde{C}_1^{\downarrow x} - \tilde{C}_1^{\downarrow x,y_2}{\tilde{C}_1^{\downarrow y_2}}^{-1}\tilde{C}_1^{\downarrow y_2,x} + C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x} \\
=&{C_1}^{\downarrow x} - (C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x}) - {C_1}^{\downarrow x,y_2}{{C_1}^{\downarrow y_2}}^{-1}{C_1}^{\downarrow y_2,x} \\
&+ (C^{\downarrow x} - I_x - C^{\downarrow x,y}{C^{\downarrow y}}^{-1}C^{\downarrow y,x}) \\
=&{C_1}^{\downarrow x} - {C_1}^{\downarrow x,y_2}{{C_1}^{\downarrow y_2}}^{-1}{C_1}^{\downarrow y_2,x}.
\end{aligned}$$

□

THEOREM **9.22**. *The algebraic structure* $(\Delta, D, d, \otimes, \mathcal{M}, \downarrow)$ *(defined in equations (9.12), (9.17), (9.18)) is a stable valuation algebra. It extends* $\mathcal{G}^*$ *and* $\mathcal{H}$. ⊘

PROOF. The axioms are verified in turn.

(A1) Associativity follows since vacuous extension and vector and matrix addition are associative. Commutativity follows since vector and matrix addition are commutative.

(A2) The labelling axiom is satisfied by definition.

(A3) The marginalisation axiom is satisfied by definition.

(A4) Let $\phi = (\nu, C)$ be a symmetric Gaussian potential with domain $d(\phi) = x \cup y$, $x \cap y = \emptyset$. Let x_1 be such that $x \subseteq x_1 \subseteq x \cup y$.
On the one hand, assume $x \in \mathcal{M}(\phi)$. Let $y_1 = d(\phi) - x_1$. Then, using Lemma 9.20,

$$(\nu, C) = e_{\mathfrak{L}}(h) \otimes (\nu_x, C_x)$$

for some Gaussian hint h with domain $d(h) = x \cup y$ and a symmetric Gaussian potential $\phi_x = (\nu_x, C_x) \in \Delta$ with domain x. Then,

$$(\nu, C) = e_{\mathfrak{L}}(h) \otimes ({\nu_x}^{\uparrow x_1}, {C_x}^{\uparrow x_1})$$

shows that $x_1 \in \mathcal{M}(\phi)$ and

$$(\nu, C)^{\downarrow x_1} = e_{\mathfrak{L}}(h^{\downarrow x_1}) \otimes ({\nu_x}^{\uparrow x_1}, {C_x}^{\uparrow x_1}) = e_{\mathfrak{L}}(h^{\downarrow x_1}) \otimes (\nu_x, C_x)$$

shows that $x \in \mathcal{M}(\phi^{\downarrow x_1})$ and, by the transitivity of the marginalisation of Gaussian hints,

$$(\nu, C)^{\downarrow x_1 \downarrow x} = e_{\mathfrak{L}}(h^{\downarrow x_1 \downarrow x}) \otimes (\nu_x, C_x) = e_{\mathfrak{L}}(h^{\downarrow x}) \otimes (\nu_x, C_x) = (\nu, C)^{\downarrow x}.$$

On the other hand, assume $x_1 \in \mathcal{M}(\phi)$ and $x \in \mathcal{M}(\phi^{\downarrow x_1})$ with $C^{\downarrow y_{12}}$ positive definite and

$$\nu_1 = \nu^{\downarrow x_1 \cup y_{11}} - C^{\downarrow x_1 \cup y_{11}, y_{12}} {C^{\downarrow y_{12}}}^{-1} \nu^{\downarrow y_{12}} = \begin{pmatrix} 0_{y_{11}} \\ \nu^{\downarrow x_1} - C^{\downarrow x_1, y_{12}} {C^{\downarrow y_{12}}}^{-1} \nu^{\downarrow y_{12}} \end{pmatrix}$$

and

$$\begin{aligned} C_1 &= C^{\downarrow x_1 \cup y_{11}} - C^{\downarrow x_1 \cup y_{11}, y_{12}} {C^{\downarrow y_{12}}}^{-1} C^{\downarrow y_{12}, x_1 \cup y_{11}} \\ &= \begin{pmatrix} 0_{y_{11}} & 0_{y_{11}, x_1} \\ 0_{x_1, y_{11}} & C^{\downarrow x_1} - C^{\downarrow x_1, y_{12}} {C^{\downarrow y_{12}}}^{-1} C^{\downarrow y_{12}, x_1} \end{pmatrix} \end{aligned}$$

and, for

$$(\nu, C)^{\downarrow x_1} = (\nu_{x_1}, C_{x_1}) = ({\nu_1}^{\downarrow x_1}, {C_1}^{\downarrow x_1}),$$

where ${C_{x_1}}^{\downarrow y_{22}}$ positive definite and

$${\nu_{x_1}}^{\downarrow x \cup y_{21}} - {C_{x_1}}^{\downarrow x \cup y_{21}, y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {\nu_{x_1}}^{\downarrow y_{22}} = \begin{pmatrix} 0_{y_{21}} \\ {\nu_{x_1}}^{\downarrow x} - {C_{x_1}}^{\downarrow x, y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {\nu_{x_1}}^{\downarrow y_{22}} \end{pmatrix}$$

and

$$\begin{aligned}
&{C_{x_1}}^{\downarrow x\cup y_{21}} - {C_{x_1}}^{\downarrow x\cup y_{21},y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {C_{x_1}}^{\downarrow y_{22},x\cup y_{21}} \\
&= \begin{pmatrix} 0_{y_{21}} & 0_{y_{21},x} \\ 0_{x,y_{21}} & {C_{x_1}}^{\downarrow x} - {C_{x_1}}^{\downarrow x,y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {C_{x_1}}^{\downarrow y_{22},x} \end{pmatrix}.
\end{aligned}$$

Partition

$$C = \begin{pmatrix}
C^{\downarrow y_{12}} & C^{\downarrow y_{12},y_{22}} & C^{\downarrow y_{12},y_{11}} & C^{\downarrow y_{12},y_{21}} & C^{\downarrow y_{12},x} \\
C^{\downarrow y_{22},y_{12}} & C^{\downarrow y_{22}} & C^{\downarrow y_{22},y_{11}} & C^{\downarrow y_{22},y_{21}} & C^{\downarrow y_{22},x} \\
C^{\downarrow y_{11},y_{12}} & C^{\downarrow y_{11},y_{22}} & C^{\downarrow y_{11}} & C^{\downarrow y_{11},y_{21}} & C^{\downarrow y_{11},x} \\
C^{\downarrow y_{21},y_{12}} & C^{\downarrow y_{21},y_{11},y_{22}} & C^{\downarrow y_{21},y_{11}} & C^{\downarrow y_{21}} & C^{\downarrow y_{21},x} \\
C^{\downarrow x,y_{12}} & C^{\downarrow x,y_{11},y_{22}} & C^{\downarrow x,y_{11}} & C^{\downarrow x,y_{21}} & C^{\downarrow x}
\end{pmatrix}.$$

Then,

$${C_1}^{\downarrow y_{22}} = C^{\downarrow y_{22}} - C^{\downarrow y_{22},y_{12}} {C^{\downarrow y_{12}}}^{-1} C^{\downarrow y_{12},y_{22}}.$$

Therefore, in light of Lemma A.7, $C^{\downarrow y_{12}\cup y_{22}}$ is positive definite if and only if $C^{\downarrow y_{12}}$ and ${C_1}^{\downarrow y_{22}}$ are both positive definite, where

$$C_1 = C^{\downarrow x_1} - C^{\downarrow x_1,y_{12}} {C^{\downarrow y_{12}}}^{-1} C^{\downarrow y_{12},x_1}.$$

Let $x_1 = x \cup y_{21} \cup y_{22}$. Then, by the definition of the marginalisation of symmetric Gaussian potentials,

$$(\nu, C)^{\downarrow x_1\cup y_{11}} = (\nu_1, C_1) = ({\nu_{x_1}}^{\uparrow x_1\cup y_{11}}, {C_{x_1}}^{\uparrow x_1\cup y_{11}}). \tag{9.27}$$

Hence, using Lemma 9.21 and (9.27), $(\nu, C)^{\downarrow x\cup y_{11}\cup y_{21}} = (\nu_{x\cup y_{11}\cup y_{21}}, C_{x\cup y_{11}\cup y_{21}})$ given by

$$\begin{aligned}
\nu_{x\cup y_{11}\cup y_{21}} &= \nu^{\downarrow x\cup y_{11}\cup y_{21}} - C^{\downarrow x\cup y_{11}\cup y_{21},y_{12}\cup y_{22}} {C^{\downarrow y_{12}\cup y_{22}}}^{-1} \nu^{\downarrow y_{12}\cup y_{22}} \\
&= {\nu_1}^{\downarrow x\cup y_{11}\cup y_{21}} - {C_1}^{\downarrow x\cup y_{11}\cup y_{21},y_{22}} {{C_1}^{\downarrow y_{22}}}^{-1} {\nu_1}^{\downarrow y_{22}} \\
&= {\nu_{x_1}}^{\uparrow x_1\cup y_{11}\downarrow x\cup y_{11}\cup y_{21}} \\
&\quad - {C_{x_1}}^{\uparrow x_1\cup y_{11}\downarrow x\cup y_{11}\cup y_{21},y_{22}} {{C_{x_1}}^{\uparrow x_1\cup y_{11}\downarrow y_{22}}}^{-1} {\nu_{x_1}}^{\uparrow x_1\cup y_{11}\downarrow y_{22}} \\
&= \left({\nu_{x_1}}^{\downarrow x\cup y_{21}} - {C_{x_1}}^{\downarrow x\cup y_{21},y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {\nu_{x_1}}^{\downarrow y_{22}}\right)^{\uparrow x\cup y_{11}\cup y_{12}} \\
&= \begin{pmatrix} 0_{y_{11}\cup y_{21}} \\ {\nu_{x_1}}^{\downarrow x} - {C_{x_1}}^{\downarrow x,y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {\nu_{x_1}}^{\downarrow y_{22}} \end{pmatrix}
\end{aligned}$$

and

$$\begin{aligned}
C_{x\cup y_{11}\cup y_{21}} &= C^{\downarrow x\cup y_{11}\cup y_{21}} - C^{\downarrow x\cup y_{11}\cup y_{21},y_{12}\cup y_{22}} {C^{\downarrow y_{12}\cup y_{22}}}^{-1} C^{\downarrow y_{12}\cup y_{22},x\cup y_{11}\cup y_{21}} \\
&= {C_1}^{\downarrow x\cup y_{11}\cup y_{21}} - {C_1}^{\downarrow x\cup y_{11}\cup y_{21},y_{22}} {{C_1}^{\downarrow y_{22}}}^{-1} {C_1}^{\downarrow y_{22},x\cup y_{11}\cup y_{21}} \\
&= \left({C_{x_1}}^{\downarrow x\cup y_{21}} - {C_{x_1}}^{\downarrow x\cup y_{11}\cup y_{21},y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {C_{x_1}}^{\downarrow y_{22},x\cup y_{21}}\right)^{\uparrow x\cup y_{11}\cup y_{21}} \\
&= \begin{pmatrix} 0_{y_{11}\cup y_{21}} & 0_{y_{11}\cup y_{21},x} \\ 0_{x,y_{11}\cup y_{21}} & {C_{x_1}}^{\downarrow x} - {C_{x_1}}^{\downarrow x,y_{22}} {{C_{x_1}}^{\downarrow y_{22}}}^{-1} {C_{x_1}}^{\downarrow y_{22},x} \end{pmatrix}.
\end{aligned}$$

Hence, $x \in \mathcal{M}(\nu, C)$ and $(\nu, C)^{\downarrow x} = (\nu, C)^{\downarrow x_1 \downarrow x}$.

(A5) Let $\phi = (\nu_1, C_1)$ and $\psi = (\nu_2, C_2)$ be symmetric Gaussian potentials with domains $x = d(\phi)$ and $y = d(\psi)$ and let $u = x \cup y$. Let

$$(\tilde{\nu}, \tilde{C}) = \phi \otimes \psi = ({\nu_1}^{\uparrow u} + {\nu_2}^{\uparrow u}, {C_1}^{\uparrow u} + {C_2}^{\uparrow u}).$$

Let s be such that $x \subseteq s \subseteq u$. Since

- $\tilde{C}^{\downarrow u-s} = {C_2}^{\downarrow u-s}$,
- $\tilde{C}^{\downarrow s, u-s} = \left({C_2}^{\downarrow s\cap y, u-s}\right)^{\uparrow s, u-s}$,
- $\tilde{\nu}^{\downarrow u-s} = {\nu_2}^{\downarrow u-s}$,

$s \in \mathcal{M}(\phi \otimes \psi)$ if and only if $s \cap y \in \mathcal{M}(\psi)$, and then also $(\phi \otimes \psi)^{\downarrow s} = \phi \otimes \psi^{\downarrow s}$.

(A6) Let $(\nu, C) \in \Delta$ with domain $d^{\Delta}(\nu, C) = x$. Then,

- $\diamond = 0_{\emptyset,\emptyset}$ is symmetric and positive definite,
- $\nu^{\downarrow x} - C^{\downarrow x,\emptyset} C^{\downarrow \emptyset^{-1}} C^{\downarrow \emptyset, x} = \begin{pmatrix} 0_\emptyset \\ \nu^{\downarrow x} - C^{\downarrow x,\emptyset} C^{\downarrow \emptyset^{-1}} C^{\downarrow \emptyset, x} \end{pmatrix} = \nu$ and
- $C^{\downarrow x} - C^{\downarrow x,\emptyset} C^{\downarrow \emptyset^{-1}} C^{\downarrow \emptyset, x} = \begin{pmatrix} 0_\emptyset & 0_{\emptyset,x} \\ 0_{x,\emptyset} & C^{\downarrow x} - C^{\downarrow x,\emptyset} C^{\downarrow \emptyset^{-1}} C^{\downarrow \emptyset, x} \end{pmatrix} = C$

shows that $x \in \mathcal{M}(\nu, C)$ and $(\nu, C)^{\downarrow x} = (\nu, C)$.

(A7) $e^* = (\diamond, \diamond)$ is the identity element since, for $(\nu, C) \in \Delta$ with domain $d^{\Delta}(\nu, C) = x$,

$$e^* \otimes (\nu, C) = (\diamond^{\uparrow x} + \nu, \diamond^{\uparrow x} + C) = (\nu, C) = (\nu + \diamond^{\uparrow x}, C + \diamond^{\uparrow x}) = (\nu, C) \otimes e_x.$$

(A8) The symmetric Gaussian potentials $e_x = (0_x, 0_{x,x})$ are neutral elements since, for $(\nu, C) \in \Delta$ with domain $d^{\Delta}(\nu, C) = x$,

$$e_x \otimes (\nu, C) = (0_x + \nu, 0_{x,x} + C) = (\nu, C) = (\nu + 0_x, C + 0_{x,x}) = (\nu, C) \otimes e^*.$$

(A9) For $e_x = (0_x, 0_{x,x})$ and $y \subseteq x$, it holds that $e_x = ({0_y}^{\uparrow x}, {0_{y,y}}^{\uparrow x})$, hence $y \in \mathcal{M}(e_x)$ and

$${e_x}^{\downarrow y} = (0_y, 0_{y,y}).$$

The extension properties follow from Theorems 9.15, 9.16, 9.17 and 9.18. □

9.6 Partially Swept Moment Matrices

As discussed in Section 3.5, the combination of moment matrices is the sum of their fully swept forms. A "fully swept" matrix $(0, 0)$ is thus the neutral element of combination. However, there is obviously no moment matrix which can be swept forward to $(0, 0)$, since its variance would have to be infinite. Therefore, (Dempster, 1990a) suggests a generalisation of moment matrices to *partially swept moment matrices*, which do neither need to have a fully swept nor a completely unswept form. It is now shown that Gaussian linear systems can be represented by such partially swept moment matrices. It turns out that such partially swept moment matrices have a close resemblance to (conditional) symmetric Gaussian potentials.

Consider a Gaussian linear system

$$x_1 = Ax_2 + \mu_1 + \omega \tag{9.28}$$

where ω is a Gaussian term with variance Σ. Then, (Dempster, 1990a) suggests the representation

$$M(x_1, \vec{x_2}) = \left(\begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & A \\ A' & 0 \end{pmatrix} \right) \tag{9.29}$$

where the x_1 are unswept and the components x_2 are swept forward. Notice that the components x_2 cannot be unswept.

This representation is a semantically motivated extension of moment matrices as follows (cf. the interpretation of moment matrices in Section 3.5).

- The conditional variance of x_1 given "$x_2 = \mathbf{x}_2$" is Σ and the conditional mean is $\mu_1 + A\mathbf{x}_2$. Hence, the variables x_1 are unswept and the corresponding entries in the matrix are μ_1 and Σ. The regression matrix A is put at x_1, x_2 in the moment matrix.
- The knowledge on x_2 is vacuous. Hence, the variables x_2 are fully swept with 0 in the vector and the matrix.

The partially swept matrix $M(x_1, \vec{x_2})$ can be fully swept forward, which results in

$$\begin{aligned} M(\vec{x_1}, \vec{x_2}) &= \triangleright(M(x_1, \vec{x_2}), x_1 = 0) \\ &= \left(\begin{pmatrix} \Sigma^{-1}\mu_1 \\ -A'\Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} -\Sigma^{-1} & \Sigma^{-1}A \\ A'\Sigma^{-1} & -A'\Sigma^{-1}A \end{pmatrix} \right). \end{aligned} \tag{9.30}$$

The form (9.29) has been called *maximally marginal representation* (Dempster, 1990a) because no more variables can be unswept. However, since head and tail need not be unique, there may be several maximally marginal representations. These may be obtained by passing through the form (9.30), which has been called *maximally conditioned representation* (Dempster, 1990a) because no more variables can be swept forward.

The symmetric Gaussian potential corresponding to the Gaussian linear system (9.28) is

$$(\nu, C) = \left(\begin{pmatrix} \Sigma^{-1}\mu_1 \\ -A'\Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma^{-1} & -\Sigma^{-1}A \\ -A'\Sigma^{-1} & A'\Sigma^{-1}A \end{pmatrix} \right),$$

which is the same as $M(\vec{x_1}, \vec{x_2})$ up to the sign in the pseudo-concentration matrix, i.e.

$$M(\vec{x_1}, \vec{x_2}) = (\nu, -C).$$

This resemblance is remarkable since the derivation of symmetric Gaussian potentials is analytically motivated, whereas that of generalised moment matrices is semantically motivated. Furthermore, it is now shown that the combination and marginalisation of generalised moment matrices as defined in (Dempster, 1990a) and of symmetric Gaussian potentials are compatible. Define

$$\Delta(\nu, -C) = (\nu, C). \tag{9.31}$$

For conditional symmetric Gaussian potentials $(\nu_1, C_1), (\nu_2, C_2)$, it holds that

$$\begin{aligned}(\nu_1, C_1) \otimes (\nu_2, C_2) &= (\nu_1 + \nu_2, C_1 + C_2) = (\nu_1 + \nu_2, -((-C_1) + (-C_2)) \\ &= \Delta(\Delta^{-1}(\nu_1, C_1) \oplus \Delta^{-1}(\nu_2, C_2)).\end{aligned} \tag{9.32}$$

Hence, combination is the *sum $\oplus$ of fully swept moment matrices* (after vacuous extension if necessary). This corresponds to the combination of generalised moment matrices as defined in (Dempster, 1990a).

Consider the conditional symmetric Gaussian potentials (ν, C) and decompose according to variables x_1 and x_2,

$$(\nu, C) = \left(\begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}, \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \right),$$

Assume $\mathrm{r}(C_{22}) = |x_2|$. Then

$$(\nu, C)^{\downarrow x_1} = \left(\nu_1 - C_{12}{C_{22}}^{-1}\nu_2, C_{11} - C_{12}{C_{22}}^{-1}C_{21} \right)$$

In terms of the corresponding partially swept generalised moment matrix, $\Delta^{-1}(\nu, C) = (\nu, -C)$, it holds that

$$\triangleleft((\nu, -C), x_2 = 0) = \left(\begin{pmatrix} \nu_1 - C_{12}{C_{22}}^{-1}\nu_2 \\ {C_{22}}^{-1}\nu_2 \end{pmatrix}, \begin{pmatrix} -C_{11} + C_{12}{C_{22}}^{-1}\nu_2 & -C_{12}{C_{22}}^{-1} \\ -C_{21}{C_{22}}^{-1} & {C_{22}}^{-1} \end{pmatrix} \right),$$

Hence, marginalisation corresponds to the elimination of unswept components or of swept vacuous components. Again, this corresponds to the marginalisation as defined in (Dempster, 1990a).

This proves the following theorem.

THEOREM **9.23**. *Generalised moment matrices corresponding to Gaussian linear systems form a valuation algebra which is isomorphic to the valuation algebra of conditional symmetric Gaussian potentials.* ⊘

The importance of the sweeping operator as a conceptual and computational tool in classical statistics was pointed out by (Goodnight, 1979).

Chapter Synopsis

By carrying the operations of combination and marginalisation over from Gaussian quotients and Gaussian hints to symmetric Gaussian potentials, the following correspondences are obtained:

- Precise Gaussian hints – Gaussian potentials – symmetric Gaussian potentials with positive definite pseudo-concentration matrix – moment matrices;
- Gaussian hints – conditional Gaussian potentials – symmetric Gaussian potentials with non-negative definite pseudo-concentration matrix (and mean vector in its column space) – generalised moment matrices corresponding to a Gaussian linear system;
- Gaussian quotients – symmetric Gaussian potentials.

Gaussian hints and Gaussian quotients are embedded in the valuation algebra of symmetric Gaussian potentials. Although marginalisation is only partially defined, conditional symmetric potentials are fully marginalisable as is the case for Gaussian hints.

Discussion

The representation of Gaussian hints or Gaussian quotients by symmetric Gaussian potentials is *canonical*: Equivalent Gaussian hints and equivalent Gaussian fractions map to the same symmetric Gaussian potential. Furthermore, it is simple to infer the symmetric Gaussian potential from a Gaussian linear system.

The following points highlight the naturalness of symmetric Gaussian potentials.

- For a GLS whose design matrix has full column rank, the inferred symmetric Gaussian potential has close resemblance to the Gaussian hint (see Theorem 6.23).
- Algebraically, the combination of Gaussian potentials corresponds to matrix addition and division to matrix subtraction, respectively. In this view, symmetric Gaussian potentials are the closure of Gaussian potentials under these two operations.
- Analytically, a symmetric Gaussian potential represents the quotient function of two Gaussian densities.

On the other hand, only a subset of symmetric Gaussian potentials has a clear semantic interpretation, namely conditional symmetric Gaussian potentials. However, the inverses of conditional Gaussian potentials are used in the Lauritzen-Spiegelhalter architecture for local computation.

Symmetric Gaussian potentials combine the advantages of both Gaussian hints and Gaussian quotients:

- full marginalisation in Gaussian hints and

- division in Gaussian quotients

are carried over to symmetric Gaussian potentials. A variable can be eliminated if

- the corresponding diagonal element of the pseudo-concentration matrix is positive, or if
- the corresponding element of the marginal mean vector and the corresponding row and column in the pseudo-concentration matrix are all zero.

The first case corresponds exactly to the rule for Gaussian quotients; the second case generalises the elimination of vacuous variables in Gaussian hints.

Finally, ordinary Gaussian potentials can be represented by symmetric Gaussian potentials with different complexity of the operations. On the one hand, combination is only the sum of the vectors and of two matrices with no inversion being required. On the other hand, the computation of the mean vector is more expensive. However, the required matrix has already to be computed for the marginal concentration matrix, so the overhead is essentially only a matrix-vector product.

Part III

Deterministic Knowledge

10 Deterministic Knowledge

Direct observations of variables are essential in the reasoning of Bayesian networks. In a Gaussian linear model, this corresponds to linear equations without a Gaussian term. Of course, assumption-based reasoning can then be applied to such Gaussian linear systems with deterministic knowledge and a hint can be derived. Therefore, algorithms for the assumption-based inference, for combination and for marginalisation will be developed in this chapter.

Chapter Outline

In Section 10.1, directly observed variables are introduced, taking a fixed value. In Section 10.2, general linear systems without a Gaussian term are added to the language of Gaussian linear information. The effect of both types of deterministic knowledge is analysed in terms of Gaussian linear systems and their associated symmetric Gaussian potentials. Symmetric Gaussian potentials with deterministic knowledge are formally defined in Section 10.3. Furthermore, combination can be defined in accordance with Dempster's Rule, and marginalisation of these potentials is derived from the projection of the corresponding focal sets (Section 10.4 and 10.5). In Section 10.6, it is shown that symmetric Gaussian potentials with deterministic linear equations form a valuation algebra.

In Section 10.7, the Gaussian linear belief function approach is presented in its full generality, by generalising the *partially swept moment matrices* for probabilistic and vacuous knowledge discussed in Section 9.6. First, moment matrices and deterministic variables (Liu, 1996a; 1999) are sketched. Then, the approach of (Dempster, 1990a; Liu et al., 2003a;b; Srivastava and Liu, 2003) is briefly outlined; it also takes deterministic equations into account. For this purpose, they use partially swept and completely unswept moment matrices. Combination and marginalisation are then only partially defined, namely if the necessary sweeping operations are defined.

10.1 Deterministic Variables

Observing Variables

If a real-valued variable X is directly observed and reported to take the value $\mathbf{x}$, this information can be captured in a deterministic hint

$$\mathcal{O}_{\mathbf{x}} = (\mathbb{R}^0, \Delta_\diamond, \Gamma_{\mathbf{x}}, \mathbb{R}^x) \tag{10.1}$$

for $x = \{X\}$ and $\Gamma_{\mathbf{x}} : \mathbb{R}^0 \to 2^{\mathbb{R}^x}$,

$$\Gamma_{\mathbf{x}}(\diamond) = \{\mathbf{x}\}. \tag{10.2}$$

Here follow two remarks about the combination of such deterministic hints by Dempster's Rule.

- Deterministic hints are *idempotent*, i.e. asserting the same piece of information several times yields nothing new, $\mathcal{O}_{\mathbf{x}} \otimes \mathcal{O}_{\mathbf{x}} = \mathcal{O}_{\mathbf{x}}$.
- The combination of different observations of the same variable yields a contradiction since $\Gamma_{\mathbf{x}}(\diamond) \cap \Gamma_{\mathbf{x}'}(\diamond) = \emptyset$ if $\mathbf{x} \neq \mathbf{x}'$. In order to close the combination of such hints, null elements z_x have to be added to the algebra for all domains $x \in D$ without specifying them further. Using this notation,

$$\mathcal{O}_{\mathbf{x}} \otimes \mathcal{O}_{\mathbf{x}'} = \begin{cases} \mathcal{O}_{\mathbf{x}} & \text{if } \mathbf{x} = \mathbf{x}' \\ z_x & \text{otherwise.} \end{cases}$$

This is something new: Without deterministic equations, the combination of Gaussian hints never yields a contradiction, and the valuation algebra of Gaussian hints does not need to be closed by contradictory elements.

A direct observation induces an event in a Gaussian hint. Technically, the observed value has to be substituted in every equation for the placeholder variable X. However, by doing so, the system may not have full row rank any more. For instance, when a value $\mathbf{x}_1$ is substituted for X_1 in the Gaussian linear equations

$$\begin{aligned} X_1 + 0.5 \cdot X_2 + \omega_1 &= z_1, \\ 0.5 \cdot X_1 + X_2 + \omega_2 &= z_2 \end{aligned}$$

for Gaussian assumptions ω_1, ω_2 and some $z_1, z_2 \in \mathbb{R}$, the system becomes

$$\begin{aligned} 0.5 \cdot X_2 + \omega_1 &= z_1 - \mathbf{x}_1, \\ X_2 + \omega_2 &= z_2 - 0.5 \cdot \mathbf{x}_1. \end{aligned}$$

The latter system (derived by substituting values for deterministic variables) has not full row rank any more, i.e. it is not a Gaussian hint. Therefore, the joint information of the Gaussian hint and the direct observation is captured in the pair consisting of

- the Gaussian hint inferred from the system obtained by replacing all occurrences of a variable by the observed value together with

- the information that X_1 takes the value $\mathbf{x}_1$.

It is necessary to store the value $\mathbf{x}_1$ explicitly after it is plugged in, since it cannot be retrieved from the modified system. In particular, when combining Gaussian hints with deterministic variables, the value will have to be substituted into other Gaussian linear system. This will be discussed in detail below.

Plugging in a value for a variable can be formalised using matrix notation as follows. Let (A, z, K) be a Gaussian linear system with $A \in \mathbb{R}(m, x)$ and assume the values $\mathbf{x}_1$ have to be substituted for the variables $x_1 \subseteq x$. In terms of the matrices $D_{x|x_1} \in \mathbb{R}(x, x - x_1)$,

$$D_{x|x_1} = \begin{pmatrix} I_{x-x_1} \\ 0_{x_1, x-x_1} \end{pmatrix}, \tag{10.3}$$

and $E_{x|x_1} \in \mathbb{R}(x, x_1)$,

$$E_{x|x_1} = \begin{pmatrix} 0_{x-x_1, x_1} \\ I_{x_1} \end{pmatrix}, \tag{10.4}$$

the resulting Gaussian linear system is

$$(AD_{x|x_1}, z - AE_{x|x_1}\mathbf{x}_1, K) \tag{10.5}$$

and the corresponding symmetric Gaussian potential is

$$\phi_{\mathbf{x}_1} = (D'_{x|x_1}A'K(z - AE_{x|x_1}\mathbf{x}_1), D'_{x|x_1}A'KAD_{x|x_1}). \tag{10.6}$$

The whole information can therefore be captured in the triplet

$$(\mathbf{x}_1, D'_{x|x_1}A'K(z - AE_{x|x_1}\mathbf{x}_1), D'_{x|x_1}A'KAD_{x|x_1}). \tag{10.7}$$

Such values $\mathbf{x}_1$ can be substituted directly into the symmetric Gaussian potential $e_{\mathfrak{L}}(A, z, K) = (A'Kz, A'KA) = (\nu, C)$ since

$$\begin{aligned} \phi_{\mathbf{x}_1} &= (D'_{x|x_1}A'Kz - D'_{x|x_1}A'KAE_{x|x_1}\mathbf{x}_1), D'_{x|x_1}A'KAD_{x|x_1}) \\ &= (D'_{x|x_1}(\nu - CE_{x|x_1}\mathbf{x}_1), D'_{x|x_1}CD_{x|x_1}). \end{aligned} \tag{10.8}$$

Symmetric Gaussian Potentials with Deterministic Variables

Triplets as in equation (10.7) are captured and generalised by the following definition.

Definition **10.1**. *A triplet*

$$(\mathbf{x}, \nu, C) \tag{10.9}$$

where $\mathbf{x} \in \mathbb{R}^x$ *(for some* $x \in D$*) and* $(\nu, C) \in \Delta$ *is a symmetric Gaussian potential with domain* $y = d(\nu, C)$ *such that* $x \cap y = \emptyset$ *is called a symmetric Gaussian potential with deterministic variables. Its domain is denoted*

$$d(\mathbf{x}, \nu, C) = x \cup y. \tag{10.10}$$

⊘

On the one hand, a symmetric Gaussian potential (ν, C) without deterministic variables can be represented by the triplet

$$(\diamond, \nu, C);$$

on the other hand, a deterministic hint $\mathcal{O}_{\mathbf{x}}$ can be represented by the triplet

$$(\mathbf{x}, \diamond, \diamond).$$

If (ν, C) is a conditional symmetric Gaussian potential, then $(\mathbf{x}, \nu, C)$ represents a hint on $\mathbb{R}^{x \cup y}$ with focal sets

$$\mathit{\Gamma}(\xi) = \{\mathbf{x}\} \times \{\mathbf{y} \in \mathbb{R}^y : C\mathbf{y} + \xi = \nu\} \tag{10.11}$$

for $\xi \in \mathit{\Xi} = \{\xi \in \mathbb{R}^y : \xi - \nu \in \mathcal{C}(C)\}$ (see Lemma 9.13). This situation is depicted in Figure 10.1. The focal sets are the points in the $\mathbf{x}$-plane obtained by intersection with parallel straight lines perpendicular to the (y_1, y_2)-plane.

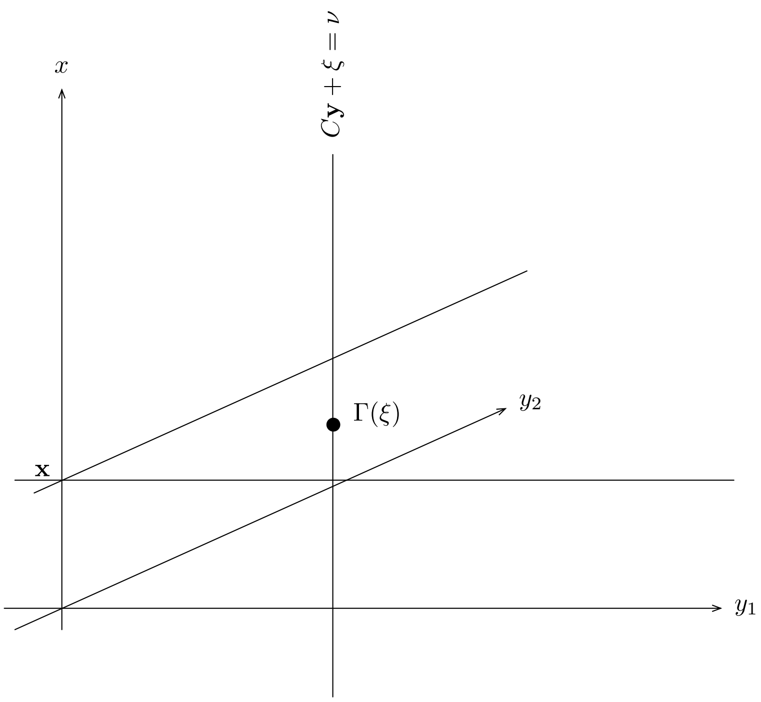


FIGURE 10.1: The focal sets $\mathit{\Gamma}(\xi)$ of a conditional symmetric Gaussian potential with deterministic variables.

Marginalisation

A conditional symmetric Gaussian potential (ν, C) with deterministic variables x fixed at $\mathbf{x}$ represents a hint on $x \cup y$. The marginal hint with respect to $s \cup t$ (for $s \subseteq x$, $t \subseteq y$) has focal sets

$$\{\mathbf{x}^{\downarrow s}\} \times \{\mathbf{y}^{\downarrow t} : \mathbf{y} \in \mathbb{R}^y, C\mathbf{y} + \xi = \nu\}. \tag{10.12}$$

The second set in equation (10.12) correspond to the projection of (ν, C) to t. Therefore, a symmetric Gaussian potential with deterministic variables $p = (\mathbf{x}, \nu, C)$ can be marginalised to $u \subseteq x \cup y$ if $u \cap y \in \mathcal{M}(\nu, C)$, i.e.

$$\mathcal{M}(\mathbf{x}, \nu, C) = \{u \subseteq x \cup y : u \cap y \in \mathcal{M}(\nu, C)\} = \{s \cup t : s \subseteq x, t \in \mathcal{M}(\nu, C)\}, \quad (10.13)$$

and the marginal is then

$$p^{\downarrow u} = (\mathbf{x}^{\downarrow u \cap x}, (\nu, C)^{\downarrow u \cap y}). \quad (10.14)$$

The right-hand side of equation (10.14) is short-hand notation for $(\mathbf{x}^{\downarrow u \cap x}, \tilde{\nu}, \tilde{C})$ with $(\tilde{\nu}, \tilde{C}) = (\nu, C)^{\downarrow u \cap y}$.

Combination

The combination of symmetric Gaussian potentials with deterministic variables is more involved. Let $(\mathbf{x}_1, \nu_1, C_1)$ and $(\mathbf{x}_2, \nu, C_2)$ be symmetric Gaussian potentials with deterministic variables x_1 and x_2, respectively. Let y_1 and y_2 be their non-deterministic variables. However, $\mathbf{x}_1$ and $\mathbf{x}_2$ may be incompatible,

$$\mathbf{x}_1 \overset{\lightning}{\bowtie} \mathbf{x}_2 \quad \Longleftrightarrow \quad {\mathbf{x}_1}^{\downarrow x_2 \cap x_1} \neq {\mathbf{x}_2}^{\downarrow x_1 \cap x_2}. \quad (10.15)$$

If $\mathbf{x}_1 \overset{\lightning}{\bowtie} \mathbf{x}_2$, there is no non-empty intersection of focal sets of p_1 and p_2. Then, the combination $p_1 \otimes p_2$ is the contradictory potential $z_{x_1 \cup x_2 \cup y_1 \cup y_2}$ on the domain $d(z_{x_1 \cup x_2 \cup y_1 \cup y_2}) = x_1 \cup x_2 \cup y_1 \cup y_2 \in D$. Of course, combining a contradiction with another piece of information yields a contradiction, hence define

$$z_x \otimes \phi = \phi \otimes z_x = z_{x \cup u} \quad (10.16)$$

for $x \in D$ and a symmetric Gaussian potential with deterministic variables ϕ of domain $u = d(\phi)$. Furthermore, a contradiction with respect to x is a contradiction with respect to any subset $y \subseteq x$, i.e. define

$${z_x}^{\downarrow y} = z_y \quad (10.17)$$

for $y \subseteq x \in D$. Thus, the contradictory elements z_x are null elements and do not represent a hint.
On the other hand, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are not incompatible, i.e. if ${\mathbf{x}_1}^{\downarrow x_2 \cap x_1} = {\mathbf{x}_2}^{\downarrow x_2 \cap x_1}$, define

$$\mathbf{x}_1 \bowtie \mathbf{x}_2 = ({\mathbf{x}_1}^{\downarrow x_1 - x_2}, {\mathbf{x}_1}^{\downarrow x_2 \cap x_1}, {\mathbf{x}_2}^{\downarrow x_2 - x_1}) = ({\mathbf{x}_1}^{\downarrow x_1 - x_2}, {\mathbf{x}_2}^{\downarrow x_1 \cap x_2}, {\mathbf{x}_2}^{\downarrow x_2 - x_1}). \quad (10.18)$$

Then, the combination is obtained by joining the deterministic variables, conditioning the symmetric Gaussian potentials and by combining the remaining non-deterministic symmetric Gaussian potentials, i.e.

$$(\mathbf{x}_1, \nu_1, C_1) \otimes (\mathbf{x}_2, \nu_2, C_2) = (\mathbf{x}_1 \bowtie \mathbf{x}_2, (\nu_1, C_1)_{{\mathbf{x}_2}^{\downarrow y_1 \cap x_2}} \otimes (\nu_2, C_2)_{{\mathbf{x}_1}^{\downarrow y_2 \cap x_1}}). \quad (10.19)$$

The neutral elements of this combination are

$$e_u = (\diamond, 0_u, 0_{u,u}) \tag{10.20}$$

since for a symmetric Gaussian potential with deterministic variables $(\mathbf{x}, \nu, C)$ with domain $u = x \cup y$ it holds that

$$e_u \otimes (\mathbf{x}, \nu, C) = (\diamond \bowtie \mathbf{x}, 0_x + \nu, 0_{x,x} + C) = (\mathbf{x}, \nu, C) \tag{10.21}$$

and, by definition,

$$e_u \otimes z_u = z_u. \tag{10.22}$$

Furthermore, these neutral elements are stable since $\mathcal{M}(e_u) = 2^u$ and

$$e_u{}^{\downarrow s} = (\diamond^{\downarrow s \cap \emptyset}, (0_u, 0_{u,u})^{\downarrow s \cap u}) = (\diamond, 0_s, 0_{s,s}) = e_s. \tag{10.23}$$

In fact, symmetric Gaussian potentials with deterministic variables form a valuation algebra. Instead of giving a direct proof, it will be shown below in Section 10.6 that they correspond to a subalgebra of symmetric Gaussian potentials with deterministic equations.

10.2 Deterministic Equations

Consider the linear equations

$$X_1 = X_2,$$
$$X_1 + \omega = 0,$$

where the term ω is Gaussian with concentration matrix K. This system represents the hint

$$(\mathbb{R}^{x_1}, \phi_{0,K}, \Gamma, \mathbb{R}^{x_1 \cup x_2})$$

where

$$\Gamma(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 = \mathbf{x}_2, \mathbf{x}_1 + \omega = 0\}$$

and $x_1 = \{X_1\}$, $x_2 = \{X_2\}$. The focal sets are thus points on a diagonal straight line as depicted in Figure 10.2. The defining equation of Γ can be decomposed as

$$\Gamma(\omega) = \Gamma_{certain} \cap \Gamma_{uncertain}(\omega) \tag{10.24}$$

where

$$\Gamma_{certain} = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 = \mathbf{x}_2\}, \qquad \Gamma_{uncertain}(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 + \omega = 0\}.$$

$\Gamma_{certain}$ has been called the certainty space (Liu, 1996a; 1999). Notice that the focal function Γ can be described in terms of other functions $\Gamma'_{uncertain}$ instead of $\Gamma_{uncertain}$, for instance in terms of

$$\Gamma'_{uncertain}(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_2 + \omega = 0\}.$$

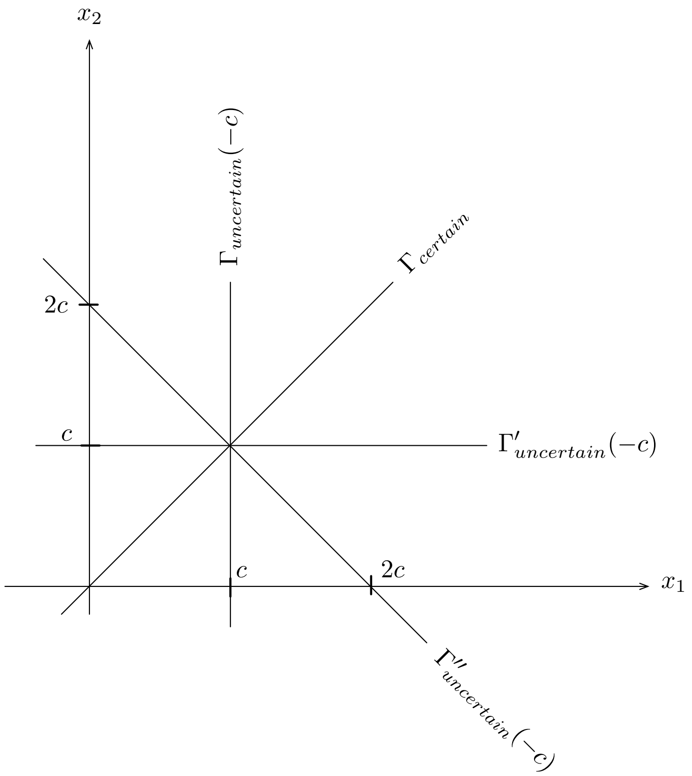


FIGURE 10.2: Different representations of the same singleton focal set

Notice that $\Gamma_{uncertain}$ and $\Gamma'_{uncertain}$ do not even involve the same variables. However, $\Gamma(\omega)$ may be empty for some assumptions ω if $\Gamma_{certain}$ and $\Gamma_{uncertain}(\omega)$ are parallel linear manifolds. In the spirit of assumption-based reasoning, the inadmissible assumptions have to be eliminated. For instance, there are no empty intersections if the sets $\Gamma_{uncertain}(\omega)$ lie perpendicular to $\Gamma_{certain}$. Of course, perpendicularity is not a necessary condition for non-empty intersections. In Figure 10.2, the function

$$\Gamma''_{uncertain}(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : 0.5 \cdot \mathbf{x}_1 + 0.5 \cdot \mathbf{x}_2 + \omega = 0\}$$

can be obtained by subtracting the first equation multiplied by 0.5 from the second equation, where the rows $(1, -1)$ and $(0.5, 0.5)$ of the compound "design matrix" are orthogonal. These preliminary considerations will now be generalised and formalised.

Gaussian Linear Systems with Deterministic Equations

DEFINITION **10.2**. *Let k be a non-negative integer, $x \in D$, $C \in \mathbb{R}(p, x)$ and $c \in \mathbb{R}^p$. Then, the pair (C, c) is called* linear system *on domain x. If $c \in \mathcal{C}(C)$, then (C, c) is called* consistent*. Further, the operations $\otimes$, $\oplus$ and $\downarrow$ can be carried over from Gaussian linear systems to (deterministic) linear systems in the obvious way (i.e.*

without considering the concentration matrix). Let (A, z, K) be a Gaussian linear system on the same domain x. Then, the quintuple (C, c, A, z, K) is called a Gaussian linear system with deterministic equations. ⊘

Let (C, c, A, z, K) be a Gaussian linear system with deterministic equations with domain $x \in D$, $C \in \mathbb{R}(p, x)$, $c \in \mathbb{R}^p$, $A \in \mathbb{R}(m, x)$, $z \in \mathbb{R}^m$, $K \in \mathbb{R}(m, m)$, and let $r = \mathrm{r}(C)$. Then, under an assumption $\omega \in \mathbb{R}^m$, the true parameter $\mathbf{x}^*$ must be in the set

$$\Gamma(\omega) = \{\mathbf{x} : A\mathbf{x} + \omega = z \text{ and } C\mathbf{x} = c\} = \{\mathbf{x} : A\mathbf{x} + \omega = z\} \cap \{\mathbf{x} : C\mathbf{x} = c\}. \quad (10.25)$$

However, the event

$$\Gamma_{certain} = \{\mathbf{x} : C\mathbf{x} = c\} \quad (10.26)$$

may rule out some assumptions as impossible, i.e. the intersection with some (or even all) sets

$$\Gamma_{uncertain}(\omega) = \{\mathbf{x} : A\mathbf{x} + \omega = z\} \quad (10.27)$$

may be empty. Assumption-based reasoning suggests deriving the admissible assumptions ω and conditioning to them. On the one hand, if the set $\Gamma_{certain}$ is empty, i.e. if $c \notin \mathcal{C}(C)$, there are no possible assumptions and the information is contradictory. On the other hand, assume $\Gamma_{certain} \neq \emptyset$. Then, an assumption ω may be inadmissible because of two different reasons, either because

(O1) $\Gamma_{uncertain}(\omega) = \emptyset$, or because

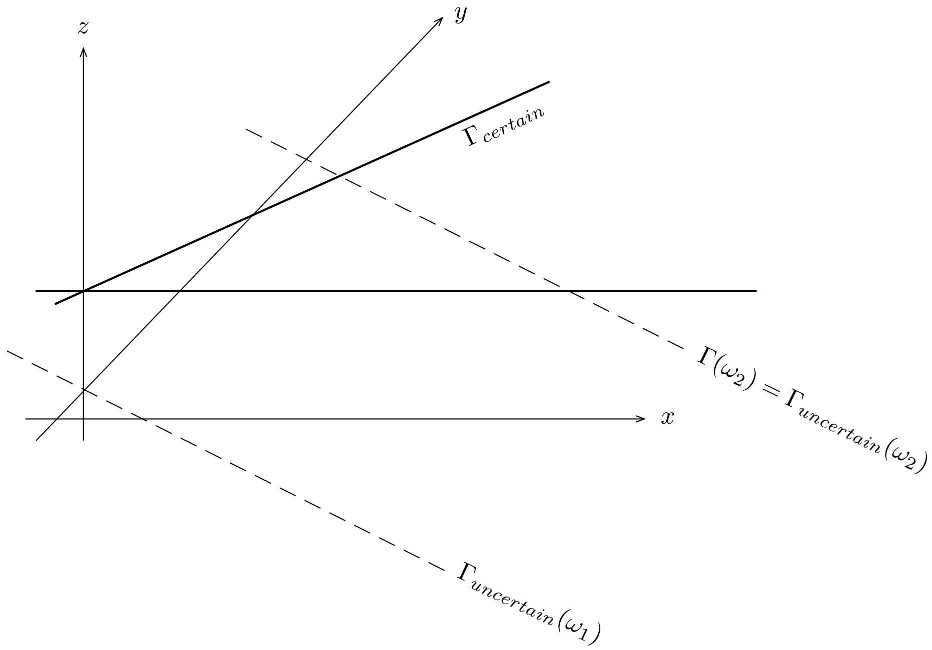
(O2) $\Gamma_{uncertain}(\omega) \neq \emptyset$, but $\Gamma_{uncertain}(\omega) \cap \Gamma_{certain} = \emptyset$.

In order to eliminate the inadmissible assumptions, the assumption-based inference may proceed in two steps as follows.

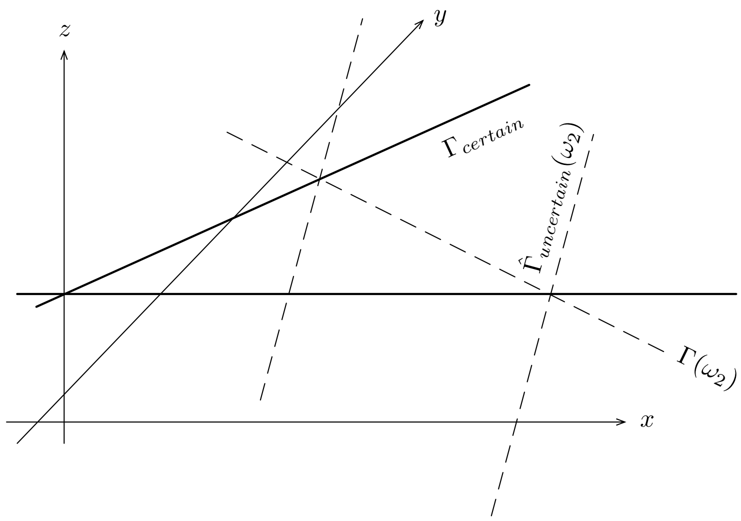
(1) In order to ensure the second condition, the deterministic knowledge is "applied" to the non-deterministic knowledge or "substituted into" the non-deterministic part. This induces a new function $\hat{\Gamma}_{uncertain}$ such that $\Gamma(\omega) = \Gamma_{certain} \cap \hat{\Gamma}_{uncertain}(\omega)$. Technically, $\hat{\Gamma}_{uncertain}$ is obtained by adding rows of (C, c) to (A, z) yielding $(\tilde{A}, \tilde{z})$ such that the row space $\mathcal{R}(C)$ of C and that of $\tilde{A}$ become essentially disjoint, i.e. $\mathcal{R}(C) \cap \mathcal{R}(A) = \{0\}$. Geometrically, this corresponds to a projection and rotation of the sets $\Gamma_{uncertain}(\omega)$. Furthermore, if the rows of A are even orthogonal to $\mathcal{R}(C)$, the non-deterministic knowledge is represented in a *unique* way.

(2) Since the non-deterministic knowledge is now "independent" from the deterministic knowledge, a Gaussian hint can be derived from the Gaussian linear system $(\tilde{A}, \tilde{z}, K)$ by using the techniques of Chapter 6.

The first step (1) is illustrated in Figure 10.3: In 10.3(a), the sets $\Gamma_{uncertain}(\omega) \neq \emptyset$ are straight lines parallel to a two-dimensional plane in the 3-dimensional space x, y, z. Here, the focal sets are straight parallel lines in the certainty plane; those ω for which $\Gamma_{uncertain}(\omega)$ lies outside the certainty plane are impossible. In 10.3(b), the new sets $\hat{\Gamma}_{uncertain}(\omega)$ are planes whose intersection with $\Gamma_{certain}$ are the straight



(a) An admissible assumption ω_2 and an inadmissible assumption ω_1.



(b) The new representation of $\Gamma(\omega_2)$ is the intersection of the certainty plane $\Gamma_{certain}$ and the plane $\hat{\Gamma}(\omega_2)$.

Figure 10.3: Applying deterministic knowledge to the non-deterministic part induces a new representation of the focal sets in the certainty plane.

lines $\Gamma(\omega)$ in the certainty plane. There are many different linear functions $\hat{\Gamma}$ with that property. However, one is free to choose the unique $\hat{\Gamma}$ such that the sets $\hat{\Gamma}(\omega)$ are perpendicular to $\Gamma_{certain}$.

In the new representation $\Gamma(\omega) = \Gamma_{certain} \cap \hat{\Gamma}_{uncertain}(\omega)$, there are no more inadmissible assumptions of type (O2). Then, the impossible assumptions of type (O1) can be ruled by inferring a hint from the Gaussian linear system $(\tilde{A}, \tilde{z}, K)$ in the usual way in step (2).

Applying Deterministic Knowledge to a Gaussian Linear System

In order to apply the deterministic knowledge, let $n = |x|$, $r = \mathrm{r}(C)$ and let $B = \binom{B_1}{B_2} \in \mathbb{R}(n, x)$ be an orthonormal matrix, i.e. $BB' = I_n$, such that the rows of C are linear combinations of $B_1 \in \mathbb{R}(r, x)$, i.e. $\mathcal{R}(B_1) = \mathcal{R}(C)$. Since the rows of A are linear combinations of the rows of B, there is a matrix $N \in \mathbb{R}(m, n)$ such that $NB = A$. Partition N into r and $n - r$ columns, $N = \big(N_1, N_2\big)$, $N_1 \in \mathbb{R}(m, r)$, $N_2 \in \mathbb{R}(m, n - r)$. Then,

$$A = NB = N_1 B_1 + N_2 B_2. \tag{10.28}$$

Since the rows of B_1 are linear combinations of the rows of C, there is a matrix $M \in \mathbb{R}(r, p)$ such that $B_1 = MC$. Substituting $Cx = c$ into Ax yields

$$z = Ax + \omega = N_1 B_1 x + N_2 B_2 x + \omega = N_1 M C x + N_2 B_2 x + \omega = N_1 M c + N_2 B_2 x + \omega.$$

Define $\tilde{C} = B_1 = MC$, $\tilde{c} = Mc$, $\tilde{A} = N_2 B_2$, $\tilde{z} = z - N_1 M c$, and

$$\hat{\Gamma}_{uncertain}(\omega) = \{\mathbf{x} : \tilde{A}\mathbf{x} + \omega = \tilde{z}\}. \tag{10.29}$$

Then, it holds that

$$\Gamma_{certain} = \{\mathbf{x} : \tilde{C}\mathbf{x} = \tilde{c}\}$$

and

$$\Gamma(\omega) = \Gamma_{certain} \cap \hat{\Gamma}_{uncertain}(\omega).$$

Here, the linear manifolds $\hat{\Gamma}_{uncertain}(\omega)$ are orthogonal to $\Gamma_{certain}$ since $\tilde{A}B_1' = N_2 B_2 B_1' = 0$. Now, the deterministic part and the non-deterministic parts are "independent", so the techniques of Chapter 6 can be applied to derive a Gaussian hint from the Gaussian linear system $(\tilde{A}, \tilde{z}, K)$.

10.3 Symmetric Gaussian Potentials with Deterministic Knowledge

Applying Deterministic Knowledge to Symmetric Gaussian Potentials

The symmetric Gaussian potential associated with the derived Gaussian linear system $(\tilde{A}, \tilde{z}, K)$ is $(\tilde{A}'K\tilde{z}, \tilde{A}'K\tilde{A})$. However, this result can be derived directly from the deterministic linear system (C, c) and the symmetric Gaussian potential $(\nu, \Lambda) = (A'Kz, A'KA)$ associated with (A, z, K) as follows. Since B is orthonormal,

$$\begin{aligned}
\tilde{A}'K\tilde{A} &= B_2'N_2'KN_2B_2 \\
&= (B_2'B_2)(B_2'N_2')K(N_2B_2)(B_2'B_2) \\
&= (B_2'B_2)(B_1'N_1' + B_2'N_2')K(N_1B_1 + N_2B_2)(B_2'B_2) \\
&= (B_2'B_2)A'KA(B_2'B_2) \\
&= (B_2'B_2)\Lambda(B_2'B_2)
\end{aligned} \tag{10.30}$$

and

$$\begin{aligned}
\tilde{A}'K\tilde{z} &= B_2'N_2'K(z - N_1Mc) \\
&= (B_2'B_2)(B_2'N_2')K(z - N_1Mc) \\
&= (B_2'B_2)(B_1'N_1' + B_2'N_2')K(z - N_1Mc) \\
&= (B_2'B_2)A'Kz - (B_2'B_2)A'KN_1Mc \\
&= (B_2'B_2)\nu - (B_2'B_2)A'K(N_1B_1 + N_2B_2)B_1'Mc \\
&= (B_2'B_2)\nu - (B_2'B_2)\Lambda B_1'Mc.
\end{aligned} \tag{10.31}$$

The following lemma shows that this result does not depend on the choice of B and M.

Lemma **10.3**. *Let (C, c) be a linear system in x, $C \in \mathbb{R}(p, x)$, $c \in \mathbb{R}^p$ such that $c \in \mathcal{C}(C)$, and let (ν, Λ) be a symmetric Gaussian potential. Further, let $B = \binom{B_1}{B_2}$ and $\tilde{B} = \binom{\tilde{B}_1}{\tilde{B}_2}$ be two orthonormal matrices such that $\mathcal{R}(B_1) = \mathcal{R}(C) = \mathcal{R}(\tilde{B}_1)$. Then, there are $M, \tilde{M}$ such that $B_1 = MC$, $\tilde{B}_1 = \tilde{M}C$. It then holds that*

$$(B_2'B_2)\Lambda(B_2'B_2) = (\tilde{B}_2'\tilde{B}_2)\Lambda(\tilde{B}_2'\tilde{B}_2)$$

and

$$(B_2'B_2)\nu - (B_2'B_2)\Lambda B_1'Mc = (\tilde{B}_2'\tilde{B}_2)\nu - (\tilde{B}_2'\tilde{B}_2)\Lambda\tilde{B}_1'\tilde{M}c.$$

⊘

Proof. It suffices to prove that $\tilde{B}_2'\tilde{B}_2 = B_2'B_2$ and $\tilde{B}_1'\tilde{M}c = B_1'Mc$.
On the one hand, notice that $\binom{B_1}{\tilde{B}_2}$ is orthonormal since

$$\begin{pmatrix} B_1 \\ \tilde{B}_2 \end{pmatrix} (B_1', \tilde{B}_2') = \begin{pmatrix} B_1B_1' & 0 \\ 0 & \tilde{B}_2\tilde{B}_2' \end{pmatrix} = I.$$

Then, by subtracting $B_1'B_1$ from

$$I = (B_1', B_2') \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = B_1'B_1 + B_2'B_2$$

and

$$I = (B_1', \tilde{B}_2') \begin{pmatrix} B_1 \\ \tilde{B}_2 \end{pmatrix} = B_1'B_1 + \tilde{B}_2'\tilde{B}_2,$$

it follows that indeed $\tilde{B}_2'\tilde{B}_2 = B_2'B_2$.
On the other hand, it can be proved in a similar way that $\tilde{B}_1'\tilde{B}_1 = B_1'B_1$. Hence,

$$B_1'MC = B_1'B_1 = \tilde{B}_1'\tilde{B}_1 = \tilde{B}_1'\tilde{M}C.$$

Since $c \in \mathcal{C}(C)$, there is an $\mathbf{x} \in \mathbb{R}^x$ such that $C\mathbf{x} = c$. Hence,

$$B_1'Mc = B_1'MC\mathbf{x} = \tilde{B}_1'\tilde{M}C\mathbf{x} = \tilde{B}_1'\tilde{M}c. \qquad \square$$

In light of the previous lemma, the following notation will be used.

DEFINITION **10.4**. *The application of a consistent linear system (C, c) to a symmetric Gaussian potential (ν, Λ) on the same domain is denoted*

$$(\nu, \Lambda)_{(C,c)} = ((B_2'B_2)\nu - (B_2'B_2)\Lambda B_1'Mc, (B_2'B_2)\Lambda(B_2'B_2)) \tag{10.32}$$

where $\binom{B_1}{B_2} \in \mathbb{R}(n, x)$ *is an orthonormal matrix such that* $\mathcal{R}(B_1) = \mathcal{R}(C)$. ⊘

Notice that $\tilde{\Lambda}C' = 0$ since $\mathcal{N}(B_2) \supseteq \mathcal{C}(B_1') = \mathcal{C}(C')$ and

$$\tilde{\Lambda}B_1' = B_2'B_2\Lambda B_2'(B_2B_1') = 0.$$

More generally, a consistent linear system may be coupled with a (not necessarily conditional) symmetric Gaussian potential. This is captured in the following definition.

DEFINITION **10.5**. *Let p be a non-negative integer. Then, a quartuple*

$$(C, c, \nu, \Lambda) \tag{10.33}$$

where

- $C \in \mathbb{R}(p, x)$, $c \in \mathbb{R}^p$ *such that* $c \in \mathcal{C}(C)$,
- (ν, Λ) *is a symmetric Gaussian potential on* $x \in D$, *and*
- $\Lambda C' = 0$

is called a consistent symmetric Gaussian potential with deterministic equations. *Further, $\Delta^\bullet$ shall denote the set of all such consistent symmetric Gaussian potentials with deterministic knowledge and all inconsistent or zero or null elements z_x $(x \in D)$. Further, define labelling $d : \Delta^\bullet \to D$ by $d(C, c, \nu, \Lambda) = x$ and $d(z_x) = x$.* ⊘

Equivalent Symmetric Gaussian Potentials with Deterministic Knowledge

There may be several symmetric Gaussian potentials with deterministic equations representing the same hint.

Lemma **10.6**. *Let* (C, c) *and* $(\tilde{C}, \tilde{c})$ *be consistent linear systems,* $C \in \mathbb{R}(m, x)$*,* $c \in \mathbb{R}^m$*,* $\tilde{C} \in \mathbb{R}(\tilde{m}, x)$*,* $c \in \mathbb{R}^{\tilde{m}}$*,* $n = |x|$ *such that*

$$\{\mathbf{x} : C\mathbf{x} = c\} = \{\mathbf{x} : \tilde{C}\mathbf{x} = \tilde{c}\}.$$

Then,

$$(\nu, \Lambda)_{(C,c)} = (\nu, \Lambda)_{(\tilde{C},\tilde{c})}.$$

⊘

Proof. It holds that $\mathcal{R}(C) = \mathcal{R}(\tilde{C})$ in light of Lemma A.2. Let $B = \binom{B_1}{B_2} \in \mathbb{R}(n, n)$ be an orthonormal matrix such that B_1 is a basis of $\mathcal{R}(C) = \mathcal{R}(\tilde{C})$. Hence, in light of Lemma 10.3, it follows that indeed $(\nu, \Lambda)_{(C,c)} = (\nu, \Lambda)_{(\tilde{C},\tilde{c})}$.

Therefore, two symmetric Gaussian potentials with deterministic equations (C, c, ν, Λ) and $(\tilde{C}, \tilde{c}, \tilde{\nu}, \tilde{\Lambda})$ are called equivalent, denoted

$$(C, c, \nu, \Lambda) \cong (\tilde{C}, \tilde{c}, \tilde{\nu}, \tilde{\Lambda}) \tag{10.34}$$

if and only if

- $\{\mathbf{x} : C\mathbf{x} = c\} = \{\mathbf{x} : \tilde{C} = \mathbf{x} = \tilde{c}\}$,
- $\nu = \tilde{\nu}$, and $\Lambda = \tilde{\Lambda}$.

Notice that the relation $\cong$ is clearly an equivalence relation. In order to keep notation simple, $\phi =_{id} \psi$ will be used for $\phi = \psi$ and $\phi = \psi$ for $\phi \cong \psi$.

Computational Aspects

The matrix B can be computed using the *singular-value decomposition* of a matrix $C \in \mathbb{R}(m, x)$ of rank $\mathrm{r}(C) = r$ (see for instance (Golub and Van Loan, 1989)), i.e.

$$C = U\Sigma V' \tag{10.35}$$

where $U \in \mathbb{R}(m, m)$ and $V \in \mathbb{R}(n, x)$ for $n = |x|$ are orthonormal matrices and

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}$$

with singular-values $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Partitioning

$$U = \begin{pmatrix} U_1, U_2 \end{pmatrix}, \qquad V' = \begin{pmatrix} V_1' \\ V_2' \end{pmatrix} \tag{10.36}$$

such that $U_1 \in \mathbb{R}(m, r)$, $U_2 \in \mathbb{R}(m, m - r)$, $V_1' \in \mathbb{R}(r, x)$ and $V_2' \in \mathbb{R}(n - r, x)$, and defining

$$\Sigma_1 = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}, \tag{10.37}$$

it holds that

$$C = U_1 \Sigma_1 V_1'. \tag{10.38}$$

The matrix

$$C^+ = V_1 {\Sigma_1}^{-1} U_1' \tag{10.39}$$

called a *pseudo-inverse* of C since $CC^+C = C$ and $C^+CC^+ = C^+$. Using this notation, substituting $B_2 = V_2'$, $B_1 = V_1'$ and $M = {\Sigma_1}^{-1} U_1'$ (since $V_1' = B_1 = MC = \Sigma^{-1} U_1' U_1 \Sigma_1 V_1'$),

$$\begin{aligned} \tilde{A}' K \tilde{A} &= (B_2' B_2) \Lambda (B_2' B_2) \\ &= (V_2 V_2') \Lambda (V_2 V_2') \end{aligned}$$

and

$$\begin{aligned} \tilde{A}' K \tilde{z} &= (B_2' B_2) \nu - (B_2' B_2) \Lambda B_1' M c \\ &= (V_2 V_2') \nu - (V_2 V_2') \Lambda V_1 {\Sigma_1}^{-1} U_1' c \\ &= (V_2 V_2') \nu - (V_2 V_2') \Lambda C^+ c. \end{aligned}$$

Whether the combination yields a contradiction, can be checked by

$$c \in \mathcal{C}(C) \iff c \in \mathcal{C}(U_1) \iff c \in \mathcal{N}(U_2) \iff U_2' c = 0 \iff U_1 U_1' c = c. \tag{10.40}$$

Notice that

$$V_2 V_2' = I_x - V_1 V_1'. \tag{10.41}$$

This can be obtained from

$$I_x = V V' = \begin{pmatrix} V_1, V_2 \end{pmatrix} = \begin{pmatrix} V_1' \\ V_2' \end{pmatrix} = V_1 V_1' + V_2 V_2'$$

Using equations (10.39), (10.40) and (10.41), applying deterministic knowledge only requires to compute U_1 and V_1 and not the whole singular-value decomposition.

Properties of Applying Deterministic Knowledge to Symmetric Gaussian Potentials

The following lemma shows that applying deterministic equations to a symmetric Gaussian potential is idempotent.

Lemma **10.7**. *Let (ν, Λ) be a symmetric Gaussian potential and let (C, c) be a consistent linear system on the same domain. Then,*

$$((\nu, \Lambda)_{(C,c)})_{(C,c)} = (\nu, \Lambda)_{(C,c)}. \tag{10.42}$$

⊘

PROOF. Let $(\tilde{\nu}, \tilde{\Lambda}) = (\nu, \Lambda)_{(C,c)}$. Let V, V_1, V_2, and C_2^+ be as above. Since V is orthonormal, $V_2'V_2 = I_{n-r}$ and $V_2'V_1 = 0_{n-r,r}$. Hence,

$$\begin{aligned}(V_2V_2')\tilde{\Lambda}(V_2V_2') &= (V_2V_2')(V_2V_2')\Lambda(V_2V_2')(V_2V_2')\\ &= V_2(V_2'V_2)V_2'\Lambda V_2(V_2'V_2)V_2'\\ &= (V_2V_2')\Lambda(V_2V_2')\\ &= \tilde{\Lambda}\end{aligned}$$

and

$$\begin{aligned}&(V_2V_2')\tilde{\nu} - (V_2V_2')\tilde{\Lambda}C^+c\\ =&(V_2V_2')\big((V_2V_2')\nu - (V_2V_2')\Lambda C^+c\big) - (V_2V_2')(V_2V_2')\Lambda(V_2V_2')C^+c\\ =&V_2(V_2'V_2)V_2'(\nu - \Lambda C^+c) - V_2(V_2'V_2)V_2'\Lambda V_2(V_2'V_1)\Sigma_1{}^{-1}U_1'c\\ =&(V_2V_2')\nu - (V_2V_2')\Lambda C^+c\\ =&\tilde{\nu}.\end{aligned}$$

□

Therefore, for $(C, c, \nu, \Lambda) \in \Delta^\bullet$, it holds that $(\nu, \Lambda)_{(C,c)} = (\nu, \Lambda)$. Furthermore, applying deterministic equations can be done step-wise.

LEMMA **10.8**. *Let (C, c) be a consistent linear system (i.e. $c \in \mathcal{C}(C)$) and let (ν, Λ) be a symmetric Gaussian potential, both on domain x, $C \in \mathbb{R}(p, x)$. Partition*

$$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

such that $C_1 \in \mathbb{R}(p_1, x)$, $C_2 \in \mathbb{R}(p_2, x)$, $c_1 \in \mathbb{R}^{p_1}$, $c_2 \in \mathbb{R}^{p_2}$. Then,

$$(\nu, \Lambda)_{(C,c)} = \big((\nu, \Lambda)_{(C_1,c_1)}\big)_{(C_2,c_2)}. \tag{10.43}$$

⊘

PROOF. Let $B_{1\cap 2}$ be an orthonormal basis of $\mathcal{R}(C_1) \cap \mathcal{R}(C_2)$. Then, there are matrices B_{1-2}, B_{2-1}, B_3 such that

$$B = \begin{pmatrix} B_{1-2} \\ B_{1\cap 2} \\ B_{2-1} \\ B_3 \end{pmatrix}, \quad B_1 = \left(\begin{smallmatrix} B_{1-2} \\ B_{1\cap 2} \end{smallmatrix}\right), \quad B_2 = \left(\begin{smallmatrix} B_{1\cap 2} \\ B_{2-1} \end{smallmatrix}\right)$$

are orthonormal bases of $\mathbb{R}^x$, $\mathcal{R}(C_1)$ and $\mathcal{R}(C_2)$, respectively. Further, let $B_{13} = \left(\begin{smallmatrix} B_1 \\ B_3 \end{smallmatrix}\right)$ and $B_{23} = \left(\begin{smallmatrix} B_2 \\ B_3 \end{smallmatrix}\right)$. Since $B_3B_2' = 0$, it holds that

$$(B_{13}'B_{13})(B_{23}'B_{23}) = (B_1', B_3')\begin{pmatrix} 0 & 0 \\ 0 & B_3B_{3'} \end{pmatrix}\begin{pmatrix} B_2 \\ B_3 \end{pmatrix} = B_3'B_3B_3'B_3 = B_3'B_3. \tag{10.44}$$

Since, $\mathcal{R}(B_1) = \mathcal{R}(C_1)$ and $\mathcal{R}(B_2) = \mathcal{R}(C_2)$, there are matrices M_1 and M_2 such that

$$M_1C_1 = B_1, \quad M_2C_2 = B_2.$$

Define $M_{12} = \left(\begin{smallmatrix} M_1 & 0 \\ 0 & M_2 \end{smallmatrix}\right)$. It holds that

$$M_{12}C = \begin{pmatrix} M_1C_1 \\ M_2C_2 \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = B_{12}.$$

Notice that

$$B_{23}'B_{23}B_2' = B_{23}' \begin{pmatrix} B_2 \\ B_3 \end{pmatrix} B_2' = (B_2', B_3') \begin{pmatrix} B_2B_2' \\ 0 \end{pmatrix} = B_2'B_2B_2' = B_2'. \tag{10.45}$$

Let $(\tilde{\nu}, \tilde{\Lambda}) = (\nu, \Lambda)_{(C_1,c_1)}$, where

$$\tilde{\nu} = (B_{23}'B_{23})(\nu - \Lambda B_1'M_1c_1), \qquad \tilde{\Lambda} = (B_{23}'B_{23})\Lambda(B_{23}'B_{23}),$$

and let $(\hat{\nu}, \hat{\Lambda}) = \big((\nu, \Lambda)_{(C_1,c_1)}\big)_{(C_2,c_2)}$, where

$$\hat{\nu} = (B_{13}'B_{13})\tilde{\nu} - (B_{13}'B_{13})\tilde{\Lambda}B_2'M_2c_2, \qquad \hat{\Lambda} = (B_{13}'B_{13})\tilde{\Lambda}(B_{13}'B_{13}).$$

Then, applying equations (10.44) and (10.45),

$$\begin{aligned}\hat{\nu} &= (B_{13}'B_{13})\tilde{\nu} - (B_{13}'B_{13})\tilde{\Lambda}B_2'M_2c_2 = \\ &= (B_{13}'B_{13})(B_{23}'B_{23})\big(\nu - \Lambda B_1'M_1c_1 - \Lambda(B_{23}'B_{23})B_2'M_2c_2\big) \\ &= (B_3'B_3)\nu - (B_3'B_3)\big(\Lambda B_1'M_1c_1 - \Lambda B_2'M_2c_2\big) \\ &= (B_3'B_3)\nu - (B_3'B_3)\Lambda B_{12}'M_{12}c\end{aligned}$$

and

$$\hat{\Lambda} = (B_{13}'B_{13})\tilde{\Lambda}(B_{13}'B_{13}) = (B_{13}'B_{13})(B_{23}'B_{23})\Lambda\big((B_{13}'B_{13})(B_{23}'B_{23})\big)' = (B_3'B_3)\Lambda(B_3'B_3).$$

Finally, since $\left(\begin{smallmatrix} B_1 \\ B_2 \end{smallmatrix}\right)$ is an orthonormal basis of $\mathcal{R}(C)$ and $M_{12}C = B_{12}$, it holds that $(\nu, \Lambda)_{(C,c)} = (\hat{\nu}, \hat{\Lambda}) = \big((\nu, \Lambda)_{(C_1,c_1)}\big)_{(C_2,c_2)}$. □

The following lemma shows that applying deterministic knowledge depends only on the variables shared with the symmetric Gaussian potential.

LEMMA **10.9**. *Let (ν, Λ) be a conditional symmetric Gaussian potential on domain x and let (C, c) be a consistent linear system on domain y. Then, it holds that*

$$(\nu, \Lambda)^{\uparrow x \cup y}{}_{(C,c)^{\uparrow x \cup y}} = (\nu, \Lambda)^{\uparrow x \cup y}{}_{((C,c)^{\downarrow x \cap y})^{\uparrow x \cup y}}. \tag{10.46}$$

⊘

PROOF. Assume without loss of generality that C has full row rank (since (C, c) is consistent and since applying deterministic knowledge does not depend on the representation in light of Lemma 10.3). Let (A, z, K) be a Gaussian linear system inducing (ν, Λ) and let T_1 be a projection matrix for C to $x \cap y$ and let T_2 be such that $T = \left(\begin{smallmatrix} T_1 \\ T_2 \end{smallmatrix}\right)$ is regular. Then, (TC, Tc) is an equivalent linear system, where

$$TC = \begin{pmatrix} T_1C^{\downarrow x \cap y} & 0 \\ T_2C^{\downarrow x \cap y} & T_2C^{\downarrow y - x} \end{pmatrix}.$$

Hence, since applying deterministic knowledge does not depend on the representation in light of Lemma 10.3, it suffices to show that

$$(\nu, \Lambda)^{\uparrow x\cup y}{}_{(TC,Tc)^{\uparrow x\cup y}} = (\nu, \Lambda)^{\uparrow x\cup y}{}_{((TC,Tc)^{\downarrow x\cap y})^{\uparrow x\cup y}}.$$

and

$$(\nu, \Lambda)^{\uparrow x\cup y}{}_{(T_1C^{\downarrow x\cap y},T_1c)^{\uparrow x\cup y}}.$$

Let B_2 be a matrix such that $\left(\begin{smallmatrix} TC^{\uparrow x\cup y} \\ B_2 \end{smallmatrix}\right)$ is regular. Then, there are matrices N_1, N_2 such that $A^{\uparrow x\cup y} = N_1TC^{\uparrow x\cup y} + N_2B_2$. Partition N_1 into N_{11} and N_{12} such that

$$N_1TC = N_{11}T_1C + N_{12}T_2C.$$

Since $T_1C^{\downarrow y-x} = 0$ and since $T_2C^{\downarrow y-x}$ must therefore have full row rank, it follows that $N_{12} = 0$ and thus $N_1TC^{\uparrow x\cup y} = N_{11}T_1C^{\uparrow x\cup y} = N_{11}T_1(C^{\downarrow x\cap y})^{\uparrow x\cup y}$. Hence, applying $(TC, Tc)^{\uparrow x\cup y}$ to $(A, z, K)^{\uparrow x\cup y}$ yields the Gaussian linear system

$$(A^{\uparrow x\cup z} - N_{11}T_1(C^{\downarrow x\cap y})^{\uparrow x\cup z}, z - N_{11}T_1c, K).$$

However, the same Gaussian linear system is obtained by applying $(C, c)^{\downarrow x\cap y} = (T_1C^{\downarrow x\cap y}, T_1c)$ to $(A, z, K)^{\uparrow x\cup y}$. This proves the claim. □

The following lemma shows that applying deterministic knowledge to a symmetric Gaussian potential and then extending it vacuously yields the same Gaussian linear system as first extending it vacuously and then applying the vacuously extended deterministic knowledge to that.

Lemma **10.10**. *Let (ν, Λ) be a conditional symmetric Gaussian potential on domain x and let (C, c) be a consistent linear system on the same domain x. Then, for u such that $x \subseteq u$, it holds that*

$$(\nu, \Lambda)^{\uparrow u}{}_{(C,c)^{\uparrow u}} = \Big((\nu, \Lambda)_{(C,c)}\Big)^{\uparrow u}. \tag{10.47}$$

⊘

Proof. Assume without loss of generality that C has full row rank (since (C, c) is consistent and since applying deterministic knowledge does not depend on the representation in light of Lemma 10.3). Let (A, z, K) be a Gaussian linear system inducing (ν, Λ) and let B_2 be a matrix such that $\left(\begin{smallmatrix} C \\ B_2 \end{smallmatrix}\right)$ is regular. Then, there are matrices N_1 and N_2 such that $A = N_1C + N_2B_2$. Applying (C, c) to (A, z, K) thus yields the Gaussian linear system $(A - N_1C, z - N_1c, K)$, and extending it to u yields

$$((A - N_1C)^{\uparrow u}, z - N_1c, K).$$

On the other hand, let

$$\begin{pmatrix} C^{\uparrow u} \\ \tilde{B}_2 \end{pmatrix}$$

be a regular matrix. Then, $A^{\uparrow u} = (A, 0_{m,u-x}) = N_1C^{\uparrow u} + \tilde{N}_2\tilde{B}_2$ for some matrix $\tilde{N}_2$ of appropriate dimensions. Hence, applying $(C, c)^{\uparrow u}$ to $(A, z, K)^{\uparrow u}$ yields

$$(A^{\uparrow u} - N_1C^{\uparrow u}, z - N_1c, K) = ((A - N_1C)^{\uparrow u}, z - N_1c, K).$$

This proves the claim. □

10.4 Combination

Let (C_1, c_1, ν_1, C_1) and (C_2, c_2, ν_2, C_2) be two consistent symmetric Gaussian potentials with deterministic equations and let $\Gamma^{(i)}_{certain}$ and $\Gamma^{(i)}_{uncertain}$ be the decomposition of their focal functions

$$\Gamma^{(i)}(\omega^{(i)}) = \Gamma^{(i)}_{certain} \cap \Gamma^{(i)}_{uncertain}(\omega^{(i)}), \qquad i \in \{1, 2\}.$$

Assume without loss of generality that they refer to the same domain x. According to Dempster's Rule, the combination of these two focal functions is the restriction of

$$\Gamma(\omega^{(1)}, \omega^{(2)}) = (\Gamma^{(1)}_{certain} \cap \Gamma^{(1)}_{uncertain}(\omega^{(1)})) \cap (\Gamma^{(2)}_{certain} \cap \Gamma^{(2)}_{uncertain}(\omega^{(2)}))$$

to the admissible assumptions. If there are no amissible assumptions, the two hints are contradictory and the result of the combination the inconsistent or contradictory element z_x. If the two hints are not contradictory,

$$\begin{aligned} &\Gamma(\omega^{(1)}, \omega^{(2)}) \\ =&(\Gamma^{(1)}_{certain} \cap \Gamma^{(2)}_{certain}) \cap (\Gamma^{(1)}_{uncertain}(\omega^{(1)}) \cap \Gamma^{(2)}_{uncertain}(\omega^{(2)})) \\ =&\{\mathbf{x} : C_1\mathbf{x} = c_1, C_2\mathbf{x} = c_2\} \cap \left(\Gamma^{(1)}_{uncertain}(\omega^{(1)}) \cap \Gamma^{(2)}_{uncertain}(\omega^{(2)})\right) \\ =&\left(\{\mathbf{x} : C\mathbf{x} = c\} \cap \Gamma^{(1)}_{uncertain}(\omega^{(1)})\right) \cap \left(\{\mathbf{x} : C\mathbf{x} = c\} \cap \Gamma^{(2)}_{uncertain}(\omega^{(2)})\right) \end{aligned}$$

where

$$C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Therefore, the techniques developed so far can be used to define the combination of two consistent symmetric Gaussian potentials by

$$\begin{aligned} &(C_1, c_1, \nu_1, \Lambda_1) \otimes (C_2, c_2, \nu_2, \Lambda_2) \\ =&\begin{cases} (C, c, (\nu_1, \Lambda_1)_{(C,c)} \otimes (\nu_2, \Lambda_2)_{(C,c)}) & \text{if } (C, c) \text{ is consistent}, \\ z_x & \text{else}. \end{cases} \end{aligned} \tag{10.49}$$

Here, $(C, c, (\nu_1, C_1)_{(C,c)} \otimes (\nu_2, C_2)_{(C,c)})$ is short-hand notation for (C, c, ν, Λ) with $(\nu, \Lambda) = (\nu_1, C_1)_{(C,c)} \otimes (\nu_2, C_2)_{(C,c)}$. Notice that combination is well defined, i.e. $(C_1, c_1, \nu_1, \Lambda_1) = (\tilde{C}_1, \tilde{c}_1, \nu_1, \Lambda_1)$ and $(C_2, c_2, \nu_2, \Lambda_2) = (\tilde{C}_2, \tilde{c}_2, \nu_2, \Lambda_2)$ imply

$$(C_1, c_1, \nu_1, C_1) \otimes (C_2, c_2, \nu_2, C_2) = (\tilde{C}_1, \tilde{c}_1, \nu_1, \Lambda_1) \otimes (\tilde{C}_2, \tilde{c}_2, \nu_2, \Lambda_2). \tag{10.50}$$

This is a consequence of Lemma 10.6. As in the case of deterministic variables, define the combination of $\phi \in \Delta^\bullet$ of domain $x = d(\phi)$ with a null element z_y by

$$\phi \otimes z_y = z_y \otimes \phi = z_{x \cup y}. \tag{10.51}$$

The following lemma provides an alternative combination rule: it suffices to apply the deterministic knowledge coming from the other factor and then to combine these two symmetric Gaussian potentials.

Lemma **10.11**. *Let $(C_1, c_1, \nu_1, \Lambda_1)$ and $(C_2, c_2, \nu_2, \Lambda_2)$ be symmetric Gaussian potentials with deterministic knowledge and let (C, c) be the combined system of (C_1, c_1) and (C_2, c_2). If (C, c) is consistent,*

$$(C_1, c_1, \nu_1, \Lambda_1) \otimes (C_2, c_2, \nu_2, \Lambda_2) = (C, c, (\nu_1, \Lambda_1)_{(C_2,c_2)} \otimes (\nu_2, \Lambda_2)_{(C_1,c_1)}). \quad (10.52)$$

⊘

Proof. Since applying deterministic equations is transitive and idempotent in light of Lemmata 10.8 and 10.7, it holds that

$$\begin{aligned} &(C_1, c_1, \nu_1, \Lambda_1) \otimes (C_2, c_2, \nu_2, \Lambda_2) \\ =&\Big(C, c, \big((\nu_1, \Lambda_1)_{(C_1,c_1)}\big)_{(C_2,c_2)} \otimes \quad \big((\nu_2, \Lambda_2)_{(C_2,c_2)}\big)_{(C_1,c_1)}\Big) \\ =&\big(C, c, (\nu_1, \Lambda_1)_{(C_2,c_2)} \otimes (\nu_2, \Lambda_2)_{(C_1,c_1)}\big). \end{aligned}$$

□

Furthermore, the following lemma shows that combining symmetric Gaussian potentials and applying deterministic knowledge commute.

Lemma **10.12**. *For symmetric Gaussian potentials $(\nu_1, \Lambda_1), (\nu_2, \Lambda_2) \in \Delta$ and a consistent linear system (C, c) on the same domain, it holds that*

$$(\nu_1, \Lambda_1)_{(C,c)} \otimes (\nu_2, \Lambda_2)_{(C,c)} = \big((\nu_1, C_1) \otimes (\nu_2, C_2)\big)_{(C,c)}. \quad (10.53)$$

⊘

Proof. Let $(\nu, \Lambda) = (\nu_1, \Lambda_1)_{(C,c)} \otimes (\nu_2, \Lambda_2)_{(C,c)}$. Using the singular-value decomposition of C, it holds that

$$\begin{aligned} \Lambda &= (V_2V_2')\Lambda_1(V_2V_2') + (V_2V_2')\Lambda_2(V_2V_2') \\ &= (V_2V_2')(\Lambda_1 + \Lambda_2)(V_2V_2') \end{aligned}$$

and

$$\begin{aligned} \nu &= ((V_2V_2')\nu_1 - (V_2V_2')\Lambda_1C^+c) + ((V_2V_2')\nu_2 - (V_2V_2')\Lambda_2C^+c) \\ &= (V_2V_2')(\nu_1 + \nu_2) - (V_2V_2')(\Lambda_1 + \Lambda_2)C^+c. \end{aligned}$$

□

10.5 Marginalisation

The marginalisation of conditional symmetric Gaussian potentials with deterministic knowledge ought to correspond to the projection of focal sets. As an example, consider again the linear equations

$$\begin{aligned} X_1 &= X_2, \\ 0.5 \cdot X_1 + 0.5 \cdot X_2 + \omega &= 0, \end{aligned}$$

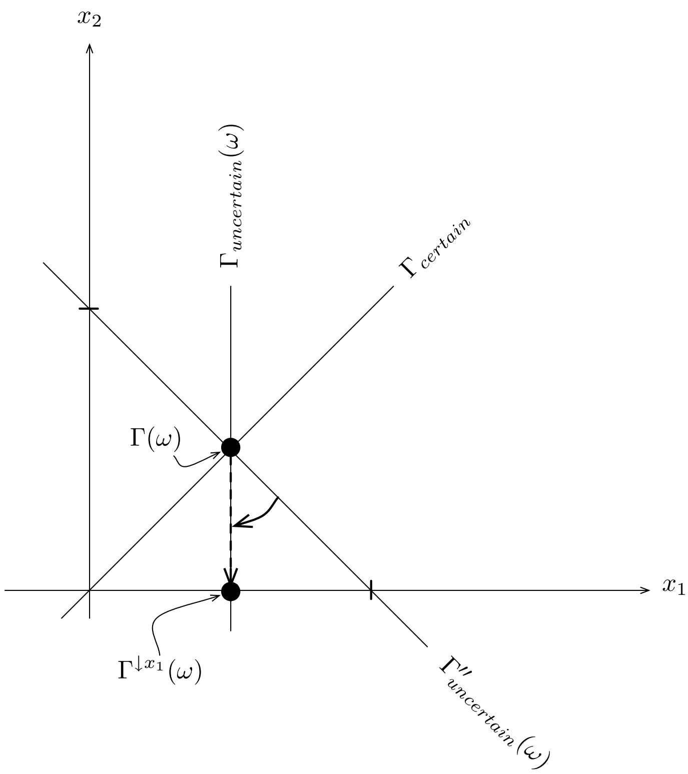


FIGURE 10.4: Projection of singleton focal sets on a straight line by "rotation" of $\Gamma_{uncertain}$

where the term ω is Gaussian with concentration K. The focal sets, which are points on a diagonal straight line, have to be projected onto the x_1-axis as depicted in Figure 10.4. In the decomposition (10.24),

$$\Gamma(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 - \mathbf{x}_2 = 0\} \cap \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 + \omega = 0\}.$$

the straight lines $\Gamma_{uncertain}(\omega) = \{(\mathbf{x}_1, \mathbf{x}_2) : \mathbf{x}_1 + \omega = 0\}$ stand orthogonal to the x_1-axis. Since x_2 does not occur in $\Gamma_{uncertain}$, $\Gamma_{certain}$ and $\Gamma_{uncertain}$ can be projected independently to x_1,

$$\begin{aligned}
\Gamma^{\downarrow x_1}(\omega) &= \{\mathbf{x}_1 : \exists \mathbf{x}_2 \text{ s.t. } \mathbf{x}_1 - \mathbf{x}_2 = 0, 2 \cdot \mathbf{x}_1 + \omega = 0\} \\
&= \{\mathbf{x}_1 : \exists \mathbf{x}_2 \text{ s.t. } \mathbf{x}_1 - \mathbf{x}_2 = 0\} \cap \{\mathbf{x}_1 : 2 \cdot \mathbf{x}_1 + \omega = 0\} \\
&= \{\mathbf{x}_1 : 2 \cdot \mathbf{x}_1 + \omega = 0\} \cap \{\mathbf{x}_1 \in \mathbb{R}^{x_1}\} \\
&= {\Gamma_{uncertain}}^{\downarrow x_1}(\omega) \cap {\Gamma_{certain}}^{\downarrow x_1}.
\end{aligned}$$

Here, the projection ${\Gamma_{certain}}^{\downarrow x_1}$ of the straight line $\Gamma_{certain}$ is the whole x_1-axis and ${\Gamma_{uncertain}}^{\downarrow x_1}$ corresponds to the usual projection of the associated Gaussian linear system.

This two-step procedure can be generalised as follows.

- In a first step, the non-deterministic part is rotated and projected such that
- the deterministic and the modified non-deterministic part can then be treated independently in the second step.

The following lemma provides a sufficient condition for projecting two subsystems of a linear system independently.

Lemma **10.13**. *Let $A_1 \in \mathbb{R}(m_1, x)$ and $A_2(m_2, x)$, and let $x_1 \cup x_2 = x$ such that $x_1 \cap x_2$. Decompose $A_1 = (A_{11}, A_{12})$ and $A_2 = (A_{21}, A_{22})$ according to x_1 and x_2, i.e. $A_{11} \in \mathbb{R}(m_1, x_1)$ and $A_{21} \in \mathbb{R}(m_2, x_2)$. Let P_1 be a projection matrix for x_1 in A_1 and let P_2 be a projection matrix for x_1 in A_2. If the rows of A_{12} and A_{22} are linearly independent,*

$$P = \begin{pmatrix} P_1 & \\ & P_2 \end{pmatrix} \tag{10.54}$$

is a projection matrix for x_1 in $A = \binom{A_1}{A_2} \in \mathbb{R}(m, x)$, $m = m_1 + m_2$. ⊘

Proof. Decompose $A = (A_{x_1}, A_{x_2})$ such that $A_{x_1} \in \mathbb{R}(m_1 + m_2, x_1)$. Let $k_1 = \mathrm{r}(A_{12})$ and $k_2 = \mathrm{r}(A_{22})$. Then,

$$PA = \begin{pmatrix} P_1 A_{11} & P_1 A_{12} \\ P_2 A_{21} & P_2 A_{22} \end{pmatrix} = (PA_{x_1}, 0).$$

Furthermore, since P is block-diagonal and since the rows of A_{12} and A_{22} are linearly independent,

$$\mathrm{r}(P) = \mathrm{r}(P_1) + \mathrm{r}(P_2) = (m_1 - k_1) + (m_2 - k_2) = (m_1 + m_2) - (k_1 + k_2) = m - \mathrm{r}(A_{x_2}).$$

This shows that P is a indeed a projection matrix for x_1 in A. □

In order to use this lemma, a non-deterministic part $\Gamma_{uncertain}(\omega)$ will be "rotated" and "projected" to $\tilde{\Gamma}_{uncertain}(\omega)$ by substituting parts of the deterministic knowledge into the non-deterministic part. Let (A, z, K) be a Gaussian linear system with $A \in \mathbb{R}(m, x)$ and let (C, c) be a consistent linear system with $C \in \mathbb{R}(p, x)$. Decompose $x = x_1 \cup x_2$ such that $x_1 \cap x_2 = \emptyset$ and $C = (C_1, C_2)$ such that $C_2 \in \mathbb{R}(p, x_2)$. Further, let $r = \mathrm{r}(C_2)$ and $n_2 = |x_2|$. Then,

$$\Gamma(\omega) = \{\mathbf{x} : C\mathbf{x} = c, A\mathbf{x} + \omega = z\}$$

Let $V' \in \mathbb{R}(n_2, x_2)$ be a regular matrix. Partition

$$V' = \begin{pmatrix} V'_1 \\ V'_2 \end{pmatrix}$$

such that $\mathcal{R}(V'_1) = \mathcal{R}(C_2)$, i.e. $V'_1 \in \mathbb{R}(r, x_2)$, $V'_2 \in \mathbb{R}(n_2 - r, x_2)$. Since $\mathcal{R}(A_2) \subseteq \mathcal{R}(V')$, there are matrices $M_1 \in \mathbb{R}(m, r), M_2 \in \mathbb{R}(m, m - r)$ such that

$$A_2 = M_1 V'_1 + M_2 V'_2.$$

Since $\mathcal{R}(C_2) = \mathcal{R}(V_1')$, there is a matrix $M^* \in \mathbb{R}(m,p)$ such that

$$M_1 V_1' = M^* C_2.$$

Then, since

$$S = \begin{pmatrix} I_p & \\ -M^* & I_m \end{pmatrix}$$

is regular, it holds that

$$\Gamma(\omega) = \Gamma_{certain} \cap \tilde{\Gamma}_{uncertain}(\omega)$$

for

$$\Gamma_{certain} = \{\mathbf{x} : C\mathbf{x} = c\}$$

and

$$\tilde{\Gamma}_{uncertain}(\omega) = \{\mathbf{x} : -M^*C\mathbf{x} + A\mathbf{x} + \omega = z - M^*c\} = \{\mathbf{x} : \tilde{A}\mathbf{x} = \tilde{z}\}$$

for $\tilde{A} = A - M^*C$ and $\tilde{z} = z - M^*c$. Partition $\tilde{A} = (\tilde{A}_1, \tilde{A}_2)$ such that $\tilde{A}_2 \in \mathbb{R}(m, x_2)$. Observe that $\tilde{A}_2 = M_2 V_2'$. Hence, $\tilde{A}$ and C satisfy the condition of Lemma 10.13: let P_C be a projection matrix for x_1 in C and let $P_{\tilde{A}}$ be a projection matrix for x_1 in $\tilde{A}$; then,

$$\begin{aligned}\Gamma^{\downarrow x_1}(\omega) &= \{\mathbf{x}_1 : \exists \mathbf{x}_2 \text{ s.t. } \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2), C\mathbf{x} = c, \tilde{A}\mathbf{x} + \omega = \tilde{z}\} \\ &= \{\mathbf{x}_1 : P_C C_1 \mathbf{x}_1 = P_C c\} \cap \{\mathbf{x}_1 : P_{\tilde{A}} \tilde{A}_1 \mathbf{x} + P_{\tilde{A}} \omega = P_{\tilde{A}} \tilde{z}\}.\end{aligned}$$

The last set corresponds to ordinary marginalisation of the Gaussian linear system $(\tilde{A}, \tilde{z}, K)$ to x_1 as discussed in Section 6.4.

These considerations are now carried over to symmetric Gaussian potentials. Let

$$(\nu, \Lambda) = (A'KA, A'Kz)$$

be the symmetric Gaussian potential corresponding to the non-deterministic part before the "rotation." Notice that

$$A'KA = \begin{pmatrix} A_1'KA_1 & A_1'KA_2 \\ A_2'KA_1 & A_2'KA_2 \end{pmatrix}.$$

For the "rotation", assume that the singular-value decomposition of C_2 is $C_2 = U\Sigma V'$ and $C_2^+ = V_1 {\Sigma_1}^{-1} U_1'$. Then, the non-deterministic part after the "rotation" corresponds to the symmetric Gaussian potential $(\tilde{\nu}, \tilde{\Lambda})$ where

$$\tilde{\nu} = (A' - C'M^{*\prime})K(z - M^*c) \tag{10.55}$$

and

$$\tilde{\Lambda} = (A' - C'M^{*\prime})K(A - M^*C). \tag{10.56}$$

Then, since $V_1'V_1 = I_r$ and $V_2'V_1 = 0$,

$$M_1 V_1' = M_1(V_1'V_1)V_1' = (M_1V_1' + M_2V_2')V_1V_1' = A_2V_1V_1',$$

and hence

$$M^* = M^* C_2 C_2^+ = M_1 V_1' C_2^+ = M_1 V_1' C_2^+ = A_2 V_1 (V_1' V_1) \Sigma^{-1} U_1' = A_2 V_1 \Sigma^{-1} U_1' \\ = A_2 C_2^+. \tag{10.57}$$

Therefore, applying (10.57) to (10.55) and (10.56),

$$\tilde{\nu} = (A' - C' C_2^{+\prime} A_2') K (z - A_2 C_2^+ c) \tag{10.58}$$
$$= \nu - C' C_2^{+\prime} \nu^{\downarrow x_2} - \Lambda^{\downarrow x_1 \cup x_2, x_2} C_2^+ c + C' C_2^{+\prime} \Lambda^{\downarrow x_2} C_2^+ c \tag{10.59}$$

and

$$\tilde{\Lambda} = (A' - C' C_2^{+\prime} A_2') K (A - A_2 C_2^+ C) \tag{10.60}$$
$$= \Lambda - C' C_2^{+\prime} \Lambda^{\downarrow x_2, x_1 \cup x_2} - \Lambda^{\downarrow x_1 \cup x_2, x_2} C_2^+ C + C' C_2^{+\prime} \Lambda^{\downarrow x_2} C_2^+ C. \tag{10.61}$$

DEFINITION **10.14**. *Let $(C, c, \nu, \Lambda) \in \Delta^\bullet$ be a consistent symmetric Gaussian potential on the domain x. Define the "rotation" for $x_1 \subseteq x$ by*

$$(\nu, \Lambda)_{\perp_{x_1}(C,c)} = (\tilde{\nu}, \tilde{\Lambda}) \tag{10.62}$$

⊘

The following lemma shows that the result of this "rotation" does not depend on the particular linear system used for the deterministic knowledge.

LEMMA **10.15**. *Let $(C, c, \nu, \Lambda) = (\tilde{C}, \tilde{c}, \nu, \Lambda)$ on domain x. Then,*

$$(\nu, \Lambda)_{\perp_{x_1}(C,c)} = (\nu, \Lambda)_{\perp_{x_1}(\tilde{C},\tilde{c})} \tag{10.63}$$

for $x_1 \subseteq x$. ⊘

PROOF. It suffices to show that $C_2^+ C_1 = \tilde{C}_2^+ \tilde{C}_1$, and $C_2^+ c = \tilde{C}_2^+ \tilde{c}$.

Since $\mathcal{R}(C) = \mathcal{R}(\tilde{C})$ in light of Lemma A.2, there is a matrix M such that $\tilde{C} = MC$ and $\tilde{C}_1 = MC_1$ and $\tilde{C}_2 = MC_2$. Then,

$$C_2^+ = \tilde{C_2}^+ \tilde{C}_2 C_2^+ = \tilde{C}_2^+ M C_2 C_2^+ = \tilde{C_2}^+ M. \tag{10.64}$$

Then, using (10.64),

$$\tilde{C_2}^+ \tilde{C}_1 = \tilde{C_2}^+ (MC_1) = C_2^+ C_1$$

and, similarly,

$$\tilde{C_2}^+ \tilde{C}_2 = \tilde{C_2}^+ (MC_2) = C_2^+ C_2.$$

Thence,

$$\begin{aligned} \tilde{C_2}^+ \tilde{c} &= \tilde{C}_2^+ \tilde{C} \mathbf{x} \\ &= \tilde{C}_2^+ M C \mathbf{x} \\ &= C_2^+ c \end{aligned}$$

for any $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in \{\mathbf{x} : C\mathbf{x} = c\} = \{\mathbf{x} : \tilde{C}\mathbf{x} = \tilde{c}\}$. Such an $\mathbf{x}$ exists since, by assumption, $c \in \mathcal{C}(C)$ and $\tilde{c} \in \mathcal{C}(\tilde{C})$. □

These considerations lead to the following (partial) definition of marginalisation of symmetric Gaussian potentials with deterministic equations.

DEFINITION **10.16**. *Let (C, c, ν, Λ) be a consistent symmetric Gaussian potential with deterministic knowledge. Let $T \in \mathbb{R}(k, k)$ be a regular matrix such that $T_1 \in \mathbb{R}(k-r, k)$ is a projection matrix for the variables x_1 in C,*

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} C = \begin{pmatrix} T_1 C_1 & 0 \\ T_2 C_1 & T_2 C_2 \end{pmatrix}. \tag{10.65}$$

Notice that $T_2 C_2$ has full row rank r. The marginals of (C, c, ν, Λ) with domain x are defined for the subdomains

$$\mathcal{M}(C, c, \nu, \Lambda) = \begin{cases} \mathcal{M}(\nu, \Lambda) & \text{if } C = 0, c = 0, \\ \{y : y \subseteq x\} & \text{if } (\nu, \Lambda) \text{ is a conditional SGP,} \\ \{x\} & \text{else} \end{cases} \tag{10.66}$$

by

$$(C, c, \nu, \Lambda)^{\downarrow y} = \begin{cases} (0, 0, (\nu, \Lambda)^{\downarrow y}), \\ \big(T_1 C_1, T_1 c, (\nu, \Lambda)_{\perp y(C,c)}{}^{\downarrow y}\big), \\ (C, c, \nu, \Lambda), \end{cases} \tag{10.67}$$

respectively. On the other hand, define marginalisation of the contradictory elements by

$$z_x{}^{\downarrow y} = z_y \tag{10.68}$$

for $y \subseteq x \in D$, i.e. $\mathcal{M}(z_x) = 2^x$. ⊘

The definition of $\mathcal{M}$ is sound since, if (ν, Λ) is a conditional symmetric Gaussian potential, then equations (10.58) and (10.60) show that $(\nu, \Lambda)_{\perp x} = (\tilde{\nu}, \tilde{\Lambda})$ is a conditional symmetric Gaussian potential as well and thus $x \in \mathcal{M}(\nu, \Lambda)$.
It has to be verified that marginalisation is well defined. The marginals defined for inconsistent elements, for empty deterministic knowledge or a non-conditional symmetric Gaussian potential with non-empty deterministic knowledge is are clearly elements of $\Delta^\bullet$ again. Finally, in the case of a conditional symmetric Gaussian potential with deterministic knowledge (C, c, ν, Λ), it has to be verified that $(C, c, \nu, \Lambda)^{\downarrow y} = (C_y, c_y, \nu_y, \Lambda_y)$ is in $\Delta^\bullet$, i.e. that $C_y \Lambda_y' = 0$. Let (A, z, K) be a Gaussian linear system inducing (ν, Λ) and let $(A, z, K)^{\downarrow y} = (PA_1, Pz, K_y)$, $K_y = (PK^{-1}P')^{-1}$. Further, let $C_y = T_1 C_1$, and $\Lambda_y = L' K_y L$ for

$$L = P(A - A_2 C_2^+ C).$$

It suffices to prove that

$$C_y L' = T_1 C_1 A' - T_1 C_1 C' C_2^{+'} A_2' = 0.$$

First, $C\Lambda' = CA'KA$ implies that $CA' = 0$ since $\mathcal{N}(C) \supseteq \mathcal{C}(A') = \mathcal{C}(A'KA)$. Then, since $CA' = 0$ implies $C_1 A' = 0$ and since $\mathcal{R}(C_2^{+'}) = \mathcal{R}(C_2)$ implies $\mathcal{N}(C_2^{+'}) = \mathcal{N}(C_2) \supseteq \mathcal{C}(A_2')$. Finally, $C_y L' = 0$ implies $C_y L' K L = C_y \Lambda_y' = 0$. Hence, marginalisation is indeed well defined.

10.6 VA of Symmetric Gaussian Potentials with Deterministic Equations

It will now be shown that symmetric Gaussian potentials form a valuations algebra.

As a preparation, the following three lemmata will be proved. They will be used in the demonstration of the combination axiom holds for symmetric Gaussian potentials with deterministic equations. The first lemma shows that the deterministic equations of domain smaller than the target domain of the marginalisation are irrelevant for the rotation. The second lemma shows that the rotation does not affect a symmetric Gaussian potential whose domain is smaller than the target domain of the marginalisation. The third lemma shows that deterministic knowledge of domain smaller than the target domain of the marginlisation can be applied before or after the marginalisation.

Lemma **10.17**. *Let (ν, Λ) be a conditional symmetric Gaussian potential on domain x, and let (C_1, c_1) and (C_2, c_2) be consistent linear systems on $y \subseteq x$ and x, respectively. Then, for z such that $y \subseteq z \subseteq x$, it holds that*

$$(\nu, \Lambda)_{\perp_z (C_1,c_1)^{\uparrow x} \oplus (C_2,c_2)} = (\nu, \Lambda)_{\perp_z (C_2,c_2)}. \tag{10.69}$$

⊘

Proof. Let (A, z, K) be a Gaussian linear system inducing (ν, Λ). Let $(C, c) = (C_1, c_1)^{\uparrow x} \oplus (C_2, c_2)$. Let k_1 and m be such that $C_1 \in \mathbb{R}(k_1, y)$ and $C_2 \in \mathbb{R}(m, x)$. Partition $C_2 = (C_{21}, C_{22})$ such that C_{22} corresponds to the variables $x - z$. Let $\binom{V_1'}{V_2'}$ be a regular matrix such that $\mathcal{R}(V_1') = \mathcal{R}(C_{22})$. Then, there are matrices M_1, M_2 such that $A^{\downarrow x-z} = M_1 V_1' + M_2 V_2'$. Hence, there is a matrix M^* such that $M_1 V_1' = M^* C_2$. Then, the rotation of (A, z, K) for z according to (C_2, c_2) is

$$(A, z, K)_{\perp_z (C_2,c_2)} (A - M^* C_2, z - M^* c_2, K)$$

[the same notation is temporarily used here for the rotation of Gaussian linear systems as for symmetric Gaussian potentials] Further, $\mathcal{R}(V_1') = \mathcal{R}(C^{\downarrow x-z})$. Define $M^{**} = \binom{0_{m,k_1}}{M^*}$. Then, it holds that $M_1 V_1' = M^{**} C$. Therefore, the rotation of (A, z, K) for z according to (C, c) is

$$(A, z, K)_{\perp_z (C_1,c_1)^{\uparrow x} \oplus (C_2,c_2)} (A - M^{**} C, z - M^{**} c, K) = (A - M^* C_2, z - M^* c_2, K).$$

This proves the claim. □

Lemma **10.18**. *Let (ν_1, Λ_1) and (ν_2, Λ_2) be conditional symmetric Gaussian potentials on domain x and y, respectively. Let (C_2, c_2) be a consistent linear system on domain y. Then, for z such that $x \subseteq z \subseteq x \cup y$, it holds that*

$$((\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2))_{\perp_z (C_2,c_2)^{\uparrow x \cup y}} = (\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2)_{\perp_{z \cap y} (C_2,c_2)}. \tag{10.70}$$

⊘

PROOF. Let (A_1, z_1, K_1) and (A_2, z_2, K_2) be Gaussian linear systems inducing (ν_1, Λ_1) and (ν_2, Λ_2), respectively. Let V_1' and V_2' be a matrices of full row rank such that $\mathcal{R}(V_1') = \mathcal{R}({C_2}^{\downarrow y-z})$ and such that $\left(\begin{smallmatrix} V_1' \\ V_2' \end{smallmatrix}\right)$ is regular. Then, there are matrices M_1, M_2 such that

$$
{A_2}^{\downarrow y-z} = M_1 V_1' + M_2 V_2'
$$

Further, there is a matrix M^* such that $M^* {C_2}^{\downarrow y-z} = M_1 V_1'$. Then,

$$
(A_2, z_2, K_2)_{\perp_{z \cap y}(C_2, c_2)} = (A_2 - M^* C_2, z_2 - M^* c_2, K).
$$

On the other hand, let $(A, z, K) = (A_1, z_1, K_1) \oplus (A_2, z_2, K_2)$ where

$$
A = \begin{pmatrix} {A_1}^{\uparrow z} & 0 \\ {A_2}^{\downarrow z \cap y \uparrow z} & {A_2}^{\downarrow y-z} \end{pmatrix}.
$$

Therefore, define M^{**} such that

$$
A^{\downarrow (x \cup y)-z} = \begin{pmatrix} 0 \\ M_1 \end{pmatrix} V_1' + \begin{pmatrix} 0 \\ M_2 \end{pmatrix} V_2' = \begin{pmatrix} 0 \\ M^* \end{pmatrix} {C_2}^{\downarrow y-z} + \begin{pmatrix} 0 \\ M_2 \end{pmatrix} V_2'.
$$

Further,

$$
{C_2}^{\uparrow x \cup y} = ({C_2}^{\downarrow z \cap y \uparrow x \cup y}, {C_2}^{\downarrow y-z}).
$$

Hence,

$$
(A, z, K)_{\perp_z (C_2, c_2)^{\uparrow x \cup y}} = (A - M^{**} C_2, z_2 - M^{**} c_2, K).
$$

Finally, it holds that

$$
A - M^{**} C_2 = \begin{pmatrix} {A_1}^{\uparrow x \cup y} \\ (A_2 - M^* C_2)^{\uparrow x \cup y} \end{pmatrix}, \quad z - M^{**} c_2 = \begin{pmatrix} z_1 \\ z_2 - M^* z_2 \end{pmatrix}.
$$

This concludes the proof. □

LEMMA **10.19**. *Let (C_2, c_2, ν, Λ) be a conditional symmetric Gaussian potential with deterministic knowledge on domain x, and let (C_1, c_1) be a consistent linear system on $y \subseteq x$ such that $(C_1, c_1)^{\uparrow x} \oplus (C_2, c_2)$ is consistent. Then, for z such that $y \subseteq z \subseteq x$, it holds that*

$$
\left((\nu, \Lambda)_{(C_1, c_1)^{\uparrow x} \perp_z (C_2, c_2)}\right)^{\downarrow z} = \left((\nu, \Lambda)_{\perp_z (C_2, c_2)}\right)^{\downarrow z}{}_{(C_1, c_1)^{\uparrow z}}. \tag{10.71}
$$

⊘

PROOF. Since (C_1, c_1) is consistent and since applying deterministic knowledge does not depend on the representation in light of Lemma 10.3, it can be assumed without loss of generality that C_1 has full row rank. Since $(C_1, c_1)^{\uparrow x}$ and (C_2, c_2) are assumed compatible and in light of the idempotency of applying deterministic knowledge (Lemma 10.7), assume without loss of generality that the row spaces of ${C_1}^{\uparrow x}$ and C_2 are essentially disjoint, i.e. that $\mathcal{R}({C_1}^{\uparrow x}) \cap \mathcal{R}(C_2) = \{0\}$. Let B_3 be a matrix such that $\begin{pmatrix} {C_1}^{\uparrow x} \\ C_2 \\ B_3 \end{pmatrix}$ is regular. It is now shown that it can be assumed without loss of generality that the row spaces of ${C_1}^{\uparrow z}$ and ${B_3}^{\downarrow z}$ are essentially disjoint. Let B_2 be a matrix such that $\left(\begin{smallmatrix} {C_1}^{\uparrow z} \\ B_2 \end{smallmatrix}\right)$ is regular. Then, ${B_3}^{\downarrow z} = S_1{C_1}^{\uparrow z} + S_2B_2$ for some matrices S_1, S_2. Then, the matrix

$$\begin{pmatrix} {C_1}^{\uparrow x} \\ C_2 \\ B_3 - S_1{C_1}^{\uparrow x} \end{pmatrix}$$

is regular as well. Therefore, it can indeed be assumed without loss of generality that ${C_1}^{\uparrow z}$ and ${B_3}^{\downarrow z}$ are essentially disjoint. On the one hand, there then are matrices N_1, N_2 and N_3 such that

$$A = N_1{C_1}^{\uparrow x} + N_2C_2 + N_3B_3.$$

Since (C_2, c_2) is already applied to (A, z, K) by assumption, it follows that $N_2 = 0$, i.e.

$$A = N_1{C_1}^{\uparrow x} + N_3B_3.$$

Hence,

$$(A, z, K)_{(C_1,c_1)^{\uparrow x}} = (A - N_1{C_1}^{\uparrow x}, z - N_1c_1, K).$$

Let then M^* be a matrix such that

$$(A, z, K)_{(C_1,c_1)^{\uparrow x}\perp_z(C_2,c_2)} = ((A - N_1{C_1}^{\uparrow x}) - M^*C_2, (z - N_1c_1) - M^*c_2, K).$$

Let T_1 be a projection matrix such that

$$\begin{aligned} &((A, z, K)_{(C_1,c_1)^{\uparrow x}\perp_z(C_2,c_2)})^{\downarrow z} \\ &= \left(T_1(A^{\downarrow z} - N_1{C_1}^{\uparrow z} - M^*{C_2}^{\downarrow z}), T_1(z - N_1c_1 - M^*c_2), (T_1K^{-1}T_1')^{-1}\right). \end{aligned}$$

On the other hand, since $A^{\downarrow x-z} = (A - N_1{C_1}^{\uparrow x})^{\downarrow x-z}$, it also holds that

$$(A, z, K)_{\perp_z(C_2,c_2)} = (A - M^*C_2, z - M^*c_2, K).$$

Further, since $(A - M^*C_2)^{\downarrow x-z} = (A - N_1{C_1}^{\uparrow x} - M^*C_2))^{\downarrow x-z}$, it also holds that T_1 is a projection matrix such that

$$(A, z, K)_{\perp_z(C_2,c_2)}{}^{\downarrow z} = (T_1(A - M^*C_2), T_1(z - M^*c_2), (T_1K^{-1}T_1')^{-1}).$$

Then,

$$T_1A^{\downarrow z} = T_1N_1{C_1}^{\uparrow z} + T_1N_3{B_3}^{\downarrow z}.$$

Here, the row spaces of $N_1{C_1}^{\uparrow z}$ and of $N_3{B_3}^{\downarrow z}$ are essentially disjoint since those of ${C_1}^{\uparrow z}$ and of ${B_3}^{\downarrow z}$ are essentially disjoint, as observed above. It then follows that

$$\begin{aligned} &(A, z, K)_{\perp_z(C_2,c_2)}{}^{\downarrow z}{}_{(C_1,c_1)^{\uparrow z}} \\ &= \left(T_1(A^{\downarrow z} - M^*{C_2}^{\downarrow z}) - (T_1N_1){C_1}^{\uparrow z}, T_1(z - M^*c_2) - (T_1N_1)c_1, (T_1K^{-1}T_1')^{-1}\right). \end{aligned}$$

This concludes the proof. □

THEOREM **10.20**. *Symmetric Gaussian potentials with* deterministic equations $\Delta^\bullet$ *form a valuation algebra with division in the groups of the valuations of the* same domain and the same deterministic knowledge. *Symmetric Gaussian potentials with* deterministic variables *form a valuation algebra which can be embedded into the valuation algebra of symmetric Gaussian potentials with deterministic knowledge by the mapping*

$$(\mathbf{x}, \nu, C) \mapsto (A, I_{n,x}\mathbf{x}, \nu^{\uparrow x \cup y}, C^{\uparrow x \cup y}) \tag{10.72}$$

⊘

where $A = (I_{n,x}, 0_{n,y}) \in \mathbb{R}(k, x \cup y)$ *and* $n = |x|$.

PROOF. The axioms of valuation algebras are verified in turn.

(A1) In order to prove the commutativity of combination, let $\phi_1, \phi_2 \in \Delta^\bullet$. Then, if ϕ_1 or ϕ_2 is null, so is the product irrespective of the order. Else, let $\phi_1 = (C_1, c_1, \nu_1, \Lambda_1)$, $\phi_2 = (C_2, c_2, \nu_2, \Lambda_2)$ and

$$\Gamma_{certain} = \{\mathbf{x} : C_1\mathbf{x} = c_1, C_2\mathbf{x} = c_2\} = \{\mathbf{x} : C_2\mathbf{x} = c_2, C_1\mathbf{x} = c_1\}.$$

If $\Gamma_{certain} = \emptyset$, then $\phi_1 \otimes \phi_2 = z_x = \phi_2 \otimes \phi_1$. Hence, it remains to prove that the combination is commutative if $\phi_1 \otimes \phi_2$ are not contradictory. Using the commutativity of combination of symmetric Gaussian potentials,

$$((\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2))_{(C,c)} = ((\nu_2, \Lambda_2) \otimes (\nu_1, \Lambda_1))_{(C,c)} .$$

In order to prove the associativity of combination, let $\phi_1, \phi_2, \phi_3 \in \Delta^\bullet$. Then, the result of $(\phi_1 \otimes) \otimes \phi_3 = \phi_1 \otimes (\phi_2 \otimes \phi_3)$ is null if one of the three factors is inconsistent. Else, let $\phi_1 = (C_1, c_1, \nu_1, \Lambda_1)$, $\phi_2 = (C_2, c_2, \nu_2, \Lambda_2)$, and $\phi_3 = (C_3, c_3, \nu_3, \Lambda_3)$. Then, the deterministic part

$$\Gamma_{certain} = \{\mathbf{x} : C_1\mathbf{x} = c_1, C_2\mathbf{x} = c_2, C_3\mathbf{x} = c_3\}$$

does not depend on the associations. Let $(C, c) = (C_1, c_1) \oplus (C_2, c_2) \oplus (C_3, c_3)$. Then, using the commutativity of applying deterministic knowledge and of combining symmetric Gaussian potentials (Lemma 10.8), the idempotency of applying deterministic knowledge (Lemma 10.7), and the associativity of the combination of symmetric Gaussian potentials,

$$\begin{aligned}
&((\nu_1, \Lambda_1)_{(C_1,c_1)\oplus(C_2,c_2)} \otimes (\nu_2, \Lambda_2)_{(C_1,c_1)\oplus(C_2,c_2)})_{(C,c)} \otimes (\nu_3, \Lambda_3)_{(C,c)}\\
=&(((\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2))_{(C_1,c_1)\oplus(C_2,c_2)})_{(C,c)} \otimes (\nu_3, \Lambda_3)_{(C,c)}\\
=&((\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2))_{(C,c)} \otimes (\nu_3, \Lambda_3)_{(C,c)}\\
=&[((\nu_1, \Lambda_1) \otimes (\nu_2, \Lambda_2)) \otimes (\nu_3, \Lambda_3)]_{(C,c)}\\
=&[(\nu_1, \Lambda_1) \otimes ((\nu_2, \Lambda_2) \otimes (\nu_3, \Lambda_3))]_{(C,c)}\\
=&\ldots\\
=&(\nu_1, \Lambda_1)_{(C,c)} \otimes ((\nu_2, \Lambda_2)_{(C_2,c_2)\oplus(C_3,c_3)} \otimes (\nu_3, \Lambda_3)_{(C_2,c_2)\oplus(C_3,c_3)})_{(C,c)}.
\end{aligned}$$

(A2) The labelling axiom holds by the definition of combination (10.49) and (10.51).

(A3) The marginalisation axiom holds by the definition of marginalisation (10.66) and (10.68).

(A4) The transitivity axiom clearly holds for inconsistent symmetric Gaussian potentials with deterministic equations. So let $(C, c, \nu, \Lambda) \in \Delta^{\bullet}$ be a consistent potential on the domain x. Let (A, z, K) be a Gaussian hint with domain x inducing (ν, Λ). Partition $A = (A_1, A_2, A_3)$ such that $A_1 \in \mathbb{R}(m, x_1)$, $A_2 \in \mathbb{R}(m, x_2)$ and $A_3 \in \mathbb{R}(m, x_3)$. It has to be shown that

$$\left(\left(\left((A, z, K)_{\perp x_1 \cup x_2 (C,c)}\right)^{\downarrow x_1 \cup x_2}{}_{\perp x_1(\tilde{C},\tilde{c})}\right)^{\downarrow x_1} = \left((A, z, K)_{\perp x_1 (C,c)}\right)^{\downarrow x_1} \tag{10.73}$$

for $(\tilde{C}, \tilde{c}) = (C, c)^{\downarrow x_1 \cup x_2}$. Here, the same notation for the rotation is used on the associated Gaussian linear system instead of symmetric Gaussian potentials.

1. First, projection matrices for the deterministic part are defined. Let $C = (C_1, C_2, C_3)$ such that $C_1 \in \mathbb{R}(k, x_1)$, $C_2 \in \mathbb{R}(k, x_2)$ and $C_3 \in \mathbb{R}(k, x_3)$ and let $C_{23} = (C_2, C_3)$ and $r_3 = \mathrm{r}(C_3)$ and $r_{23} = \mathrm{r}(C_{23})$. Let $S = \left(\begin{smallmatrix} S_{12} \\ T_3 \end{smallmatrix}\right)$ be a regular matrix such that S_{12} is a projection matrix of rank $k - r_3$ for C_3 in C. Since S is regular, the matrix

$$\begin{pmatrix} S_{12} \\ T_3 \end{pmatrix} (C_2, C_3) = \begin{pmatrix} S_{12}C_2 & 0 \\ T_3C_2 & T_3C_3 \end{pmatrix}$$

preserves the rank r_{23} of (C_2, C_3). Since T_3C_3 has full row rank r_3, it follows that $S_{12}C_2$ has rank $r_{23} - r_3$. Further, let $R = \left(\begin{smallmatrix} R_1 \\ R_2 \end{smallmatrix}\right)$ be a regular matrix such that R_1 is projection matrix of rank $k - r_{23}$ for $S_{12}C_2$. Hence, define the regular matrix $T \in \mathbb{R}(k, k)$

$$T = \begin{pmatrix} R_1 & 0 \\ R_2 & \\ 0 & I_{r_3} \end{pmatrix} \begin{pmatrix} S_{12} \\ T_3 \end{pmatrix} = \begin{pmatrix} R_1 S_{12} \\ R_2 S_{12} \\ T_3 \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix},$$

$T_1 \in \mathbb{R}(k - r_{23}, k)$, $T_2 \in \mathbb{R}(r_{23} - r_3, k)$, $T_3 \in \mathbb{R}(r_3, k)$. Then, it holds that

$$TC = \begin{pmatrix} T_1C_1 & & \\ T_2C_2 & T_2C_2 & \\ T_3C_1 & T_3C_2 & T_3C_3 \end{pmatrix}.$$

2. Now, the non-deterministic parts can be "rotated" to get $(A, z, K)_{\perp x_1 \cup x_2 (C,c)}$ and $(A, z, K)_{\perp x_1 (C,c)}$. There is a matrix $W_2' \in \mathbb{R}(w, k)$ for $w = |x_2 \cup x_3| - r_{23}$ such that

$$\begin{pmatrix} T_2C_2 & \\ T_3C_2 & T_3C_3 \\ \multicolumn{2}{c}{W_2'} \end{pmatrix}$$

is regular. Then, there are matrices $N_1 \in \mathbb{R}(m, r_{23}), N_2 \in \mathbb{R}(m, w)$ such that

$$(A_2, A_3) = N_1 \begin{pmatrix} T_2C_2 & \\ T_3C_2 & T_3C_3 \end{pmatrix} + N_2 W_2'.$$

Decompose $N_1 = (N_{11}, N_{12})$ and $N_2 = (N_{21}, N_{22})$ such that

$$A_2 = N_{11}T_2C_2 + N_{12}T_3C_2 + N_{21}W_2', \quad A_3 = N_{12}T_3C_3 + N_{22}W_2'.$$

Define

$$M_{23} = N_1 \begin{pmatrix} T_2 \\ T_3 \end{pmatrix}, \quad M_3 = N_{12}T_3 \quad \text{and} \quad M_2 = P_3(N_{11}, 0).$$

Therefore, rotating the Gaussian hint (A, z, K) with respect to (C, c) for x_1 yields the Gaussian linear system

$$(A, z, K)_{\perp x_1(C,c)}(A - M_{23}C, z - M_{23}z, K);$$

similarly, rotating for $x_1 \cup x_2$ yields the Gaussian linear system

$$(A, z, K)_{\perp x_1 \cup x_2(C,c)} = (\tilde{A}, \tilde{z}, K) = (A - M_3C, z - M_3z, K).$$

3. Now, $(A, z, K)_{\perp x_1 \cup x_2(C,c)} = (\tilde{A}, \tilde{z}, K)$ can be dealt with. Let P_3 be a projection matrix for $(\tilde{A}, \tilde{z}, K)$ eliminating the variables x_3 such that

$$P_3(A_3 - M_3C_3) = 0.$$

Define

$$(\tilde{C}, \tilde{c}) = (T_{12}(C_1, C_2), T_{12}c)$$

for $T_{12} = \binom{T_1}{T_2}$. Then, it holds that

$$P_3(A_2 - M_3C_2) = P_3(N_{11}T_2C_2 + N_{21}W_2').$$

Therefore, M_2 rotates the Gaussian linear system $(\tilde{A}, \tilde{z}, K)^{\downarrow x_1 \cup x_2}$ for x_1 with respect to $(\tilde{C}, \tilde{c})$. Let P_2 be a projection matrix to x_1 such that

$$P_2\left(\tilde{A}_2 - M_2\tilde{C}_2 = 0\right)$$

where $\tilde{A}_2$ and $\tilde{C}_2$ are the columns of $\tilde{A}$ and $\tilde{C}$ corresponding to the variables x_2.

4. It has can now be verified that (10.73) holds. Define $P_{23} = P_2P_3$. It holds that

$$\begin{aligned}
P_{23}(A_2 - M_{23}C_2) &= P_2P_3\left(A_2 - N_1\begin{pmatrix} T_2C_2 \\ T_3C_2 \end{pmatrix}\right) \\
&= P_2P_3\left(A_2 - N_{12}T_3C_2 - N_{11}T_2C_2\right) \\
&= P_2P_3\left(A_2 - N_{12}T_3C_2 - N_{11}T_2C_2\right) \\
&= P_2\left(P_3(A_2 - M_3C_3) - M_2T_{12}C_2\right) \\
&= P_2\left(P_3(A_2 - M_3C_3) - M_2\tilde{C}_2\right) \\
&= P_2(\tilde{A}_2 - M_2\tilde{C}_2) \\
&= 0
\end{aligned}$$

and

$$\begin{aligned} P_{23}(A_3 - M_{23}C_3) &= P_2P_3(A_3 - N_1\begin{pmatrix} 0 \\ T_3C_3 \end{pmatrix}) \\ &= P_2P_3(A_3 - M_3C_3) \\ &= 0. \end{aligned}$$

By the same reasoning as above for T, it can be verified that P_{23} has full row rank $m-\mathrm{r}(A_{23}-M_{23}C_{23})$. Hence, P_{23} is a projection matrix for $(A-M_{23}C)$ to x_1. Further,

$$\begin{aligned} P_2(P_3(A_1 - M_3C_1) - M_2\tilde{C}_1) &= P_2((P_3(A_1 - M_3C_1) - P_3N_{11}T_2C_1) \\ &= P_2P_3(A_1 - N_{12}T_3C_1 - N_{11}T_{12}C_1) \\ &= P_{23}(A_1 - M_{23}C_1). \end{aligned}$$

Finally,

$$\begin{aligned} (P_{23}K^{-1}P_{23})^{-1} &= (P_2((P_3K^{-1}P_3')^{-1})^{-1}P_2')^{-1} \\ &= (P_{23}K^{-1}P_{23}'). \end{aligned}$$

(A5) On the one hand, the combination trivially holds if at least one of the factors is null since the result is then null in every case.

On the other hand, let $\phi_1 = (C_1, c_1, \nu_1, \Lambda_1)$ and $\phi_2 = (C_2, c_2, \nu_2, \Lambda_2)$ be symmetric Gaussian potentials with deterministic equations on domains s and t respectively. Let m_1 and m_2 be such that $C_1 \in \mathbb{R}(m_1, s)$ and $C_2 \in \mathbb{R}(m_2, t)$.

Let $z \in \mathcal{M}(\phi_1 \otimes \phi_2)$ such that $s \subseteq z \subseteq s \cup t$. Let (C, c) be the combined linear system where $C = (\begin{smallmatrix} C_1^{\uparrow s\cup t} \\ C_2^{\uparrow s\cup t} \end{smallmatrix})$. Let P_2 be a projection matrix to $t \cap z$ for (C_2, c_2), i.e. a matrix of full row rank $m_2 - \mathrm{r}(C_2{}^{\downarrow t-z})$ such that $P_2C^{\downarrow t-z} = 0_{m_2,t-z}$. Define $m = m_1 + m_2$ and the regular matrix $P \in \mathbb{R}(m, m)$ by

$$P = \begin{pmatrix} I_{m_1} & 0_{m_1,m_2} \\ 0_{m_2,m_1} & P_2 \end{pmatrix}.$$

Observe that (C, c) is inconsistent if and only if $(PC^{\downarrow z}, Pc)$ is inconsistent. This holds since

$$\begin{aligned} \{\mathbf{u}^{\downarrow z} : C\mathbf{u} = c\} &= \{\mathbf{u}^{\downarrow z} : C_1{}^{\uparrow s\cup t}\mathbf{u} = c_1\} \cap \{\mathbf{u}^{\downarrow z} : C_2{}^{\uparrow s\cup t}\mathbf{u} = c_2\} \\ &= \{\mathbf{z} : C_1{}^{\uparrow z}\mathbf{z} = c_1\} \cap \{\mathbf{z} : P_2C_2{}^{\to z}\mathbf{z} = Pc_2\} \\ &= \{\mathbf{z} : PC^{\downarrow z}\mathbf{z} = Pc\}. \end{aligned}$$

Observe that $(\phi_1 \otimes \phi_2)^{\downarrow z} = z_z$ only if $\phi_1 \otimes \phi_2 = z_{s\cup t}$. Therefore,

$$(\phi_1 \otimes \phi_2)^{\downarrow z} = z_z \iff \phi_1 \otimes \phi_2{}^{\downarrow t\cap z} = z_z.$$

Assume that (C, c) is consistent. In order to prove the combination axiom in this case, it will now be shown that

$$\left(PC^{\downarrow z}, Pc, \left((\nu_1, \varLambda_1) \otimes (\nu_2, \varLambda_2)\right)_{(C,c)\perp_z(C,c)}{}^{\downarrow z}\right)$$
$$=(C_1, c_1, \nu_1, \varLambda_1) \otimes (C_2, c_2, \nu_2, \varLambda_2)^{\downarrow t\cap z}.$$

It holds that

$$PC = \begin{pmatrix} {C_1}^{\uparrow z} & 0_{m_1,t-z} \\ \left(P_2 {C_2}^{\downarrow t\cap z}\right)^{\uparrow z} & 0_{m_2,t-z} \end{pmatrix}.$$

Hence, P is a projection matrix for C since it has full row rank $m - \mathrm{r}(C^{\downarrow t-z}) = m_1 + m_2 - \mathrm{r}({C_2}^{\downarrow t-z})$ and since $(PC)^{\downarrow t-z} = 0_{m,t-z}$. Therefore,

$$(C, c)^{\downarrow z} = (C_1, c_1)^{\uparrow z} \oplus (C_2, c_2)^{\downarrow t\cap z\uparrow z}.$$

Furthermore,

$$\left((\nu_1, \varLambda_1) \otimes (\nu_2, \varLambda_2)\right)_{(C,c)\perp_z(C,c)}{}^{\downarrow z}$$
$$\overset{(1)}{=}\left((\nu_1, \varLambda_1)^{\uparrow s\cup t}{}_{(C_2,c_2)^{\uparrow s\cup t}} \otimes (\nu_2, \varLambda_2)^{\uparrow s\cup t}{}_{(C_1,c_1)^{\uparrow s\cup t}}\right)_{\perp_z(C,c)}{}^{\downarrow z}$$
$$\overset{(2)}{=}\left((\nu_1, \varLambda_1)^{\uparrow s\cup t}{}_{(C_2,c_2)^{\uparrow s\cup t}} \otimes (\nu_2, \varLambda_2)^{\uparrow s\cup t}{}_{(C_1,c_1)^{\uparrow s\cup t}}\right)_{\perp_z(C_1,c_1)^{\uparrow s\cup t}\oplus(C_2,c_2)^{\uparrow s\cup t}}{}^{\downarrow z}$$
$$\overset{(3)}{=}\left((\nu_1, \varLambda_1)^{\uparrow s\cup t}{}_{(C_2,c_2)^{\uparrow s\cup t}} \otimes (\nu_2, \varLambda_2)^{\uparrow s\cup t}{}_{(C_1,c_1)^{\uparrow s\cup t}}\right)_{\perp_z(C_2,c_2)^{\uparrow s\cup t}}{}^{\downarrow z}$$
$$\overset{(4)}{=}\left((\nu_1, \varLambda_1)^{\uparrow s\cup t}{}_{(C_2,c_2)^{\downarrow s\cap t\uparrow s\cup t}} \otimes (\nu_2, \varLambda_2)^{\uparrow s\cup t}{}_{(C_1,c_1)^{\downarrow s\cap t\uparrow s\cup t}}\right)_{\perp_z(C_2,c_2)^{\uparrow s\cup t}}{}^{\downarrow z}$$
$$\overset{(5)}{=}\left((\nu_1, \varLambda_1)_{(C_2,c_2)^{\downarrow s\cap t\uparrow s}}{}^{\uparrow s\cup t} \otimes (\nu_2, \varLambda_2)_{(C_1,c_1)^{\downarrow s\cap t\uparrow t}}{}^{\uparrow s\cup t}\right)_{\perp_z(C_2,c_2)^{\uparrow s\cup t}}{}^{\downarrow z}$$
$$\overset{(6)}{=}\left((\nu_1, \varLambda_1)_{(C_2,c_2)^{\downarrow s\cap t\uparrow s}} \otimes (\nu_2, \varLambda_2)_{(C_1,c_1)^{\downarrow s\cap t\uparrow t}}\right)_{\perp_z(C_2,c_2)^{\uparrow s\cup t}}{}^{\downarrow z}$$
$$\overset{(7)}{=}\left((\nu_1, \varLambda_1)_{(C_2,c_2)^{\downarrow s\cap t\uparrow s}} \otimes (\nu_2, \varLambda_2)_{(C_1,c_1)^{\downarrow s\cap t\uparrow t}\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z}$$
$$\overset{(8)}{=}(\nu_1, \varLambda_1)_{(C_2,c_2)^{\downarrow s\cap t\uparrow s}} \otimes \left((\nu_2, \varLambda_2)_{(C_1,c_1)^{\downarrow s\cap t\uparrow t}\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z\cap t}$$
$$\overset{(9)}{=}(\nu_1, \varLambda_1)_{(C_2,c_2)^{\downarrow s\cap t\uparrow s}} \otimes \left((\nu_2, \varLambda_2)_{\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z\cap t}{}_{(C_1,c_1)^{\downarrow s\cap t\uparrow z\cap t}}$$
$$\overset{(10)}{=}(\nu_1, \varLambda_1)^{\uparrow z}{}_{(C_2,c_2)^{\downarrow z\cap t\uparrow z}} \otimes \left((\nu_2, \varLambda_2)_{\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z\cap t\uparrow z}{}_{(C_1,c_1)^{\downarrow s\cap t\uparrow z}}$$
$$\overset{(11)}{=}(\nu_1, \varLambda_1)^{\uparrow z}{}_{(C_2,c_2)^{\downarrow z\cap t\uparrow z}} \otimes \left((\nu_2, \varLambda_2)_{\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z\cap t\uparrow z}{}_{(C_1,c_1)^{\uparrow z}}$$
$$\overset{(12)}{=}\left((\nu_1, \varLambda_1)^{\uparrow z} \otimes \left((\nu_2, \varLambda_2)_{\perp_{z\cap t}(C_2,c_2)}\right)^{\downarrow z\cap t\uparrow z}\right)_{(C,c)^{\downarrow z}},$$

using the definition of vacuous extension and Lemma 10.11 in (1) and (12), using Lemma 10.17 in (3), using Lemma 10.9 in (4) and (11) [observing that $s \cap (z \cap t)= (s \cap z) \cap t = s \cap t$], using the definition of vaucous extension and of neutral elements, the transitivity of vacuous extension, and Lemma 10.10 in (5) and (10), using the definition of vacuous extension in (6), using Lemma 10.18 in (7), using the combination axiom of symmetric Gaussian potentials in (8), and using Lemma 10.19 in (9) [observing that $s \cap t \subseteq z \cap t$].

(A6) On the one hand, since the deterministic part does not change and since $(\nu, \Lambda)_{(C,c)} = (\nu, \Lambda)$ in light of Lemma 10.7, it holds that $x \in \mathcal{M}(\nu, \Lambda) = \mathcal{M}(C, c, \nu, \Lambda)$ and $(C, c, \nu, \Lambda)^{\downarrow x} = (C, c, (\nu, \Lambda)^{\downarrow x}) = (C, c, \nu, \Lambda)$. On the other hand, $x \in \mathcal{M}(z_x)$ and $z_x{}^{\downarrow x} = z_x$. This shows that the domain axiom holds.

(A7) The element $e = (\diamond, \diamond, \diamond, \diamond)$ with $d(\diamond, \diamond, \diamond, \diamond) = \emptyset$ is the identity element.

(A8) The elements $e_x = (0_x, 0_x, 0_x, 0_x)$ are neutral elements for the domain $x \in D$. Since $e_x \otimes e_y = (0_{x\cup y}, 0_{x\cup y}, 0_{x\cup y}, 0_{x\cup y}) = e_{x\cup y}$, the neutrality axioms holds, too.

(A9) Stability holds since $y \in \mathcal{M}(e_x) = \mathcal{M}(0_x, 0_x)$ for $y \subseteq x$ and

$$e_x{}^{\downarrow y} = (0_x, 0_x, 0_x, 0_x)^{\downarrow y} = (0_y, 0_y, 0_y, 0_y) = e_y.$$

(A10) The nullity axiom follows from the definition of marginalisation of null elements, equation (10.68).

The symmetric Gaussian potentials of the same domain with the same deterministic knowledge (up to equivalence) form a group.
It can be verified that the mapping (10.72) is injective and compatible with combination and marginalisation. □

10.7 Gaussian Belief Functions

Gaussian belief functions (GBF) or linear belief functions (LBF) are moment matrices with deterministic knowledge (Dempster, 1990a; Liu, 1996a;b; 1999; Liu et al., 2003a;b; Srivastava and Liu, 2003). Two representations are going to be discussed: moment matrices with deterministic variables and partially swept moment matrices.

Moment Matrices with Deterministic Variables

In (Liu, 1996a;b; 1999), variables are assumed to fall in three categories:

- *deterministic* variables,
- *uncertain* variables, and
- *vacuous* variables.

The knowledge about the deterministic variables is represented by the vector of their fixed values and the knowledge about the uncertain variables by a moment matrix. In the terminology of the preceding sections, this means that the sets $\Gamma_{certain}$ and

$\Gamma_{uncertain}(\omega)$ are spanned by different sets of variables and that the uncertain variables form a precise Gaussian hint (or, equivalently, a symmetric Gaussian potential with positive definite concentration matrix). See Figure 7.2.

Formally, moment matrices with deterministic variables are quadruples

$$M = (\mathbf{x}, \mu, \Sigma, s) \tag{10.74}$$

where

- $\mathbf{x} \in \mathbb{R}^x$,
- $\mu \in \mathbb{R}^y$, $\Sigma \in \mathbb{R}(y, y)$ symmetric and positive definite,
- $x \cap y = \emptyset$, and
- $x \cup y \subseteq s \in D$.

The variables x are *deterministic*, the variables y *uncertain*, and the variables $s - (x \cup y)$ are *vacuous*. M corresponds to a symmetric Gaussian potential with deterministic variables

$$(\mathbf{x}, \left(\Sigma^{-1}\mu\right)^{\uparrow s-x}, \left(\Sigma^{-1}\right)^{\uparrow s-x}). \tag{10.75}$$

The rules of combination and marginalisation derived in Section 10.1 can be easily carried over to moment matrices. The marginal of M with respect to $t \subseteq s$ is then

$$(\mathbf{x}^{\downarrow t\cap x}, \mu^{\downarrow t\cap y}, \Sigma^{\downarrow t\cap y}, t). \tag{10.76}$$

The combination of $(\mathbf{x}_1, \mu_1, \Sigma_1, s_1)$ and $(\mathbf{x}_2, \mu_2, \Sigma_2, s_2)$ is given by

- the deterministic vector $\mathbf{x}_1 \bowtie \mathbf{x}_2$ (if $\mathbf{x}_1$ and $\mathbf{x}_2$ are compatible),
- the moment matrix $\left[\triangleright(M_1, {\mathbf{x}_2}^{\downarrow y_1\cap x_2})\right]^{\downarrow_M y_1 - x_2} \otimes_M \left[\triangleright(M_2, {\mathbf{x}_1}^{\downarrow y_2\cap x_1})\right]^{\downarrow_M y_2 - x_1}$ for $M_1 = (\mu_1, \Sigma_1)$, $M_2 = (\mu_2, \Sigma_2)$, and
- the label $s_1 \cup s_2$.

Notice that for $M = \sigma(\phi)$, it holds that

$$\sigma(\phi_{x|\mathbf{z}}) = \triangleright(M, \mathbf{z})^{\downarrow_M x}. \tag{10.77}$$

These rules for combination and marginalisation reproduce those in (Liu, 1996a; 1999), where combination was derived directly from Dempster's Rule.

Partially Swept Moment Matrices with Deterministic Knowledge

As discussed in Section 9.6, Gaussian linear systems can be represented by *partially swept moment matrices*, which can neither be fully swept nor completely unswept. Furthermore, (Dempster, 1990a) claims that even deterministic linear equations without Gaussian term can be fitted into this framework. In fact, he implicitly defines *partially swept moment matrices* as the combination of building blocks of three types:

- deterministic knowledge,
- probabilistic Gaussian knowledge, and
- vacuous knowledge.

The approach is taken up in (Liu et al., 2003a;b; Srivastava and Liu, 2003).

Consider a linear system

$$x_1 = Ax_2 + \mu_1. \tag{10.78}$$

(Dempster, 1990a) suggests representing such such a linear system by

$$M(x_1, \vec{x_2}) = \left(\begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0_{x_1} & A \\ A' & 0_{x_2} \end{pmatrix} \right),$$

where the variables x_1 are unswept and the variables x_2 are swept forward. As it stands, such a matrix $M(x_1, \vec{x_2})$ can neither be fully swept nor fully unswept. However, regarding $M(x_1, \vec{x_2})$ as the limit

$$M(x_1, \vec{x_2}) = \lim_{\epsilon \to 0} \left(\begin{pmatrix} \mu_1 \\ 0 \end{pmatrix}, \begin{pmatrix} \epsilon \cdot I_{x_1} & A \\ A' & 0_{x_2} \end{pmatrix} \right),$$

The corresponding fully swept matrices are

$$M(\vec{x_1}, \vec{x_2})_\epsilon = \left(\begin{pmatrix} \epsilon^{-1} \cdot \mu_1 \\ -\epsilon^{-1} \cdot A'\mu_1 \end{pmatrix}, \begin{pmatrix} \epsilon^{-1} \cdot I_{x_1} & \epsilon^{-1} \cdot A \\ \epsilon^{-1} \cdot A' & -\epsilon^{-1} \cdot A'A \end{pmatrix} \right)$$

for $\epsilon > 0$. He conjectures that such a matrix $M(\vec{x_1}, \vec{x_2})_\epsilon$ can be combined in the usual way with another partially swept moment matrix: by taking the sum, by sweeping backwards, and by replacing the ϵ's by 0. In this way, he defines generalised moment matrices implicitly as combinations of these building blocks: (regular) moment matrices, vacuous variables, and deterministic equations.

As shown above, the representation of probabilistic and of vacuous knowledge has a close resemblance to symmetric Gaussian potentials. In contrast, the representation of deterministic knowledge is tricky as discussed in the following two conjectures.

Conjecture **10.21**. *Only the forms without ϵ are a unique representation of the Gaussian linear information. Therefore, the combination of generalised moment matrices with deterministic knowledge is not always well defined. However, Dempster's ϵ-trick works fine for switching between these representations.* ⊘

Consider the following linear equation

$$Y = -0.5 \cdot X + 1$$

Multiplying it by -2, one obtains the equivalent linear equation

$$X = -2 \cdot Y + 2.$$

The corresponding partially swept matrices are

$$M(\vec{X}, Y) = \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 & -0.5 \\ -0.5 & 0 \end{pmatrix}\right)$$

and

$$\tilde{M}(X, \vec{Y}) = \left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & -2 \\ -2 & 0 \end{pmatrix}\right).$$

Then, the ϵ-swept matrices are not the same:

$$M(\vec{X}, \vec{Y})_\epsilon = \left(\begin{pmatrix} \epsilon^{-1} \cdot 0.5 \\ \epsilon^{-1}. \end{pmatrix}, \begin{pmatrix} -\epsilon^{-1} \cdot 0.25 & -\epsilon^{-1} \cdot 0.5 \\ -\epsilon^{-1} \cdot 0.5 & -\epsilon^{-1} \end{pmatrix}\right)$$

and

$$\tilde{M}(\vec{X}, \vec{Y})_\epsilon = \left(\begin{pmatrix} 2\epsilon^{-1}. \\ 4\epsilon^{-1}. \end{pmatrix}, \begin{pmatrix} -\epsilon^{-1} & -\epsilon^{-1} \cdot 2 \\ -\epsilon^{-1} \cdot 2 & -\epsilon^{-1} \cdot 4 \end{pmatrix}\right)$$

(notice that $M(\vec{X}, \vec{Y})_\epsilon$ is a multiple of $\tilde{M}(\vec{X}, \vec{Y})_\epsilon$). However, sweeping backwards yields

$$M(X, \vec{Y}) = \tilde{M}(X, \vec{Y})$$

and

$$\tilde{M}(\vec{X}, Y) = M(\vec{X}, Y).$$

CONJECTURE **10.22**. *If the common variables are swept forward without Dempster's ϵ-trick, the combined information is the sum of these matrices.* ⊘

As shown in Section 3.5, the combination of moment matrices only requires sweeping forward on the common variables. Furthermore, (Dempster, 1990a) gives an example of a partially swept moment matrix (without fully swept representation) where the combination works if the common variables are swept forward. For instance, a Gaussian linear system $x_1 = Ax_2 + \mu_1 + \omega$ with covariance Σ discussed above can be built from the probabilistic knowledge about x_1 and the deterministic knowledge about x_1 given x_2: Combining

$$M_{prob}(\vec{\omega}) = \left[\left(\Sigma^{-1}\mu_1\right), \left(-\Sigma^{-1}\right)\right]$$

and

$$M_{det}(x_1, \vec{x_2}, \vec{\omega}) = \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & A & I \\ A' & 0 & 0 \\ I & 0 & 0 \end{pmatrix}\right].$$

Then, sweeping vacuously extending $M_{prob}(\vec{\omega})$ and building the sum with $M_{det}(\vec{x_1}, \vec{x_2}, \vec{\omega})_\epsilon$ yields

$$M(\vec{x_1}, \vec{x_2}, \vec{\omega})_\epsilon = \left[\begin{pmatrix} 0 \\ 0 \\ \Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} -\epsilon^{-1} \cdot I & \epsilon^{-1} \cdot A & \epsilon^{-1} \cdot I \\ \epsilon^{-1} \cdot A' & -\epsilon^{-1} \cdot A'A & -\epsilon^{-1} \cdot A' \\ \epsilon^{-1} \cdot I & -\epsilon^{-1} \cdot A & -\epsilon^{-1} \cdot I - \Sigma^{-1} \end{pmatrix}\right].$$

Then, sweeping backwards on x_1

$$M(x_1, \vec{x_2}, \vec{\omega})_\epsilon = \left[\begin{pmatrix} 0 \\ 0 \\ \Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \epsilon \cdot I & A & I \\ A' & 0 & 0 \\ I & 0A & -\Sigma^{-1} \end{pmatrix} \right]$$

and, letting $\epsilon = 0$,

$$M(x_1, \vec{x_2}, \vec{\omega}) = \left[\begin{pmatrix} 0 \\ 0 \\ \Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} 0 & A & I \\ A' & 0 & 0 \\ I & 0 & -\Sigma^{-1} \end{pmatrix} \right]$$

Here, it holds that

$$M(x_1, \vec{x_2}, \vec{\omega}) = M_{det}(x_1, \vec{x_2}, \vec{\omega}) \oplus M_{prob}(\vec{\omega})$$

Finally, sweeping backwards on ω,

$$M(x_1, \vec{x_2}, \omega) = \left[\begin{pmatrix} \mu_1 \\ 0 \\ \Sigma^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma & A & \Sigma \\ A' & 0 & 0 \\ \Sigma & 0 & \Sigma \end{pmatrix} \right],$$

and eliminating ω yields the same partially swept moment matrix as above in equation (9.29).

If marginalisation corresponds to retaining the pertinent entries of a fully unswept matrix, similar inconsistencies arise as in the case of combination, i.e. the fully unswept form is not well defined in general.

Chapter Synopsis & Discussion

Two approaches to deterministic knowledge have been discussed in this chapter:

- Either some variables may take a fixed value or, more generally,
- linear combinations of variables may take a fixed value.

(Liu, 1996a; 1999) showed how Dempster's Rule of combination can be applied to moment matrices with variables taking a fixed value. The restriction to this setting was motivated by problems with Dempster's more general approach of partially swept moment matrices (Dempster, 1990a).

In this chapter, algorithms for the general case of deterministic equations have been developed for Gaussian hints and symmetric Gaussian potentials as follows.

- *Application of deterministic knowledge*: Deterministic equations induce an event in Gaussian linear equations since some assumptions are ruled out by the deterministic knowledge. Technically, the deterministic knowledge has to be "substituted into" the Gaussian linear system. Geometrically, this corresponds to a projection and a rotation of the non-deterministic part. The overall Gaussian linear information can be represented by a symmetric Gaussian potential with deterministic knowledge. This representation is unique up to equivalence in the deterministic part.

- *Combination* amounts to applying the joint deterministic knowledge to both symmetric Gaussian potentials and their combination. This complies with Dempster's Rule of intersection of focal sets.

- *Marginalisation*: Geometrically, marginalisation corresponds to the projection of focal sets. This projection can be achieved by first rotating and projecting the non-deterministic part. Then, the deterministic and the modified non-deterministic part can be treated independently.

Part IV

Applications and Implementation

11

Kalman Filter Models and Local Computation

Temporal Gaussian linear models are widely used, for instance in control theory (Kalman, 1960) and coding theory (MacKay, 2003). The basic model is often the following: The state of a system cannot be monitored directly but only through measurements with additive Gaussian noise at discrete moments of time. The state evolves over time according to a linear transition function of the previous state and additive Gaussian noise only.

Although terminology and interpretation in these different fields vary considerably, the basic model and inference algorithms turn out to be essentially the same (Roweis and Ghahramani, 1999). As shown in (Monney, 2003), assumption-based reasoning reproduces the results of inference based on the least squares or the maximum likelihood principle or by applying Bayes' rule. However, the interpretation is completely different. The application of the theory of Gaussian belief functions to these Kalman filter models has already been discussed in (Dempster, 1990a;b), and the application to that of Gaussian hints in (Monney, 2003; Kohlas and Monney, 2008).

Based on the noisy observations, there are three basic inference tasks in such a model: *filtering* the current value, *smoothing* past values, and *predicting* future values. These three poblems can be formulated as projection problems, which can be solved by local computation using the techniques developed in parts I and II.

Chapter Outline

The Kalman filter model as well as the filtering, smoothing, and prediction problems, will be formally defined in Section 11.1. The recursion in the Kalman filter, smoothing and prediction algorithms can easily be translated into the message-passing scheme of the collect algorithm. This is the topic of Sections 11.2 to 11.4. The different parts of the Kalman filter model are represented by symmetric Gaussian potentials.

11.1 The Kalman Filter Model

According to (Roweis and Ghahramani, 1999), the basic model of discrete time linear dynamical systems with additive Gaussian noise is as follows:

$$x_{k+1} = A_k x_k + \omega_k, \tag{11.1}$$

$$y_k = H_k x_k + \nu_k, \tag{11.2}$$

$$y_k = \mathbf{y}_k, \tag{11.3}$$

$$x_1 = \omega_0 \tag{11.4}$$

for $k \in \{1, 2, \ldots\}$ and where the disturbances ω_k and ν_k are distributed normally with mean 0 and variance-covariance Q_k and R_k, respectively. Such a Gaussian linear system is often called a Kalman filter model. Here, equation (11.1) defines a first-order Markov state evolution process, where the matrix $A_k \in \mathbb{R}(x_{k+1}, x_k)$ is called transition matrix and the state vector x_k of real-valued variables. Furthermore, equations (11.3) and (11.2) define the observation process: At each time step k, an output or observation $\mathbf{y}_k \in \mathbb{R}^{y_k}$ is obtained from the unknown state $\mathbf{x}_k$ through the observation matrix (also measurement or generative matrix) $H_k \in \mathbb{R}(p, k)$. Both noise sources are independent from each other as well as from time step to time step. Finally, the initial state $\mathbf{x}_1$ in equation (11.4) is given by the Gaussian noise ω_0 which is assumed Gaussian with mean μ_0 and covariance Q_0. Notice that the restriction to zero-mean noise sources does not infringe upon generality as observed by (Roweis and Ghahramani, 1999; p.307) since a non-zero mean can always be simulated by adding a dimension to the state or the observation model. Of course, the equations of the Kalman filter model can be brought into the standard form of a Gaussian linear system as follows:

$$(A_k, -I) \begin{pmatrix} x_k \\ x_{k+1} \end{pmatrix} + \omega_k = 0, \tag{11.5}$$

$$H_k x_k + \nu_k = \mathbf{y}_k, \tag{11.6}$$

$$x_1 = \omega_0 \tag{11.7}$$

Notice that the deterministic equation 11.3 has been directly substituted into (11.2) without using the theory of Chapter 10. This is possible since the variable y_k only occurs here (see Figures 11.2 and 11.3 below).

Figure 11.1 shows the block diagram of the functional model of equations (11.1) and (11.2): Boxes represent functions and arrows input and output of these functions; branching points are denoted by a black dot.

If the state and the observation model are not entirely known, a full model which best explains the observed data has first to be discovered. The problem here is to *learn* the model or to *identify the system*. Typically, this is the case in speech recognition or in social sciences, where economical and performative models have be found. However, this is not an inference task, and it will therefore not be discussed here.

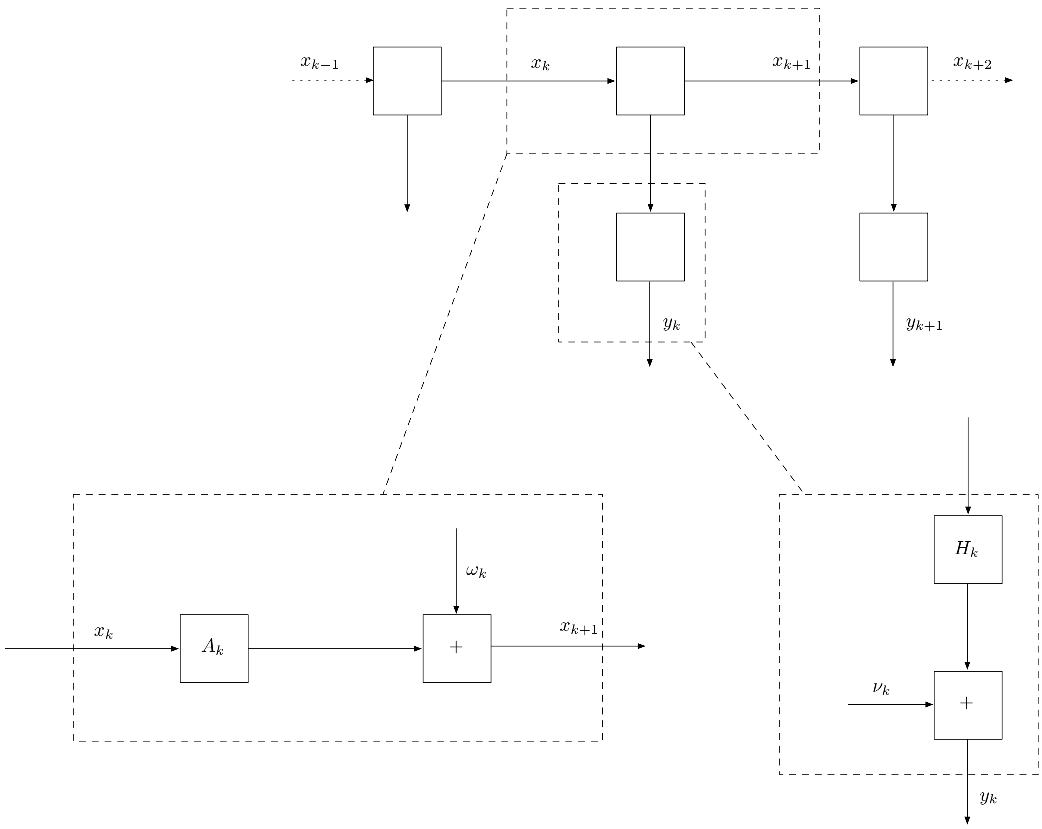


Figure 11.1: A block diagram for the Kalman filter model

The whole model can be decomposed into independent factors. Let $\mathcal{O}(k)$ be the symmetric Gaussian potential inferred from equations (11.2) and (11.3) on x_k, which is

$$\mathcal{O}(k) = (H'_k {R_k}^{-1} H_k, H'_k {R_k}^{-1} \mathbf{y}_k). \tag{11.8}$$

On the other hand, the state transition equation (11.1) yields

$$\mathcal{S}(k, k+1) = (\begin{pmatrix} A'_k {Q_k}^{-1} A_k & -A'_k {Q_k}^{-1} \\ -{Q_k}^{-1} A_k & {Q_k}^{-1} \end{pmatrix}, 0). \tag{11.9}$$

If an initial distribution (μ_0, Q_0) is given for x_1, then define

$$\mathcal{S}(0) = ({Q_0}^{-1}, {Q_0}^{-1}\mu_0); \tag{11.10}$$

on the other hand, if no initial distribution is given on x_1, define

$$\mathcal{S}(0) = e_{x_1}. \tag{11.11}$$

Notice that

$$\mathcal{S}(k, k+1)^{\downarrow k} = e_k. \tag{11.12}$$

Assume that the elements A_i, H_i, Q_i and R_i are all known for $k \in \{1, 2, \ldots\}$. Further, assume the measurements $\mathbf{y}_i$ are given up to time k, i.e.

$$\mathcal{O}(l) = e$$

for $l > k$. What can be inferred on some x_i from this information? There are three different cases:

- $i = k$: filtering about the current value,
- $i < k$: smoothing about past values, and
- $i > k$: prediction of future values.

All three cases are projection problems of the form

$$\mathcal{H}(k, i) = \left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{max(k,i)-1} \mathcal{S}(s, s+1)) \otimes (\bigotimes_{l=1}^{k} \mathcal{O}(l)) \right]^{\downarrow x_i}, \tag{11.13}$$

It can be verified for the filtering and smoothing cases $i \leq k$ that $\mathcal{S}(s, s+1)$ for $s \geq k$ give no information about x_i.

The valuation network (Shenoy, 1992) or factor graph (Kschischang et al., 2001) of the factorisation is shown in Figure 11.2: A valuation network is a bipartite graph with a box for each factor and a circle for each variable; edges are drawn from a factor to every variable occurring in its domain. The picture is simplified by drawing only one edge between the sets y_k and x_k and the corresponding factors. This is justified since these variables only occur together. Finally, a covering join tree and a possible assignment mapping is shown in Figure 11.3. All these graphical representations show a linear backbone, which is the transition model, and a series of equidistant pins attached to that backbone, which are the observations.

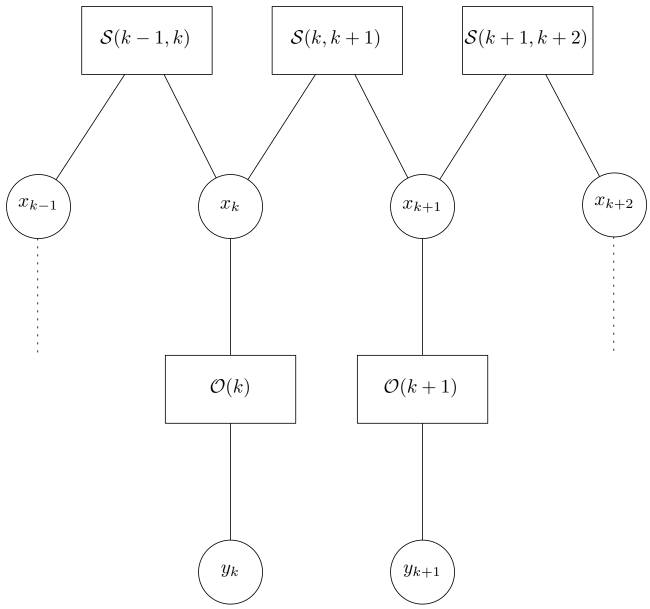


Figure 11.2: A valuation network for the Kalman filter model

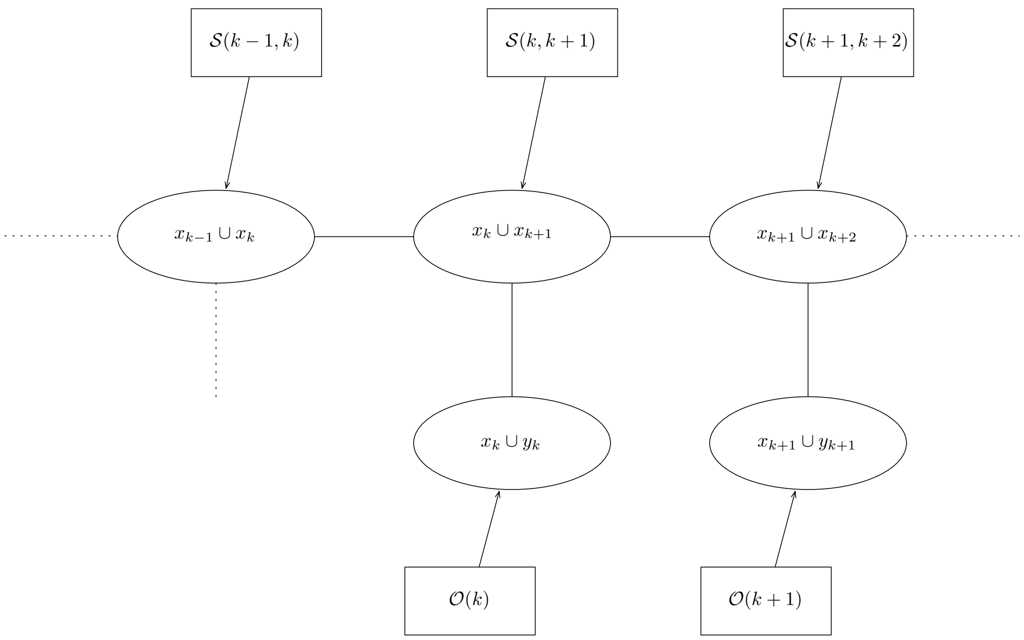


Figure 11.3: A join tree for the Kalman filter model

11.2 Filtering

The filter solution $\mathcal{H}(k,k)$ can be solved by an execution of the collect algorithm in the subtree rooted at the node containing $\mathcal{S}(k,k+1)$ and a final marginalisation to x_k. According to Theorem 4.18, after the collect algorithm in that subtree up, the root node contains

$$\mathcal{H}_k = \left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{h} \mathcal{O}(l))\right]^{\downarrow x_k \cup x_{k+1}} \tag{11.14}$$

$$= \left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{h} \mathcal{O}(l))\right]^{\downarrow x_k \cup x_{k+1}} \otimes \mathcal{S}(k,k+1)^{\downarrow x_k}$$

$$= \left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{h} \mathcal{O}(l))\right]^{\downarrow x_k \cup x_{k+1}} \tag{11.15}$$

using Lemma 2.4 and (11.12). Then, using the transitivity axiom, a final marginalisation yields ${H_k}^{\downarrow x_k} = \mathcal{H}(k,k)$. This scheme can also be described recursively by

$$\mathcal{H}(k,k) = \mathcal{H}(k-1,k) \otimes \mathcal{O}(k) \tag{11.16}$$

$$= (\mathcal{H}(k-1,k-1) \otimes \mathcal{S}(k-1,k) \otimes \mathcal{O}(k))^{\downarrow x_k}. \tag{11.17}$$

The messages of this execution of the collect algorithm are shown in Figure 11.4. The messages $\mathcal{H}(k-1,k)$ are called *one-step forward prediction*. They can be computed from the filter solution for $k-1$ and from the state model from $k-1$ to k,

$$\mathcal{H}(k-1,k) = \big(\mathcal{H}(k-1,k-1) \otimes \mathcal{S}(k-1,k)\big)^{\downarrow x_k}. \tag{11.18}$$

Let $\mathcal{H}(k-1,k-1) = (\nu(k-1,k-1), K(k-1,k-1))$. Then, the one-step prediction $\mathcal{H}(k-1,k) = (\nu(k-1,k), K(k-1,k))$ is given by

$$\nu(k-1,k) = A_{k-1}\nu(k-1,k-1) \tag{11.19}$$

and

$$K(k-1,k) = {Q_{k-1}}^{-1} - {Q_{k-1}}^{-1}A_{k-1}\big(A'_{k-1}{Q_{k-1}}^{-1}A_{k-1} + K(k-1,k-1)\big)^{-1} \cdot A'_{k-1}{Q_{k-1}}^{-1} \tag{11.20}$$

using equations (11.8) and (11.9). From the one-step prediction $\mathcal{H}(k-1,k) = (\nu(k-1,k), K(k-1,k))$, it is easy to compute the filter solution for k by

$$\mathcal{H}(k,k) = \big(\nu(k-1,k) + H'_k{R_k}^{-1}\mathbf{y}_k, H'_k{R_k}^{-1}H_k + K(k,k+1)\big). \tag{11.21}$$

If A is regular, as is usually assumed, then

$$\begin{aligned} K(k-1,k)^{-1} =& {Q_{k-1}}^{-1} - \\ & {Q_{k-1}}^{-1}\big({Q_{k-1}}^{-1} + {A_{k-1}}^{-1}K(k-1,k-1){A'_{k-1}}^{-1}\big)^{-1}{Q_{k-1}}^{-1} \\ =& Q_{k-1} + \big({A_{k-1}}^{-1}K(k-1,k-1){A'_{k-1}}^{-1}\big)^{-1} \\ =& Q_{k-1} + A'_{k-1}K(k-1,k-1)^{-1}A_{k-1} \end{aligned} \tag{11.22}$$

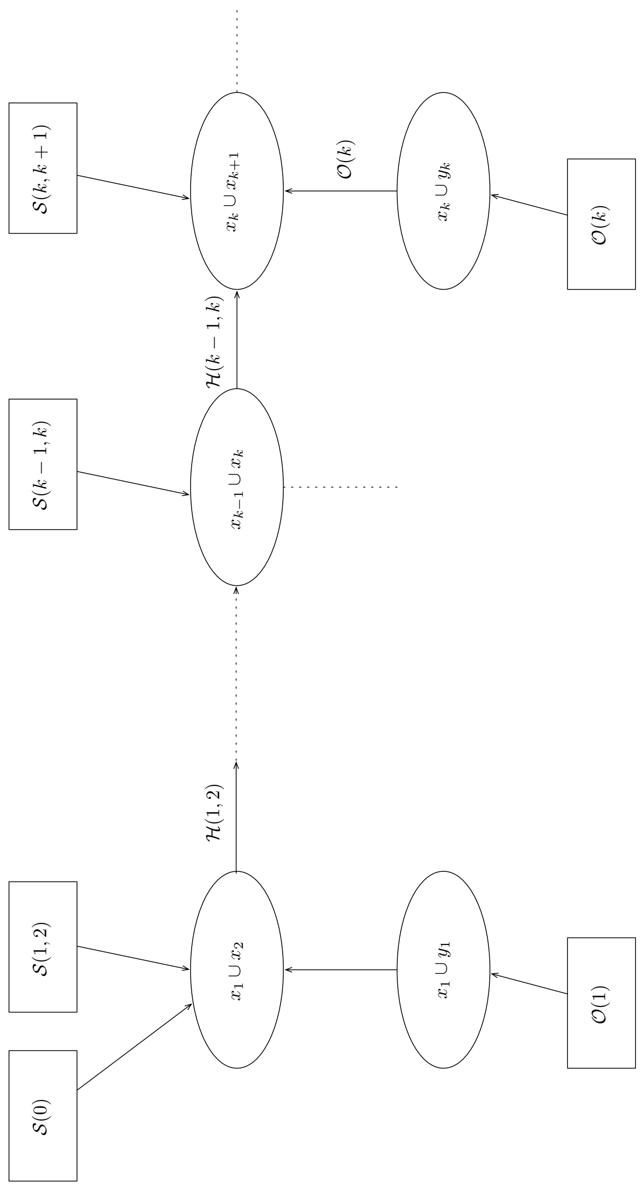


Figure 11.4: Messages for the filter problem

since for regular matrices A and B it holds that $A - A(A-B)^{-1}A = (A^{-1} - B^{-1})^{-1}$ in light of Theorem 18.2.4 of (Harville, 1997; p.420). This corresponds to the usual computation of the one-step prediction in terms of variance-covariance matrices.

The initial distribution on x_1 is technically not needed for the assumption-based inference; in contrast, it is needed when the Kalman filter is derived by using Bayes' rule or the least-squares principle. (Monney, 2003) points out that a precise filter solution may be found even without an initial distribution. Assume that all matrices A_k are regular. Then, if $\mathcal{H}(k,k)$ corresponds to a precise Gaussian hint, so do $\mathcal{H}(k,k+1)$, $\mathcal{H}(k+1,k+1)$, $\mathcal{H}(k+1,k+2)$,... This follows from equations (11.22) and (11.21), since the sum of a positive definite and a non-negative definite matrix is positive definite and since the inverse of symmetric positive definite matrix is positive definite. A sufficient condition for $\mathcal{H}(k,k)$ to be precise is that $\mathcal{H}_k$ is regular. Hence, the filter $\mathcal{H}(k,k)$ may become precise after a few steps without an initial distribution on x_1.

11.3 Prediction

The prediction problem $\mathcal{H}(k,k+t)$ can be solved by propagating further towards the node containing $x_{k+t} \cup x_{k+t+1}$. It holds that

$$\begin{aligned}
\mathcal{H}(k,k+t) &= \left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k+t-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k+t}} \\
&= \left(\left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k+t-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k} \mathcal{O}(l))\right]^{\downarrow \bigcup_{l=k+1}^{k+t} x_l}\right)^{\downarrow x_{k+t}} \\
&= \left(\left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k+1}} \otimes (\bigotimes_{s=k+1}^{k+t-1} \mathcal{S}(s,s+1))\right)^{\downarrow x_{k+t}} \\
&= \left(\mathcal{H}(k,k+1) \otimes (\bigotimes_{s=k+1}^{k+t-1} \mathcal{S}(s,s+1))\right)^{\downarrow x_{k+t}}
\end{aligned}$$

using the transitivity axiom and the combination axioms. Here, $\mathcal{H}(k,k+1) = \mathcal{H}_k{}^{\downarrow x_{k+1}}$ is the first message sent from the old root node containing $\mathcal{H}_k$ to the next node. This scheme is shown in Figure 11.5.

11.4 Smoothing

The smoothing problem $\mathcal{H}(k,k-t)$ can be solved by first using the collect algorithm to compute the filter solution at k and then distribute to propagating outwards, resp. backwards to $k-t$. Of course, the messages involved are exactly those from

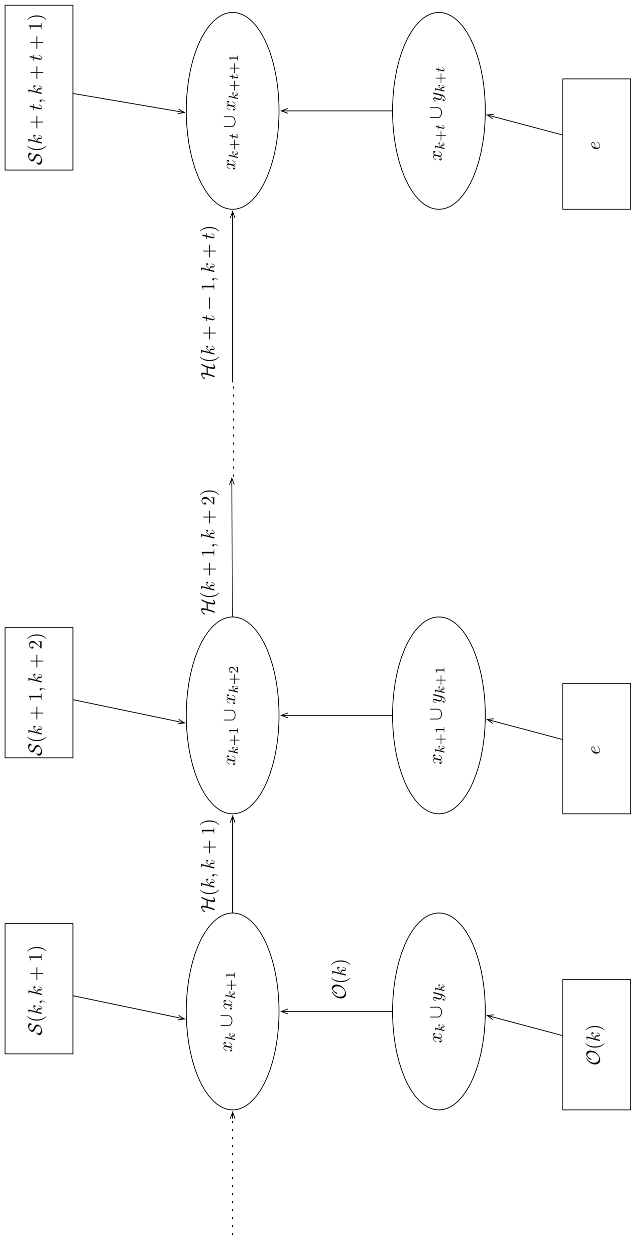


Figure 11.5: Messages for the prediction problem

the collect algorithm towards the root node containing $\mathcal{S}(k-t, k-t+1)$. Define the backward filter at time $k-t$ as

$$\mathcal{B}(k,k-t) = \left[(\bigotimes_{s=k-t}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=k-t}^{k} \mathcal{O}(l)) \right]^{\downarrow x_{k-t}}. \tag{11.23}$$

It contains the information on the time $k-t$ from the times $k-t$ to k. Then, it holds that

$$\begin{aligned}
&\mathcal{H}(k,k-t)\\
=&\left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k-t}}\\
=&\left(\left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k-t-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k-t-1} \mathcal{O}(l))\right] \otimes \left[(\bigotimes_{s=k-t}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=k-t}^{k} \mathcal{O}(l))\right]\right)^{\downarrow x_{k-t}}\\
=&\left(\left[\mathcal{S}(0) \otimes (\bigotimes_{s=1}^{k-t-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=1}^{k-t-1} \mathcal{O}(l))\right]^{\downarrow x_{k-t}} \otimes\right.\\
&\quad \left.\left[(\bigotimes_{s=k-t}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=k-t}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k-t}}\right)^{\downarrow x_{k-t}}\\
=&\left(\mathcal{H}(k-t-1,k-t) \otimes \left[(\bigotimes_{s=k-t}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=k-t}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k-t} \cup x_{k-t+1}}\right)^{\downarrow x_{k-t}}\\
=&\left(\mathcal{H}(k-t-1,k-t) \otimes \mathcal{S}(k-t,k-t+1) \otimes \mathcal{O}(k-t) \otimes\right.\\
&\quad \left.\left[(\bigotimes_{s=k-t+1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=k-t+1}^{k} \mathcal{O}(l))\right]^{\downarrow x_{k-t+1}}\right)^{\downarrow x_{k-t}}\\
=&\big(\mathcal{H}(k-t-1,k-t) \otimes \big(\mathcal{S}(k-t,k-t+1) \otimes \mathcal{O}(k-t)\big) \otimes \mathcal{B}(k,k-t+1)\big)^{\downarrow x_{k-t}}\\
=&\big(\mathcal{H}_{k-t} \otimes \mathcal{O}(k-t) \otimes \mathcal{B}(k,k-t+1)\big)^{\downarrow x_{k-t}}
\end{aligned}$$

using the combination and the transitivity axioms. Here, $\mathcal{H}(k-t-1, k-t)$ is the message from the earlier times to $k-t$, $\mathcal{S}(k-t, k-t+1) \otimes \mathcal{O}(k-t)$ is the information on state $t-k$ and $\mathcal{B}(k, k-t+1)$ is the information on $k-t+1$ from the times $k-t+1$ to k. The backward filters can be computed recursively as

$$\begin{aligned}
\mathcal{B}(k,i) &= \left[(\bigotimes_{s=i+1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=i+1}^{k} \mathcal{O}(l))\right]^{\downarrow x_i}\\
&= \left(\left[(\bigotimes_{s=i+1}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=i+1}^{k} \mathcal{O}(l))\right]^{\downarrow x_i \cup x_{i+1}}\right)^{\downarrow x_i}
\end{aligned}$$

$$= \left(\mathcal{S}(i+1,i+2) \otimes \mathcal{O}(i+1) \otimes \left[(\bigotimes_{s=i+2}^{k-1} \mathcal{S}(s,s+1)) \otimes (\bigotimes_{l=i+2}^{k} \mathcal{O}(l)) \right]^{\downarrow x_{i+1}} \right)^{\downarrow x_i}$$
$$= \big(\mathcal{S}(i+1,i+2) \otimes \mathcal{O}(i+1) \otimes \mathcal{B}(k,i+1)\big)^{\downarrow x_i}$$

using the transitivity and the combination axioms. Notice that

$$\mathcal{B}(k,k) = \mathcal{O}(k) = \mathcal{O}(k) \otimes e_k = \mathcal{O}(k) \otimes \mathcal{S}(k,k+1)^{\downarrow x_k} = \big(\mathcal{O}(k) \otimes \mathcal{S}(k,k+1)\big)^{\downarrow x_k}$$

using equation (11.12) and the combination axiom. Therefore, the backward filters correspond to the messages of the backward distribute phase after collect towards the root node containing $\mathcal{S}(k,k+1)$ or, equivalently, to the messages of the collect algorithm towards the node containing $\mathcal{S}(k-t,k-t+1)$ as shown in Figure 11.6.

Chapter Synopsis & Discussion

It has been shown that the Kalman filter, prediction and smoothing algorithms fit nicely into the framework of assumption-based reasoning and local computation: At the end of the Shenoy-Shafer algorithm, the corresponding nodes contain the solution of the filtering, the prediction, and the smoothing problem, respectively. This sheds new light on the original algorithms, which were intended for optimal estimation with respect to the expected quadratic loss (Kalman, 1960).

The assumptions of an initial distribution and on the ranks of the state transition and observation matrices can be dropped for the assumption-based inference. The filter solution may still be precise. Furthermore, the computation of the filter solution from the one-step predictions is very easy in terms of symmetric Gaussian potentials, whereas the one-step forward predictions are more complicated. In contrast, if Gaussian hints are represented by the variance-covariance matrix instead of the concentration matrix, it is more expensive to compute the filter solution from the one-step forward predictions than to compute the one-step forward predictions. See (Monney, 2003; Kohlas and Monney, 2008).

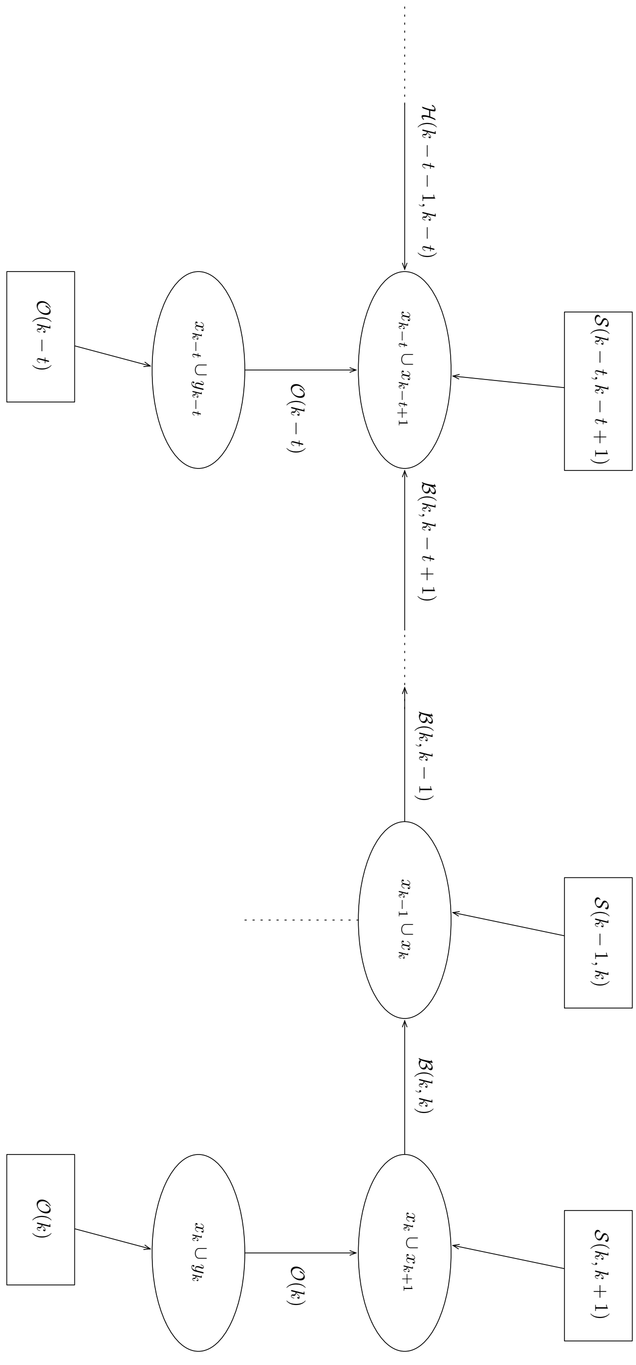


FIGURE 11.6: Messages for the smoothing problem

12

Implementation

Introduction and Chapter Outline

In this chapter, a prototypical architecture for analysing Gaussian linear systems is presented as an application and illustration of the theoretical concepts presented so far, in particular of symmetric Gaussian potentials. An assumption-based analysis of Gaussian linear systems may proceed in the following steps:

1. Gaussian linear systems and queries are expressed in a human-readable way;
2. a knowledge base of symmetric Gaussian potentials is generated;
3. the queries are answered by local computations;
4. the results are presented in a human-readable form.

In this chapter, such an architecture is presented: Models and queries formulated in ABEL (assumption-based evidential language) (Anrig et al., 1997; Haenni et al., 1998; Lehmann, 2005) are passed on to the *Gauss solver*, which uses the NENOK framework for local computation (Pouly, 2004; 2006; 2008) to solve the projection problems. This architecture is shown in Figure 12.1. Here, arrows represent actions, and boxes stand for the intermediate representations of the processed input; the dashed boxes group the actions performed by the same software unit. ABEL and the Gauss solver communicate via file exchange, whereas the Gauss solver uses the NENOK framework by passing objects in JAVA method calls. This implementation is going to be discussed as follows: How Gaussian linear models can be formulated in ABEL and how this input is passed on to the Gauss solver is discussed in Section 12.1. How the Gauss solver parses generates the knowledge base and the queries in the NENOK framework, is the subject of Section 12.2.

12.1 Model Formulation in Abel

The ABEL language (Anrig et al., 1997; Haenni et al., 1998; Lehmann, 2005) is an extension of the Common Lisp language[1] by several commands which allow to

[1] ANSI standard INCITS 226-1994 (R2004) *Information Technology – Programming Language – Common Lisp*; see also (Steele, 1990).

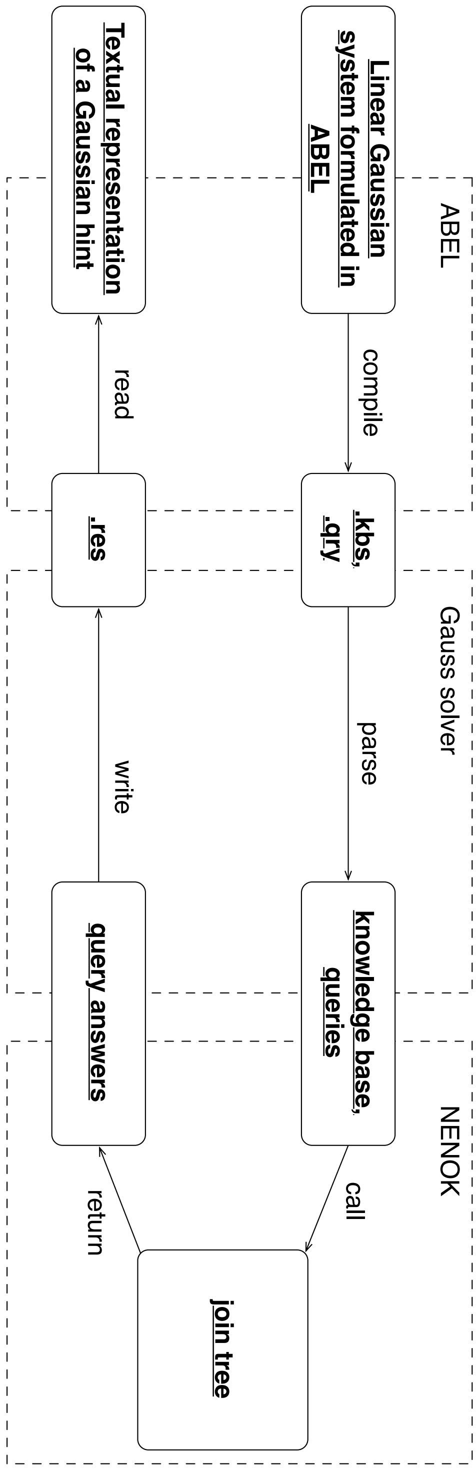


FIGURE 12.1: Overview of the architecture

manage a knowledge base. Abel can be integrated in XEmacs[2] and run in an interactive way from there.

In the following paragraphs, the Abel-specific commands are going to be discussed. The formal syntax for Abel is written in a simple extended BNF[3]:

- Although not mandatory, non-terminal symbols are marked by angle brackets (<some-non-terminal-symbol>);
- "string" and 'string' match the literal string given inside double and single quotes, respectively;
- (expression) matches expression, which is treated as a unit and may be combined as described below;
- A? matches the expression A or nothing (i.e. A is optional);
- A B matches expressions A followed by B (*concatentation)*;
- A | B matches A or B (*alternation*);
- A+ matches one or more occurrences of A;
- A* matches zero or more occurrences of the expression A.

Here, + and * have higher precedence than concatenation and concatenation higher precedence than alternation, i.e.

- A B | C D is identical to (A B) | (C D), and
- A+ | B+ and A* | B* are identical to (A+) | (B+) and (A*) | (B*), respectively.

The formal syntax declarations will abstract from the *lexical structure* of programs, i.e. how they are *tokenised*. Abel's lexical structure is inherited from the underlying Common Lisp language.[4]

Knowledge Assertion: the tell and observe Commands

First, the knowledge base is defined using the commands tell and observe. They add variable and assumption declarations, relations and observations to the (global) knowledge base. They allow to distinguish between static knowledge of the model (such as rules, relations or dependencies) and facts or observations in an actual situation, which may change over the time:

[2] http://www.xemacs.org/, accessed 2009/2/27

[3] The version used is adapted from http://www.w3.org/TR/REC-xml/#sec-notation, accessed 2009/2/27.

[4] There is an exception: Indexed variables (for instance var[25]) are not Common Lisp identifiers and are only valid in the body of Abel assertions (which are Lisp macros and not ordinary functions).

The Syntax

```
<assertion> ::=
    <tell> | <observe>;
<tell> ::=
    "("  "tell"
     ( "(" <key> ")" )?
     <declaration or statement>+
    ")" ;
<observe> ::=
    "("  "observe"
     ( "(" <key> ")" )?
     <declaration or statement>+
   ")";
<key> ::=
    <identifier>
```

The <key> is an optional Lisp identifier and can be used for partial retraction of knowledge (see the empty command). Notice that tell and observe differ only on the syntactical level; internally, they are treated completely identically. Therefore, the Gauss solver will not be able to distinguish between knowledge asserted using the tell and the observe commands (see also Section 12.1 below for details).

The knowledge assertions are either declarations or statements:

The Syntax

```
<declaration or statement> ::=  <declaration> | <statement>
```

Declarations

In the case of Gaussian linear models, (arrays of) <variable>s, (arrays of) <assumption variable>s, <indexing set>s, and covariances can be declared in a <declaration>:

The Syntax

```
<declaration> ::=
    <variable declaration> | <assumption declaration> |
    <covariance declaration> |Ê<indexing set declaration>;
<variable declaration>  ::=
    "(" "var"  <identifier> ("[" <indexing set>+ "]")? "real" ")";
<assumption declaration> ::=
    "("  "ass"  <identifier> ("[" <indexing set>+ "]")?  "real"
        "("  "gauss" <number>+  ")"
    ")";
<indexing set declaration> ::=
    "("  "type" <indexing set>
      "(" "integer"  "1"  <positive integer>")"
    ")"
<covariance declaration> ::=
    "("  "cov"  <assumption variable>  <assumption variable>
                <covariance> ")"
```

```
<variable> ::=
    <identifier>("[" ( <integer> | <integer-bound variable> )+ < "]")?
<assumption variable> ::=
    <identifier>("[" ( <integer> | <integer-bound variable> )+ "]")?
<indexing set> ::=
    <identifier>
<array of variables> ::=
    <identifier>
<array of assumption variables> ::=
    <identifier>
```

In an assumption declaration, the <number>s are the variance of the declared assumption variables:

- Either there is only one <number>, in which case all the indexed assumption variables have the same variance <number>, or
- there are exactly as many <number>s as indexed assumption variables, in which case the assumption variable <identifier>[<i>]s has the ith variance from the <number>s.

Example **12.1**. As an example, the following tell assertion declares a real-valued variable V and Gaussian assumptions O1 with variance 100 and O2 with variance 200, both with mean 0; the Abel macros added to basic Lisp as well as some keywords are boldfaced:

```
(tell
  (var V real)
  (ass O1 O2 real (gauss 100 200))
)
```

Alternatively,

```
(tell
  (type some-indexing-set (integer 1 3))
)
```

defines the indexing set some-indexing-set with integer values from 1 to 3. Indexed variables and assumption variables can be defined in the following way:

```
(tell
 (var some-indexed-var[some-indexing-set] real)
 (ass some-indexed-assumption[some-indexing-set]
     real (gauss 50))
 (ass some-other-indexed-assumption[some-indexing set]
     real (gauss 1 2 3))
)
```

This declares the array of variables some-indexed-var containing the variables some-indexed-var[1] to some-indexed-var[3]. For Gaussian assumptions, two forms are possible: The assumptions some-indexed-assumption[1] to some-indexed-

`assumption[3]` all have the same variance `50`, whereas in the second form the assumptions `some-other-indexed-assumption[1]` to `some-other-indexed-assumption[3]` have the individual variances $1, 2, 3$, respectively. Doubly indexed variables and assumptions can be declared as follows:

```
(tell
  (type another-indexing-set (integer 1 5))
  (var doublette[some-indexing-set another-indexing-set]
      real)
)
```

This defines the $3 \cdot 5 = 15$ variables `doublette[1 1]`, `doublette[1 2]`, ..., `doublette[3 5]` in the doubly indexed array `doublette`. ⊘

Statements

Statements containing Gaussian linear equations have the following form:

The Syntax

```
<statement> ::=
     <equation> | <forall>;
<forall> ::=
     "(" "forall" <identifier>  <indexing set> ")"
     <statement>+;
<equation> ::=
    "(" "=" <sum> <sum> ")";
<sum> ::=
    ("(" <identifier> <indexing set> ")")?
        "(" ( "+" | "-" ) <sum>+ ")" |
    <term>;
<term> ::=
    <real number> | <observed variable> | <product> |
    <assumption variable>;
<product> ::=
    "(" "*" <observed variable>  <variable> ")" |
    "(" "*" <variable>  <observed variable> ")"
```

Of course, at most one of the terms in every equation must be an assumption variable, and the order of the terms is irrelevant semantically. An `<observed variable>` is one having a `<direct observation>`, i.e. an `<equation>` of the form

The Syntax

```
<direct observation> ::=
    "(" "=" <variable> <number> ")" |
    "(" "=" <number> <variable> ")"
```

Notice that directly observed variables are treated as constants in an equation with a Gaussian term and do not appear in that equation's domain.

EXAMPLE **12.2**. *Indexed* statements can be formulated using `forall`. For instance, the price for all vendors can be fixed at 20 for all 5 vendors:

```
(tell
  (type vendors (integer 1 5))
  (var vendor[vendors] real)
  (forall (vendor vendors)
          (= price[vendor] 20)))
```

Here, 5 statements are generated, one for each value $1, \ldots, 5$, and the occurrences of the local variable `vendor` are replaced by that value. This is of course equivalent to

```
(tell
  (type vendors (integer 1 5))
  (var vendor[vendors] real)
  (= price[1] 20)
  (= price[2] 20)
  (= price[3] 20)
  (= price[4] 20)
  (= price[5] 20)
)
```

Similarly, indexed sums can be expressed using `sum`. For instance,

```
(sum (factor factors)
   impact[factor]))
```

sums up the impacts over all `factor`s in the indexing set `factors`. ⊘

Querying: the `ask` Command

Then, queries about the knowledge base can be formulated using the command `ask`.

The Syntax

```
<ask> ::=
     "(" "ask" <query>+ ")";
<query> ::=
    <variable> | "("  <variable>+ | <array of variables> ")";
```

A query on a Gaussian linear model is either a single variable or a list of variables. For instance,

The Syntax

```
(ask price)
```

solves the projection problem for the single variable `price`. However, several queries can be asked in the same ask statement.

Knowledge Retraction: the `empty` Command

Finally, the knowledge base can be deleted (partially or completely) using `empty`: Either it is emptied completely, or only parts can be retracted from the knowledge base:

The Syntax

```
(empty)
(empty observe)
(empty tell)
(empty <key>)
```

The first form clears the whole knowledge base, the second one all observations, the third everything added using the `tell` command, and the fourth the statements corresponding to the key `<key>`.

EXAMPLE **12.3**. The Gaussian linear system

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + o_i, \qquad i \in \{1,2\}$$ ⊘

where o_1 and o_2 are Gaussian with mean 0 and variances 50 and covariance 30 can be expressed in ABEL in the following way:

```
(empty)
(tell
  (type I-set (integer 1 2))
  (type J-set (integer 1 2))
  (var a[I-set J-set] REAL)
  (var X[J-set] REAL)
  (var Z[I-set] REAL)
  (ass o[I-set] REAL (GAUSS 50))
  (cov o[1] o[2] 30)
)

(tell
  (forall (i I-set)
          (= Z[i]
             (+ (sum (j J-set)
                     (* a[i j] X[j]))
                o[i]))))

(observe
  (= 5 a[1 1])
  (= 6 a[1 2])
  (= 7 Z[1])
  (= 8 a[2 1])
  (= 9 a[2 2])
  (= 10 Z[2])
)
```

The .kbs and .qry Files

When an `ask` statement is evaluated, ABEL's knowledge base is written in a `.kbs` (knowledge base) file, which is an ordinary ASCII text file with the following structure:

- the `ID` section just contains the marker `KBS`;
- the `VARIABLES` section declares all variables and
- the `ASSUMPTIONS` section all assumptions.
- The covariance of two Gaussian assumptions is given in the `COVARIANCES` section.
- Then, the equations and the observed variables are given in the `CLAUSES` section.
- Finally, the end of the knowledge base is marked by the `END` tag.

Example **12.4**. The model of Example 12.3 produces the following `.kbs` file:

```
{ID} KBS

{VARIABLES}
A[1 1] REAL
A[1 2] REAL
A[2 1] REAL
A[2 2] REAL
X[1] REAL
X[2] REAL
Z[1] REAL
Z[2] REAL

{ASSUMPTIONS}
O[1] REAL (GAUSS 50)
O[2] REAL (GAUSS 50)

{COVARIANCES}
O[1] O[2] 30

{CLAUSES}
| (= (- Z[1] (+ (* A[1 1] X[1]) (* A[1 2] X[2]) O[1])) 0)
| (= (- Z[2] (+ (* A[2 1] X[1]) (* A[2 2] X[2]) O[2])) 0)
| (= A[1 1] 5)
| (= A[1 2] 6)
| (= Z[1] 7)
| (= A[2 1] 8)
| (= A[2 2] 9)
| (= Z[2] 10)

{END}
```

The reason for the name `CLAUSES` is that Abel was originally designed for assumption-based reasoning on logical models. However, Abel can also be used in a natural way to cover Gaussian linear models. ⊘

The queries are encoded in a separate `.qry` file, which contains the domains.

EXAMPLE **12.5**. When asking

```
      (ask X)
```

on the model of Example 12.3, the following `.qry` file is passed on to the Gauss solver:

```
{ID} QRY

{QUERY} MARGINAL
{CLAUSES}
(X[1] X[2])

{END}
```

With Gaussian linear models, only `MARGINAL` queries can be asked. When dealing with logical models, ABEL can also be used for other queries evaluating hypotheses (quantitatively and qualitatively). ⊘

12.2 Implementing an Algebra of Gaussian Linear Information in Nenok

When the Gauss solver gets the `.kbs` and `.qry` files from ABEL, it has to extract the knowledge base and answer the queries. For this purpose, the Gauss solver extends the NENOK framework (Pouly, 2004; 2006; 2008) for local computation and uses the Jama library[5] for matrix computations. These dependencies are shown in the UML package diagram of Figure 12.2. Further, Figure 12.3 shows how a

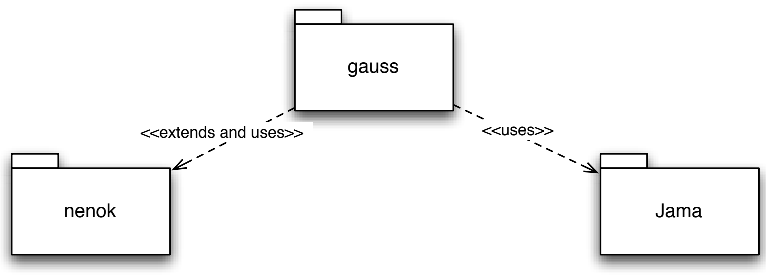


FIGURE 12.2: The Gauss solver package `gauss` depends on the packages `Jama` and `nenok`.

valuation algebra of Gaussian linear information can be integrated in the NENOK framework: The class `GLI` implements NENOK's `Valuation` interface, and the class `GaussVariable` implements the `Variable` interface. Thereby, NENOK will recognise the generated symmetric Gaussian potentials as valuations with the operations of a valuation algebra.

[5] version 1.0.2, `http://math.nist.gov/javanumerics/jama/`, accessed 2009/2/27

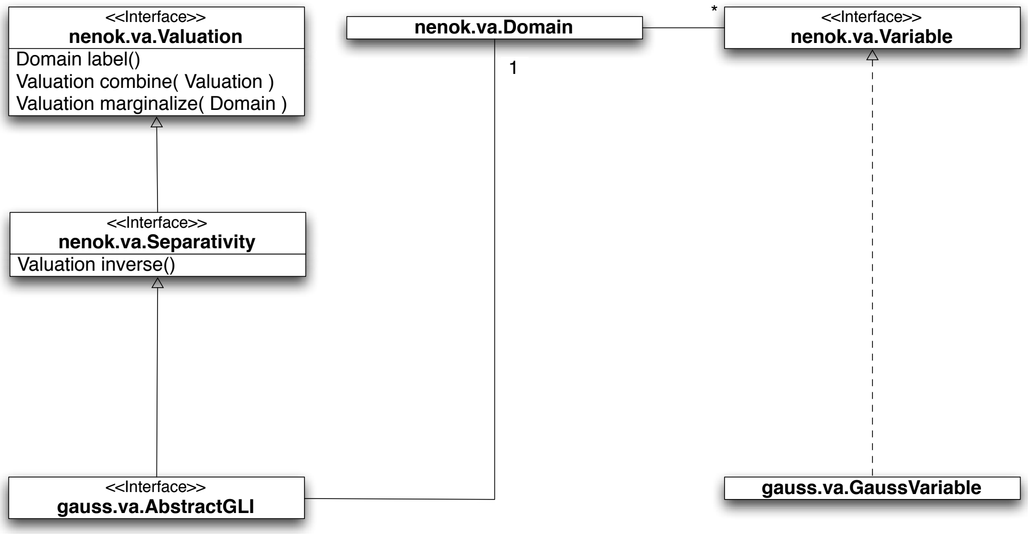


Figure 12.3: Integrating symmetric Gaussian potentials in the Nenok framework (simplified UML class diagram)

Applying Deterministic Equations to a Symmetric Gaussian Potential

The method `applyDeterministicEquations` is called in the constructor of the class `GLI`, which implements the interface `AbstractGLI`. In Listing 12.1, the symmetric Gaussian potential is stored in the fields `nu` and `Lambda`.

Combination

The combination is implemented by component-wise addition of the mean vector and the pseudo-concentration matrix, as shown in Listing 12.2.

Marginalisation

Listings 12.3 and 12.4 show only extracts of the `marginalize` method of the class `GLI`. The buffers `nu_buf` and `C_buf` contain the mean and the pseudo-covariance matrix such that lines and columns corresponding to the domain of marginalisation are top and left, respectively. Then, the symmetric Gaussian potential has first to be "rotated" and the variables are then eliminated one by one from bottom to top and from right to left in the outmost loop until the lower bound `mMarg` (i.e. the cardinality of the domain of marginalisation) is reached. For every variable index `l`, the variable elimination requires a loop over the lines (cf. running variable `i`) and a nested loop over the columns (cf. running variable `j`); since the pseudo-concentration matrix remains symmetric at every step, only the diagonal and the upper part are used (condition `j<i`).

After that, the top left parts of the buffers are used to instantiate the marginal valuation.

```java
private void applyDeterministicEquations(){
  int k = c.getRowDimension();
  int n = varArr.length;

  if(k>0 && n>0){
    // apply
    Matrix U,V; double[] s; int r;
    if(k>=n){ // svd algorithm works only for k>=n
      SingularValueDecomposition svd = C.svd();
      r = svd.rank(); U = svd.getV(); V = svd.getU();
      s = svd.getSingularValues();
    } else{
      SingularValueDecomposition svd = C.transpose().svd();
      r = svd.rank(); U = svd.getV(); V = svd.getU();
      s = svd.getSingularValues();
    }

    // check whether for contradiction: U_1 U_1' c = c?
    if(r==0){
      if(c.get(0, 0)!=0) { // contradiction
        throw new IllegalArgumentException("contradiction␣(rank␣0)");
      }
    } else {
      Matrix U_1 = U.getMatrix(0, k-1, 0, r-1);
      Matrix U_1prime = U_1.transpose();
      Matrix U_1U_1primec = U_1.times(U_1prime.times(c));
      for(int i=0; i<k; i++){
        if(!DhbMath.equals(U_1U_1primec.get(i, 0),c.get(i, 0))){
          throw new IllegalArgumentException("contradiction␣at␣line␣" + i);
        }
      }
    }
    // compute V2 V2' = I_n - V_1 V_1' (V2 is not computed in the svd)
    Matrix V_1 = V.getMatrix(0, n-1, 0, r-1); // V is n x n, V_1 is n x r
    Matrix V_1TimesV1prime = V_1.times(V_1.transpose());
    double[][] buf = new double[n][n];
    for(int i=0; i<n; i++){
      for(int j=0;j<n;j++){
        buf[i][j] = - V_1TimesV1prime.get(i, j);
      }
      buf[i][i] += 1;
    }
    Matrix V_2TimesV_2prime = new Matrix(buf);

    // compute C+
    Matrix Cplus = svdUtil.pseudoinverse(U,s,V,r);
    Cplus.print(Logger.getOutputStream(), 5, 5);
    C.times(Cplus.times(C)).print(Logger.getOutputStream(), 5, 5);

    // LambdaTilde = (V_2 V_2') Lambda (V_2 V_2')
    // nuTilde = (V_2 V_2') nu - (V_2 V_2') Lambda C+ c
    Matrix temp = V_2TimesV_2prime.times(Lambda);
    Matrix LambdaTilde = temp.times(V_2TimesV_2prime);
    Matrix nu_1 =
      V_2TimesV_2prime.times(nu);
    Matrix nu_2 = temp.times(Cplus.times(c));
    Matrix nuTilde = nu_1.minus(nu_2);

    nu = nuTilde;
    Lambda = LambdaTilde;
  }
}
```

LISTING 12.1: Applying deterministic knowledge to a symmetric Gaussian potentials

```
public static SymmetricGaussianPotential
  combine(SymmetricGaussianPotential val1,
  SymmetricGaussianPotential val2){
  /*
   * (nu1,C1) x (nu2,C2) = (nu1+nu2,C1+C2)
   */

  // prepare domain and empty vector and matrix...
  Domain newDomain = Domain.union(val1.label(), val2.label());
  int m = newDomain.size();

  List<GaussVariable> newVarsList = new ArrayList<GaussVariable>();
  for(Variable v : newDomain){
    newVarsList.add((GaussVariable) v);
  }
  GaussVariable[]  newVarsArr = new GaussVariable[m];
  newVarsList.toArray(newVarsArr);

  Matrix newK = new Matrix(m,m);
  Matrix newmu = new Matrix(m,1);

  // ... and sum up everything into them
  for(int i=0; i<m; i++){
    GaussVariable v_i = newVarsArr[i];
    newmu.set(i, 0, val1.getMu(v_i) + val2.getMu(v_i));
    for(int j=0; j<m; j++){
      GaussVariable v_j = newVarsArr[j];
      newK.set(i, j, val1.getK(v_i, v_j) + val2.getK(v_i, v_j));
    }
  }
  return new SymmetricGaussianPotential(
    newVarsList.toArray(new GaussVariable[newVarsList.size()]),
    newK, newmu );
}
public Valuation combine(Valuation val) {
  return combine(this, (SymmetricGaussianPotential) val);
}
```

Listing 12.2: The combination of GLIs

```
//rotate
Matrix margC = new Matrix(0,mMarg);
Matrix margc = new Matrix(0,0);
if(k>0 && m>0){
  Matrix U,V; double[] s; int r;
  if(k>=m-mMarg){ // svd algorithm works only for k>=m-mMarg
    Matrix U,V; double[] s; int r;
    try{
      /* in order to get U2.
       * However, svd algorithm may not work if k>=m-mMarg.
       */
      SingularValueDecomposition svd = C2.transpose().svd();
      r = svd.rank();
      U = svd.getV();
      V = svd.getU();
      s = svd.getSingularValues();
    } catch(Exception e){
      try{
        SingularValueDecomposition svd = C2.svd();
        r = svd.rank();
        U = svd.getU();
        V = svd.getV();
        s = svd.getSingularValues();
      } catch(Exception e2){
        throw new RuntimeException("Cannot compute the svd. I give up.");
    }
  }

  // rotate if necessary
  if(r>0){
    Matrix C2plus = svdUtil.pseudoinverse(U,s,V,r);
    Matrix C2plusC = C2plus.times(C_buf);
    Matrix C2plusc = C2plus.times(c);

    Matrix Lambda22 = Lambda_buf.getMatrix(mMarg,m-1, mMarg,m-1);
    Matrix Lambda2 = Lambda_buf.getMatrix(0,m-1, mMarg,m-1);
    Matrix nu2 = nu_buf.getMatrix(mMarg,m-1, 0,0);

    Matrix CprimeC2plusprime = C2plusC.transpose();
    Matrix CprimeC2plusprimeLambda22 = CprimeC2plusprime.times(Lambda22);
    Matrix Lambda2CplusC = Lambda2.times(C2plusC);


    // nu -> nu - C' C_2+' nu_2 - Lambda_2 C_2+ c + C' C_2+ ' Lambda22 C_2+ c
    nu_buf.minusEquals(CprimeC2plusprime.times(nu2));
    nu_buf.minusEquals(Lambda2.times(C2plusc));
    nu_buf.plusEquals(CprimeC2plusprimeLambda22.times(C2plusc));

    // Lambda -> Lambda - C' C_2+' Lambda_2 - Lambda2' C_2+ C -
    //                           C' C_2+' Lambda_22 C_2+ C
    Lambda_buf.minusEquals(Lambda2CplusC.transpose());
    Lambda_buf.minusEquals(Lambda2CplusC);
    Lambda_buf.plusEquals(CprimeC2plusprimeLambda22.times(C2plusC));
  }

  // setting up the deterministic system of the marginal
  if(k-r>0){
    Matrix U_2 = U.getMatrix(0, k-1, r, k-1);
    Matrix U_2prime = U_2.transpose();
    margC = U_2prime.times(C1);
    margc = U_2prime.times(c);
  }
}
```

LISTING 12.3: "Rotation" of a symmetric Gaussian potential in preparation of marginalisation

```
// eliminate the variables in the probabilistic part
int cuts = 0;
// eliminate variable l
/*
 * (nu,C)^{\downarray s} = (nu[i] - C[i,l] C[l]^{-1} nu[l], C[i,j] -
 *            C[i,l] C[l]^{-1} C[l,j])
 */
for(int l=m-1; l>= mMarg; l--){
  if(DhbMath.equals(Lambda_buf.get(l, l), 0)){
    for(int i=0;i<l;i++){
      if(!DhbMath.equals(Lambda_buf.get(l, i),0) ||
         !DhbMath.equals(nu_buf.get(l, 0),0)){
        throw new VAException("Cannot␣marginalise␣non-vacuous␣variable␣with␣"
          + "concentration␣0␣at␣var␣"+l + "␣while␣marginalisation␣to␣"
          + margDom + "\n"  + this.toString());
      }
    }
  } else {
    double c_ll_inv = 1 / Lambda_buf.get(l,l);
    double nu_l = nu_buf.get(l, 0);
    for(int i=0;i<l;i++){
      double c_il = Lambda_buf.get(l, i);
      nu_buf.set(i, 0, nu_buf.get(i, 0) - c_il * nu_l * c_ll_inv);
      Lambda_buf.set(i, i, Lambda_buf.get(i, i)- c_il * c_il * c_ll_inv);
      for(int j=0; j<i; j++){
        Lambda_buf.set(i, j, Lambda_buf.get(i, j) -
          c_il * Lambda_buf.get(l, j) * c_ll_inv);
      }
    }
  }
}
```

Listing 12.4: Marginalisation of `GLI`s by variable elimination

12.3 Implementing the Gauss Solver

In order to answer the queries, the Gauss solver proceeds in 8 steps:

1. parse the `.kbs` file;
2. build the Gaussian linear systems;
3. build the symmetric Gaussian potentials from them;
4. read the `.qry` file;
5. generate the join tree;
6. put the deterministic knowledge on the appropriate nodes;
7. propagate and answer queries;
8. generate the `.res` file.

In the parsing steps 1 and 4, the files are read line by line in order to extract the necessary information and details are omitted here (see the source code for details[6]). Some further information is now given on the other steps.

Step 2: Building the Gaussian Linear systems

In order to answer the queries using local computations, a knowledge base with a set of labelled valuations is required. However, in the `.kbs` file, the whole model is "flat" in the sense that the `.kbs` file does not specify which equations can and should be grouped together in the same system. The equations which *must* be grouped together *without transforming* the system are those whose assumptions are correlated. Algorithm 3 groups together these equations in the following way: At the beginning of the algorithm, every assumption (variable) is its one and only neighbour. For every covariance entry of two assumptions, their neighbourhoods are joined, and that new neighbourhood is assigned to all members. Thus, by iterating over the covariance entries, the neighbourhoods are transitively closed in each step.

Step 3: Building Symmetric Gaussian Potentials and Neutral Elements for the Deterministic Equations

From these Gaussian linear systems, symmetric Gaussian potentials without deterministic equations are generated.

Step 5: Generating the Join Tree

Once the Symmetric Gaussian Potentials are built from the corresponding equations, a covering join tree is generated by a NENOK heuristics, and the queries are answered in NENOK, as shown in Listing 12.6.

[6]The source code is available under `http://diuf.unifr.ch/tcs/christian.eichenberger/GLI`.

```
/*
 * 3. build the potentials (neutral for deterministic)
 */

AbstractGLI[] potentials = new GLI[GLSs.length+detEqs.length];
{
  int i=0;
  for (GaussianLinearSystem g : GLSs){
    potentials[i] = g.convertToGLI();
    Logger.log(Logger.DEBUG, "␣--␣Built␣symmetric␣GP:␣");
    if(Logger.logDEBUG()){
      potentials[i].print(Logger.getOutputStream());
    }
    i++;
  }
  for(int k=0;k<detEqs.length;k++,i++){
    Domain dom = detEqs[k].getDomain();
    potentials[i] = new NeutralGLI(dom);
  }
}
```

Listing 12.5: Building symmetric Gaussian potentials and neutral elements for the deterministic equations

```
/*
 * 5. generate the join tree based on the domains
 */

Knowledgebase<?> kb = Knowledgebase.create(potentials, "the␣knowledgebase");

LCFactory factory = new LCFactory(Architecture.Lauritzen_Spiegelhalter);
JoinTree jt = factory.create(kb, queries);
```

Listing 12.6: Join tree generation using Nenok in the main class `gauss.GaussSolver`

Algorithm 3: *GroupEquations*

→ **input:** A set<assumption> *assumptions*,
a set<unordered-pair<assumption,assumption>> *covariances*

← **output:** a class table<assumption,set<assumption>>

neighbours := **new** table<assumption,set<assumption>>

loop for *ass* **in** *assumptions*
do
    neighbours.put(*ass*,{*ass*})
done

loop for $(ass1, ass2)$ **in** *covariances*
do

1. $fillIn$:= $neighbours$.get($ass1$) $\cup$ $neighbours$.get($ass2$)
2. **for** *ass* **in** $fillIn$
    $neighbours$.put(ass,$fillIn$)
   done

done
return neighbours

Step 6: Applying the Deterministic Knowledge

Every deterministic equation is added to a covering node: First, a covering node is looked up for every deterministic equation, and the equation is added to a linear system associated with that node. Finally, the resulting linear system is converted to a deterministic `GLI` and combined into the corresponding node content. In this way, the deterministic linear equations do not have to be applied individually, which may reduce the number of required singular-value decompositions.

Step 7: Propagation

Now, the generic local computation algorithms of the NENOK framework can be used to answer the queries as shown in Listing 12.7.

Step 8: The `.res` File

Finally, a textual representation of the query answers are printed to a `.res` file. The `.res` file is then read by ABEL and printed on the screen as is. Some examples are shown in Chapter 13.

```java
/*
 * 7. propagate and answer queries
 */

Logger.log(Logger.INFO,
"\n\n=== 7. Answer the Queries using Local Computation ==========");

jt.propagate();
AbstractGLI answers[] = new AbstractGLI[queries.size()];
Domain[] queryArr = new Domain[queries.size()];
queries.toArray(queryArr);
for(int i=0; i<answers.length; i++){
  Logger.logINFO("Answering query " + queryArr[i]);
  answers[i] = (AbstractGLI) jt.answer(queryArr[i]);
}
```

LISTING 12.7: Propagation using NENOK in the main class `gauss.GaussSolver`

Chapter Synopsis & Discussion

The choice of NENOK reflects the fact that Gaussian linear models fit into the generic algebraic and algorithmic framework of valuation algebras. The realisation of symmetric Gaussian potentials adds a further item to the catalogue of instances (Pouly, 2004; Eichenberger, 2004; Langel, 2004; Schneuwly, 2007; Pouly, 2008).

On the one hand, (Eichenberger, 2004; Lehmann et al., 2005) proposed an environment for the inference of Gaussian hints from Gaussian linear models written in the language *LPL* (linear programming language) (Kohlas and Hürlimann, 1988); this mathematical modelling language is equipped with a powerful indexing mechanism. *LPL* was designed for optimisation problems in linear models and not for assumption-based reasoning and the administration of a knowledge base. On the other hand, ABEL (assumption-based evidential language) was developed for assumption-based reasoning on logical and discrete models. However, it had lacked an indexing mechanism till version 3.0. With this modification, ABEL became usable for Gaussian linear models with only slight extensions.[7]
Although *LPL* could just as well have served as the modelling language for Gaussian linear models, ABEL was chosen in order to have an independent all-in-one tool for assumption-based reasoning.

As it stands, the implementation is not truly interactive: Whenever an `ask` expression is evaluated, the whole current knowledge base is passed on to the Gauss solver, and the queries are answered on a new join tree. However, since symmetric Gaussian potentials have inverses, non-deterministic information could be retracted from the knowledge base on the join tree by *updating* the join tree using the techniques from (Schneuwly, 2007). Furthermore, (Schneuwly, 2007) also proposes tree modification algorithms for queries which are not covered by the current tree. These extensions would make the environment truly interactive.

[7] In particular, the `sum` construct has been added and the `ask` command extended for indexed sets by Norbert Lehmann.

13 Examples

In the previous chapter, it has been shown how ABEL can be used to formulate Gaussian linear models and queries about variables in the model. Four simple Gaussian linear models will now be discussed and analysed in the spirit of assumption-based reasoning, using ABEL as inference machine.

Chapter Outline

The following four examples will be discussed.

- A simple measurement model: An unknown quantity is measured several times.
- A wholesale price estimation model (Pearl, 1988; Lehmann et al., 2005; Kohlas and Monney, 2008): On the one hand, mean profits and asking prices of the vendors can be used for diagnostic estimation. On the other hand, expert's knowledge on production costs and marketing costs can be used for predictive estimation. Finally, both types of knowledge can be combined.
- A porfolio estimation example (Liu et al., 2003a): A portfolio is modelled as a linear combination of asset variables, whose mean return has a Gaussian distribution.
- A 2-D tracking model (Russell and Norvig, 2003; p.555ff.): Noisy observations of a trajectory in a plane are filtered and smoothed.

13.1 A Simple Measurement Model

The simple measurement model from Example 6.14 can be formulated in ABEL as follows:

```
(empty)
(load "~/Desktop/cl-statistics.lisp")

(tell
 (const n 300)
```

```
)

(tell
 (type I (integer 1 n))
)

(tell
 (var x real)
 (var z[I] real)
 (ass o[I] real (gauss 1))
)

(tell
 (forall
  (k I)
  (= z[k] (+ x o[k]))
  ))

(defmacro sample-observations (n)
  (defun observation-iter (i)
    (if (= i 0)
      nil
      (cons `(= z ,(values (read-from-string
                          (concatenate
                            'string "[" (write-to-string i) "]")))
        ,(random-normal :mean 0 :sd 1))
        (observation-iter (- i 1)))))
  `(observe
    ,@(observation-iter n)))

(sample-observations 300)


(ask x)
```

Here, 300 values are sampled from independent Gaussian variables with mean 0 and variance 1.[1] By applying assumption-based reasoning, this Gaussian linear model then yields the sample mean $\sum_{i=1}^{300} z_i$ and the variance $\frac{1}{300}$. Therefore, ABEL gives the following output:

```
[X]
C =



c =



Sigma =
```

[1]The function `random-normal` is used from the package `cl-statistics.lisp`. See `http://compbio.uchsc.edu/Hunter_lab/Hunter/cl-statistics.lisp`, accessed 2009/2/20.

```
 0.00333

mu =

 -0.00666

Lambda =

 300.00000

nu =

 -1.99697
```

This output has to be interpreted in the following way: The information concerns the vector of variables $x = (X)$; the deterministic part is given by the system $Cx = c$ (which is empty in this case); the non-deterministic part is given by the symmetric Gaussian potential (Λ, ν) and the corresponding mean vector μ and variance-covariance matrix Σ if they exist.

13.2 A Wholesale Price Estimation Model

This example considers a small causal model for estimating the wholesale price of a car (Pearl, 1988; Lehmann et al., 2005; Kohlas and Monney, 2008).

In this model, there are observations of quantities that influence this wholesale price (like production cost and marketing cost) and quantities that are influenced by the wholesale price (like dealer asking prices). Besides, each observation has an associated Gaussian random term simulating the variation that estimation and profits can have. Then, inferences are made on the wholesale price, i.e. what is the wholesale price of the car given the costs or/and the final selling prices asked by dealers.

More precisely, the wholesale price is influenced by the production cost, the marketing cost, and the industry profit. On the other hand, the wholesale price influences the asking prices on the market; the wholesale price is thus estimated on the basis of two dealers' asking prices. In summary, the following variables are used:

Main variable to be inferred on

X**:** Wholesale price

Quantities influencing the wholesale price

U_1**:** Production cost

U_2**:** Marketing cost

U_3**:** Industry profit

Quantities influenced by the wholesale price

Y_1**:** Dealer-1 asking price

Y_2: Dealer-2 asking price

Furthermore, there is information about how some of these quantities can be computed, namely the production and marketing costs and the dealer asking prices:

- There are two independent experts' estimations for both the production cost and the marketing cost;
- the mean profit of each dealer over the past few years and its variance are known;
- there is a known mean of the industry profit.

In summary, the following observations will be used:

Estimates by experts for the production cost

I_1: Expert 1

I_2: Expert 2

Estimates by experts for the marketing cost

J_1: Expert 1

J_2: Expert 2

Dealers mean profit over past years

Z_1: Dealer-1

Z_2: Dealer-2

Some of the quantities defined above may have a certain degree of error or imprecision. Their degree of reliability is measured by attributing a Gaussian random variable with each estimation. This then induces the following system of equations:

$$\left\{ \begin{array}{lcl} W & = & U_1 + U_2 + U_3 + \Omega_W \\ Y_1 & = & X + Z_1 + \Omega_{Y_1} \\ Y_2 & = & X + Z_2 + \Omega_{Y_2} \\ I_1 & = & U_1 + \Omega_{I_1} \\ I_2 & = & U_1 + \Omega_{I_2} \\ J_1 & = & U_2 + \Omega_{J_1} \\ J_2 & = & U_2 + \Omega_{J_2}, \end{array} \right. \tag{13.1}$$

Here, the assumption variables are distributed normally with zero mean and variance as following:

Variable	Ω_W	Ω_{I_1}	Ω_{I_2}	Ω_{J_1}	Ω_{J_2}	Ω_X	Ω_{Y_1}	Ω_{Y_2}
Standard deviation	σ_W	σ_{I_1}	σ_{I_2}	σ_{J_1}	σ_{J_2}	σ_X	σ_{Y_1}	σ_{Y_2}

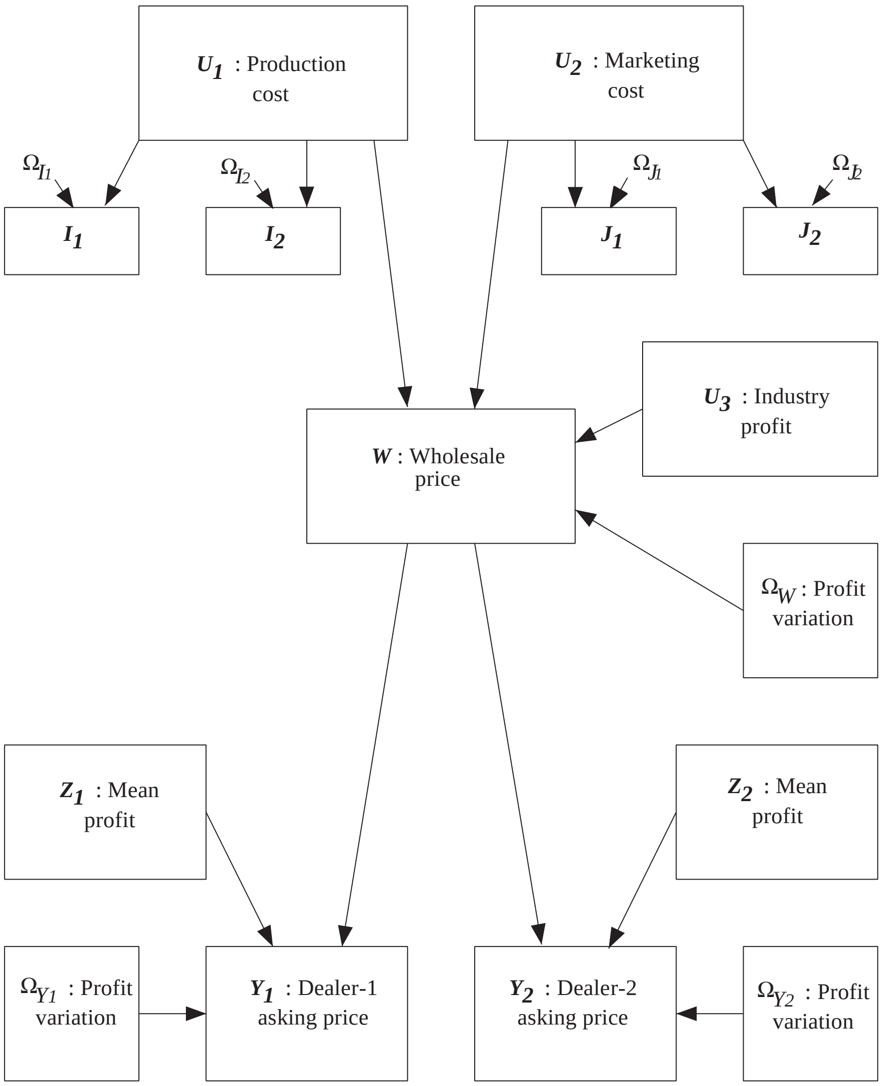


FIGURE 13.1: A causal model used for estimating the wholesale price X of a car

This is illustrated in Figure 13.1 by a directed acyclic graph where a variable in a node is the sum of all the variables in the in-going nodes. The model described here will be used to infer on the wholesale price W (i.e. the value of W will be estimated in the light of different observations) in three ways: a *diagnostic estimation*, using information observed on variables influencing the wholesale price W; a *predictive estimation*, using information observed on variables which are influenced by the wholesale price W; *combined diagnostic and predictive estimates*, using all the available information.

In ABEL, the inference about W can be set up in the following way:

```
(defun wholesale-setup ()
  (empty)
  (tell
    (var W real))            ; the price to be estimated
  )
```

Here, a function `setup` is defined which calls the ABEL macros `empty` and `tell`: First, the knowledge base is emptied and the real-valued variable `W` is declared in the ABEL knowledge base.

Diagnostic Estimation

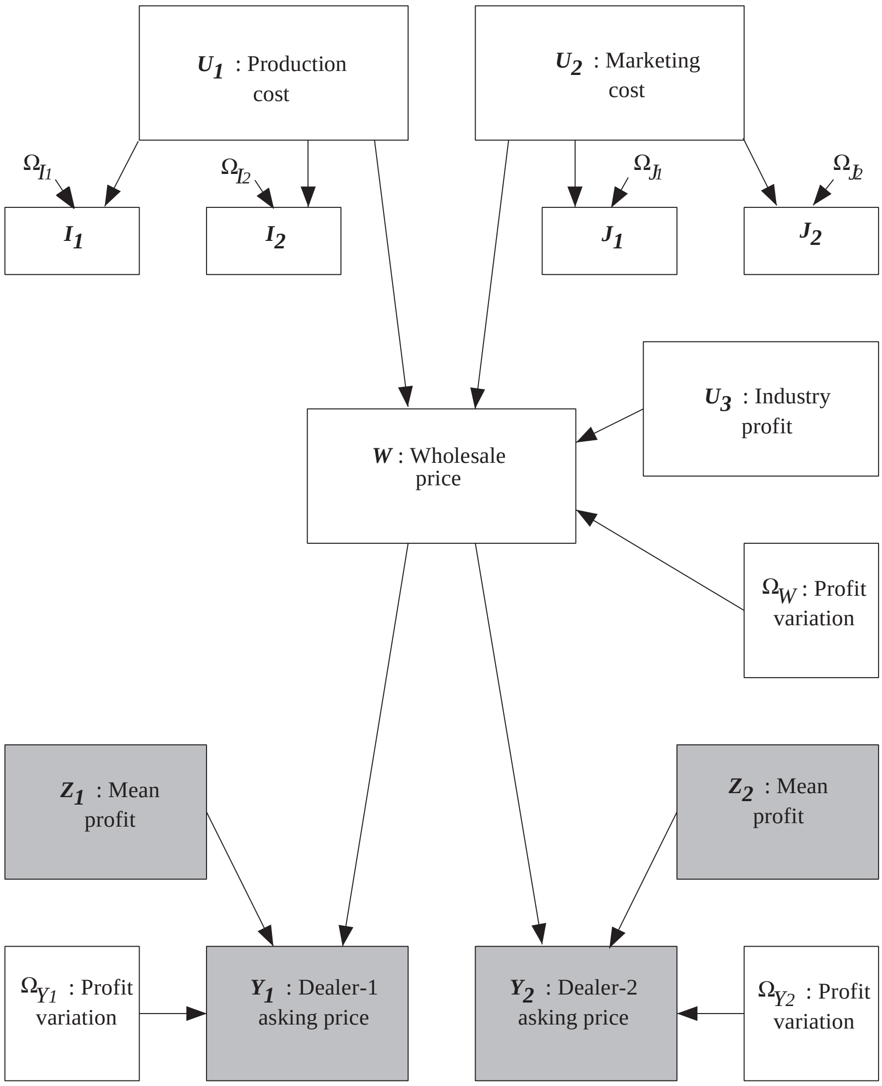
In the first case of diagnostic estimation, there are observations of the variables which are influenced by the wholesale price, thus the name diagnostic. Figure 13.2 shows the values of the observed variables and a graphical representation. Here, the nodes of the variables which have been observed are shaded. The query that computes the diagnostic estimation of the variable W can be formulated in ABEL in the following way:

```
(defun diagnostic-model ()
  (tell
    (const number-of-vendors 2))
  (tell
    (type vendors (integer 1 number-of-vendors)))
  (tell
    (var Z[vendors] real)  ; mean profit
    (var Y[vendors] real)  ; asking prices
    (ass o_y[vendors] real (gauss 90000 1000000)))
  (tell
    (forall
     (vendor vendors)
     (= Y[vendor]
        (+ W Z[vendor] o_y[vendor]))))
  (observe
    (= Y[1] 8000)
    (= Y[2] 10000)
    (= Z[1] 1000)
    (= Z[2] 1000))
)

(defun ask-diagnostic ()
```

Variable	Value
Y_1	8000 $
Y_2	10000 $
Z_1	1000 $
Z_2	1000 $
σ_{Y_1}	1000 $
σ_{Y_2}	300 $

(a) Data in the diagnostic problem



(b) Graphical representation

Figure 13.2: Diagnostic estimation

```
  (wholesale-setup)
  (diagnostic-model)
  (ask W)
)

(ask-diagnostic)
```

Since the first dealer is less shaky about in the asking price, the first estimation should be given more importance in the combined estimation. Intuitively, the price estimate should be between the two asking prices, but closer to the value of the second dealer. But what about the reliability of this combined estimation? Since the first dealer is much less reliable than the second, the reliability of the combined estimate cannot be much greater than that of the second dealer. More technically, the Gaussian linear system corresponding to the diagnostic case is

$$\begin{cases} W + \Omega_{Y_1} & = Y_1 - Z_1 \\ W + \Omega_{Y_2} & = Y_2 - Z_2. \end{cases} \tag{13.2}$$

Then, the first equation can be subtracted from the second to form the equivalent system

$$\begin{cases} W + \Omega_{Y_1} & = Y_1 - Z_1 \\ \Omega_{Y_2} - \Omega_{Y_1} & = (Y_2 - Z_2) - (Y_1 - Z_1). \end{cases} \tag{13.3}$$

This corresponds to a transformation of (13.2) by the regular matrix

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

Let

$$\Omega_Y = \begin{bmatrix} \Omega_{Y_1} \\ \Omega_{Y_2} \end{bmatrix} \sim \mathcal{N}(0, \Sigma_Y), \qquad \Sigma = \begin{bmatrix} \sigma_{Y_1}^2 & 0 \\ 0 & \sigma_{Y_2}^2 \end{bmatrix}.$$

Then, the transformed disturbance variable $\Xi = B\Omega_Y$ is distributed normally according to $\mathcal{N}(0, B\Sigma B')$. However, since the second component Ξ_2 of Ξ is constant, the distribution can be conditioned on this event. This yields an estimated price

$$\mu(W; Y_1, Y_2, Z_1, Z_2) = \Sigma(W; Y_1, Y_2, Z_1, Z_2) \left(\frac{Y_1 - Z_1}{\sigma_{Y_1}^2} + \frac{Y_2 - Z_2}{\sigma_{Y_2}^2} \right) \tag{13.4}$$

with

$$\Sigma(W; Y_1, Y_2, Z_1, Z_2) = \left(\frac{1}{\sigma_{Y_1}^2} + \frac{1}{\sigma_{Y_2}^2} \right)^{-1} = \frac{\sigma_{Y_1}^2 \sigma_{Y_2}^2}{\sigma_{Y_1}^2 + \sigma_{Y_2}^2}. \tag{13.5}$$

For the given numbers, the ABEL output for the diagnostic problem is:

```
[W]
C =
```

```
c =

Sigma =

 82568.80734

mu =

 7165.13761

Lambda =

 0.00001

nu =

 0.08678
```

Predictive Estimation

The variables U_1, U_2, U_3 all influence the wholesale price. Therefore, the following model is called predictive. The production cost U_1 and the marketing cost U_2 are estimated by two experts; one estimation is given for the industry profit U_3:

$$\begin{cases} I_1 & = U_1 + \Omega_{I_1} \\ I_2 & = U_1 + \Omega_{I_2} \\ J_1 & = U_2 + \Omega_{J_1} \\ J_2 & = U_2 + \Omega_{J_2}, \\ W & = U_1 + U_2 + U_3 + \Omega_W \end{cases} \tag{13.6}$$

Figure 13.3 shows the values of the observed variables in a graphical representation where the nodes of the variables which have been observed are shaded.

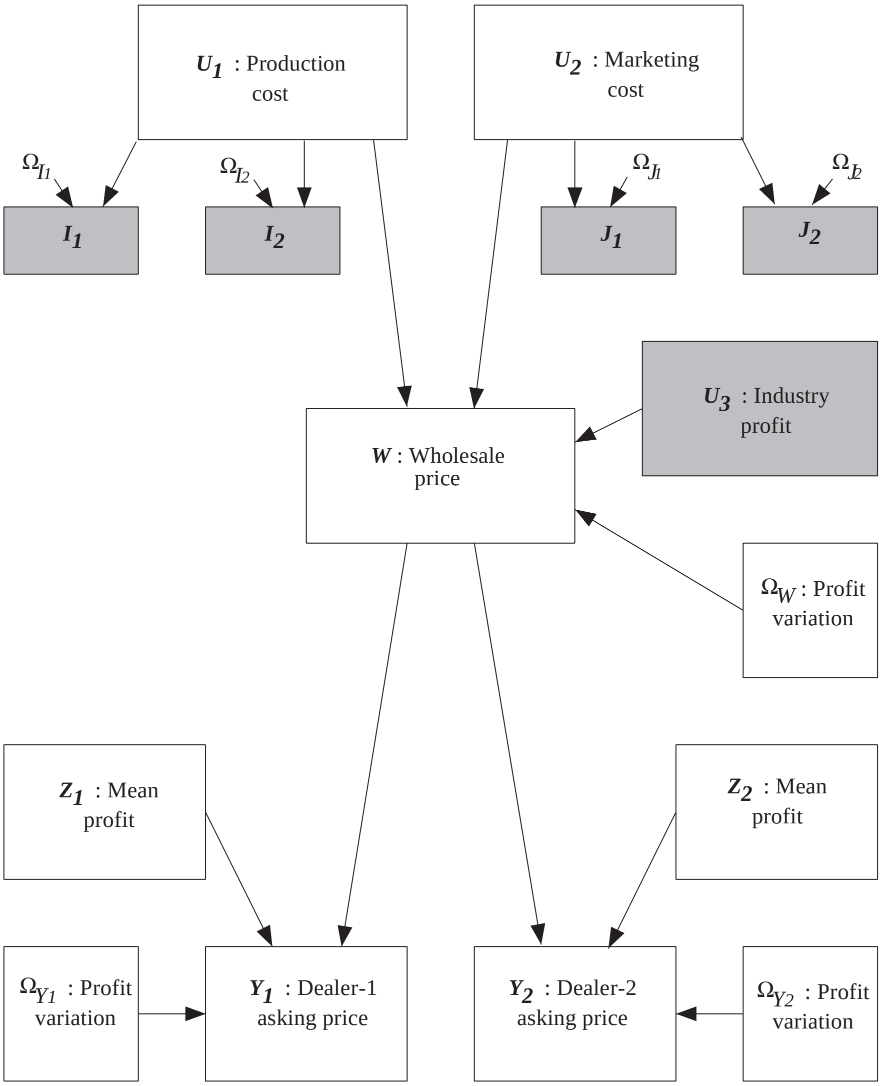
The query that computes the predictive estimation of the variable W can be formulated in Abel in the following way:

```
(defun predictive-model ()
  (tell
    (const number-of-experts 2))
  (tell
    (type experts (integer 1 number-of-experts)))

  (tell
    (var I[experts] real)   ; production cost estimation
    (var J[experts] real)   ; marketing cost estimation
    (var U3 real)           ; manufacturer profit estimation
    (var U1 real)           ; production cost
    (var U2 real)           ; marketing cost
    (ass o_i[experts] real
```

Variable	Value	Standard deviation	Value
U_3	1000 \$	σ_W	300 \$
I_1	5000 \$	σ_{I_1}	200 \$
I_2	6500 \$	σ_{I_2}	300 \$
J_1	500 \$	σ_{J_1}	50 \$
J_2	600 \$	σ_{J_2}	20 \$

(a) Data in the predictive estimation problem



(b) Graphical representation

FIGURE 13.3: Predictive estimation

```
          (gauss 40000 90000)); production cost estimation
    (ass o_j[experts] real
          (gauss 2500 400)) ; marketing cost estimation
    (ass o real (gauss 90000))) ; manufacturer profit
  (tell
    (forall
     (expert experts)
     (= I[expert]
        (+ U1 o_i[expert])
        ))
    (forall
     (expert experts)
     (= J[expert]
        (+ U2 o_j[expert])))
    (= W
       (+ U1 U2 U3 o)))
  (observe
    (= I[1] 5000)
    (= I[2] 6500)
    (= J[1] 500)
    (= J[2] 600)
    (= U3 1000))
)


(defun ask-diagnostic ()
  (wholesale-setup)
  (diagnostic-model)
  (ask W)
)

(defun ask-predictive ()
  (wholesale-setup)
  (predictive-model)
  (ask W)

(ask-predictive)
```

This problem is analysed similarly, so the respective estimates for U_1 and U_2 are in this case

$$\mu(U_1; I_1, I_2) = \Sigma(U_1; I_1, I_2)\left(\frac{I_1}{\sigma^2_{I_1}} + \frac{I_2}{\sigma^2_{I_2}}\right), \tag{13.7}$$

$$\mu(U_2; J_1, J_2) = \Sigma(U_2; J_1, J_2)\left(\frac{J_1}{\sigma^2_{J_1}} + \frac{J_2}{\sigma^2_{J_2}}\right) \tag{13.8}$$

with

$$\Sigma(U_1; I_1, I_2) = \frac{\sigma^2_{I_1}\sigma^2_{I_2}}{\sigma^2_{I_1} + \sigma^2_{I_2}}, \tag{13.9}$$

$$\Sigma(U_2; J_1, J_2) = \frac{\sigma^2_{J_1}\sigma^2_{J_2}}{\sigma^2_{J_1} + \sigma^2_{J_2}}. \tag{13.10}$$

These results can then be used to deal with the Gaussian linear system

$$\begin{cases} W - U_1 - U_2 + \Omega_W = U_3 \\ U_1 + \Xi_{I_1} = \mu(U_1; I_1, I_2) \\ U_2 + \Xi_{U_2} = \mu(U_2; J_1, J_2), \end{cases} \tag{13.11}$$

where $\Omega_W \sim \mathcal{N}(O, \sigma^2_W)$, $\Xi_{U_1} \sim \mathcal{N}(O, \Sigma(U_1; I_1, I_2)$ and $\Xi_{U_2} \sim \mathcal{N}(O, \Sigma(U_2; I_1, I_2)$. The result of the predictive estimation is then

$$\mu(W; I_1, I_2, J_1, J_2, U_3) = \mu(U_1; I_1, I_2) + \mu(U_2; J_1, J_2) + U_3, \tag{13.12}$$

$$\Sigma(W; I_1, I_2, J_1, J_2, U_3) = \Sigma(U_1; I_1, I_2) + \Sigma(U_2; J_1, J_2) + \sigma^2_W. \tag{13.13}$$

For the given numbers, the ABEL output for the predictive problem is:

```
[W]
C =


c =


Sigma =

 118037.13528

mu =

 7047.74536

Lambda =

 0.00001

nu =

 0.05971
```

Combined Diagnostic and Predictive Estimation

Here, the values of all the variables are observed, of those which influence the wholesale price, as well as of the ones influenced by it. Figure 13.4 shows a graphical representation where the nodes of the variables which have been observed are shaded. Then, inference can be either made from the whole model or by combining the two submodels, and both methods yield the same result. The former case can be handled by ABEL by the following query:

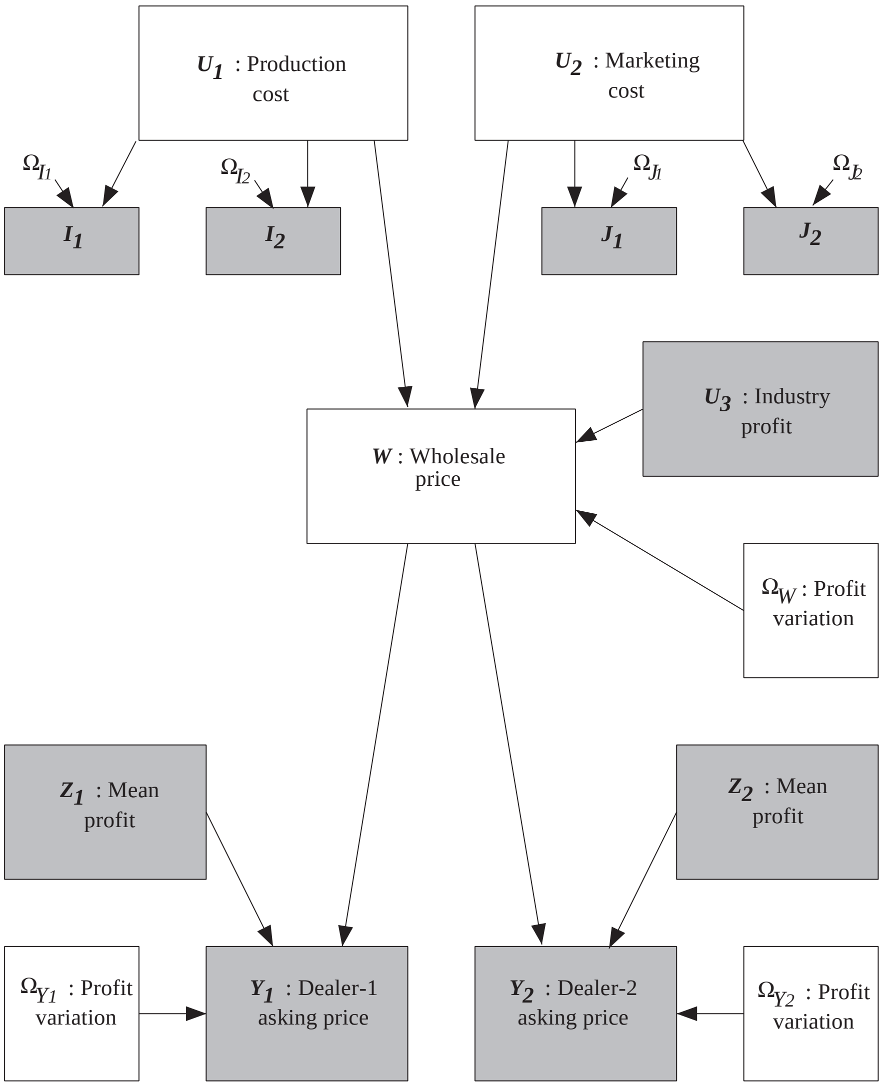


Figure 13.4: Combined diagnostic and predictive estimation

```
(defun ask-combined ()
   (wholesale-setup)
   (diagnostic-model)
   (predictive-model)
   (ask W)
)

(ask-combined)
```

In terms of the intermediate results, the combined estimate is

$$\begin{aligned} &\mu(W; Y_1, Y_2, Z_1, Z_2, I_1, I_2, J_1, J_2) \\ =&\Sigma(W; Y_1, Y_2, Z_1, Z_2, I_1, I_2, J_1, J_2) \\ &\left(\frac{\mu(W; Y_1, Y_2, Z_1, Z_2)}{\Sigma(W; Y_1, Y_2, Z_1, Z_2)} + \frac{\mu(W; I_1, I_2, J_1, J_2, U_3)}{\Sigma(W; I_1, I_2, J_1, J_2, U_3)}\right) \end{aligned} \tag{13.14}$$

with

$$\begin{aligned} &\Sigma(W; Y_1, Y_2, Z_1, Z_2, I_1, I_2, J_1, J_2) \\ =&\frac{\Sigma(W; Y_1, Y_2, Z_1, Z_2)\Sigma(W; I_1, I_2, J_1, J_2, U_3)}{\Sigma(W; Y_1, Y_2, Z_1, Z_2) + \Sigma(W; I_1, I_2, J_1, J_2, U_3)} \end{aligned} \tag{13.15}$$

The ABEL output looks as follows:

```
[W]
C =


c =


Sigma =

 48583.73264

mu =

 7116.81931

Lambda =

 0.00002

nu =

 0.14649
```

13.3 Portfolio Estimation

This model describes a portfolio estimation based on the expected performance of the stocks it is composed of (Liu et al., 2003a).

i	α_i	β_{iG}	β_{iM}	σ_i
1	0.03	0.60	0.40	0.08
2	0.03	0.45	0.25	0.04
3	0.03	0.50	0.30	0.05

i	μ_i	σ_i
G	-5%	2%
M	1%	8%

Table 13.1: Sample data for multiple regression

A financial asset is characterised by a mean return and a Gaussian error term. A portfolio is modelled as a linear combination of asset variables. Then, for a given portfolio composition, one can infer on the expected return and on the reliability of this estimation. Liu et al. (2003a) analyse this problem in terms of linear belief functions. Here, assumption-based reasoning is applied, which leads to the same results. In particular, since the model is Gaussian linear, this leads to a Gaussian hint (Monney, 2003).

A portfolio is evaluated using a multifactor regression model for stocks i and factors k,

$$r_i = \alpha + \sum_k \beta_{ik} F_k + \epsilon_i \tag{13.16}$$

where r_i is the return on stock i, β_{ik} the responsiveness of the stock i to factor k, and the $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ are stochastically independent random components. Furthermore, information available for the individual factors F_k can be given by $F_k = \mu_{F_k} + \Omega_k, \Omega_k \sim \mathcal{N}(0, \sigma_{F_k}^2)$ or just $F_k = \mu_{F_k}$.

Consider the following scenario from (Liu et al., 2003a): A portfolio consists of three gold mining stocks S_1, S_2 and S_3. Each stock S_i $(i = 1, \ldots, 3)$ is given by a mean α_i and is assumed to be influenced by three factors: by the forecast of the change of the market return M, by the forecast of the price of gold G, and by a firm specific unknown term $F_i \sim \mathcal{N}(0, \sigma_i^2)$. The modelled percentage of change of the gold price is denoted by μ_G with a tolerance $F_G \sim \mathcal{N}(0, \sigma_G^2)$. Similarly, the relative change of the stock market return is given by μ_M with a tolerance $F_M \sim \mathcal{N}(0, \sigma_M^2)$. The responsiveness of stock i to the gold price is given by β_{iG} and to the stock market return by β_{iM}. This induces the following model:

$$\begin{cases} G & = \mu_G + F_G \\ M & = \mu_M + F_M \\ S_i & = \alpha_i + \beta_{iG} G + \beta_{iM} M + F_i, \qquad i = 1, \ldots, 3. \end{cases} \tag{13.17}$$

A central bank is selling a large amount of gold. Based on historical data or personal experience, it can be expected that this transaction negatively impacts the gold price by $\mu_G = 5\%$ on the average. However, the actual rate of change could vary with standard deviation $\sigma_G = 2\%$. Before China joined the WTO, one might have speculated that this could boost the stock market by $\mu_M = 10\%$ on the average with a wide spread $\sigma_M = 8\%$. The data of this regression model is summarised in Table 13.1. This can be written in Abel as follows:

```
(empty)
```

```
(tell
 (const number-of-stock-items 3)
 (const number-of-factors 2)
)

(tell
 (type stock (integer 1 number-of-stock-items))
 (type factors (integer 1 number-of-factors))
)

(tell
 (var mean-return[stock] real)
 (var responsiveness[stock factors] real)
 (var portfolio[stock] real)
 (var mean-impact[factors] real)
 (var impact[factors] real)
 (var stock-return[stock] real)
 (var P real)
 (ass eS[stock] real (gauss 0.0064 0.0016 0.0025))
 (ass eI[factors] real (gauss 0.0004 0.0064))
)


(observe
  (= mean-return[1] 0.03)
  (= mean-return[2] 0.03)
  (= mean-return[3] 0.03)
  (= mean-impact[1] -0.05)
  (= mean-impact[2] 0.1)
  (= portfolio[1] 0.2)
  (= portfolio[2] 0.7)
  (= portfolio[3] 0.1)
  (= responsiveness[1 1] 0.60)
  (= responsiveness[2 1] 0.45)
  (= responsiveness[3 1] 0.50)
  (= responsiveness[1 2] 0.40)
  (= responsiveness[2 2] 0.25)
  (= responsiveness[3 2] 0.30)
 )

 (tell
  (forall
   (factor factors)
   (= impact[factor] (+ mean-impact[factor] eI[factor])))

  (forall (stock-item stock)
    (= stock-return[stock-item]
       (+ mean-return[stock-item]
          (sum (factor factors)
               (* responsiveness[stock-item factor]
                  impact[factor]))
          eS[stock-item])))
```

$i \setminus j$	μ_i	S_1	S_2	S_3	P
S_1	0.0400	0.0076	0.0007	0.0009	0.0021
S_2	0.0325		0.0021	0.0006	0.0017
S_3	0.0350			0.0032	0.0009
P	0.0343				0.0017

Table 13.2: The result of the portfolio estimation

```
  (= P
     (sum (stock-item stock)
          (* portfolio[stock-item]
             stock-return[stock-item])))
)
```

Applying assumption-based reasoning leads to the results summarised in Table 13.2. In Abel, the corresponding query is for the marginal on S_1, S_2, S_3 is:

```
(ask (stock-return))
```

The output is given in Table 13.2:

```
[STOCK-RETURN[2], STOCK-RETURN[1], STOCK-RETURN[3]]
C =


c =


Sigma =

 0.00208 0.00075 0.00057
 0.00075 0.00757 0.00089
 0.00057 0.00089 0.00318

mu =

 0.03250
 0.04000
 0.03500

Lambda =

 517.76836 -41.63738 -81.28274
 -41.63738 139.96564 -31.66127
 -81.28274 -31.66127 338.30176

nu =

 12.31708
 3.13727
```

```
 7.93242
```

If one is only interested in P, one can ask the query

```
(ask P)
```

and gets

```
[P]
C =


c =


Sigma =

 0.00168

mu =

 0.03425

Lambda =

 595.61747

nu =

 20.39990
```

If one asks about all four variables by

```
(ask (stock-return P))
```

then the pseudo-concentration matrix of the resulting symmetric potential is not regular and the potential can thus not be converted into a Gaussian potential:

```
[STOCK-RETURN[2], STOCK-RETURN[1], P, STOCK-RETURN[3]]
C =

 -0.70000 -0.20000 1.00000 -0.10000

c =

 0.00000

Lambda =

 252.49200 -70.53463 154.30253 -83.34936
 -70.53463 145.10814 -22.90784 -25.55230
 154.30253 -22.90784 100.51595 -29.14250
 -83.34936 -25.55230 -29.14250 343.12510
```

```
nu =

 7.75224
 1.83303
 6.52120
 7.28030
```

13.4 Kalman Filtering and Smoothing for a Simple Tracking Problem

The following example is inspired by (Russell and Norvig, 2003; p.555ff.)[2]: An object moves in a two-dimensional plane; its position is measured at constant intervals, but the measurements are noisy. Further, it is assumed that the accelerations in both directions are constant in each interval and follow a Gaussian distribution. This can be modelled in Abel as follows:

```
(tell
 (const k 10)
 (const n 15)
)

(tell
 (type T (integer 1 n))
 (type T-observed (integer 1 k))
 (type T-without-last (integer 1 (- n 1)))
 (type T-observed-without-last (integer 1 (- k 1)))
)

(tell
 (var x[T] real)
 (var x-vel[T] real)
 (var y[T] real)
 (var y-vel[T] real)
 (var x-obs[T-observed] real)
 (var y-obs[T-observed] real)
 (ass omega-x[T-without-last] real (gauss 3))
 (ass omega-y[T-without-last] real (gauss 3))
 (ass x-acc[T-without-last] real (gauss 2))
 (ass y-acc[T-without-last] real (gauss 2))
 (var noise-x[T] real)
 (var noise-y[T] real)
)


(tell
 ;; movement model
 (forall
  (i T-without-last)
  (= x[(+ i 1)] (+ x[i] x-vel[i]))
```

[2] See also `http://en.wikipedia.org/wiki/Kalman_Filter#Example`, accessed 2009/3/9.

```
   (= y[(+ i 1)] (+ y[i] y-vel[i]))
   (= x-vel[(+ i 1)] (+ x-vel[i] x-acc[i]))
   (= y-vel[(+ i 1)] (+ y-vel[i] y-acc[i]))
   )

 ;; observation model
 (forall
   (i T-without-last)
   (= noise-x[(+ i 1)]
      (+ (* 0.50  noise-x[i]) omega-x[i]))
   (= noise-y[(+ i 1)]
      (+ (* 0.50  noise-y[i]) omega-y[i]))
   )
 (forall
   (i T-observed)
   (= x-obs[i]
      (+ x[i] noise-x[i]))
   (= y-obs[i]
      (+ y[i] noise-y[i]))
   )
 )


(observe
   (= x-obs[1] 5) (= y-obs[1] 100)
   (= x-obs[2] 13) (= y-obs[2] 90)
   (= x-obs[3] 23) (= y-obs[3] 86)
   (= x-obs[4] 32) (= y-obs[4] 77)
   (= x-obs[5] 41) (= y-obs[5] 78)
   (= x-obs[6] 52) (= y-obs[6] 60)
   (= x-obs[7] 63) (= y-obs[7] 48)
   (= x-obs[8] 75) (= y-obs[8] 51)
   (= x-obs[9] 86) (= y-obs[9] 44)
   (= x-obs[10] 100) (= y-obs[10] 60)
)



;; smoothing
(ask (x[1] y[1]))
(ask (x[2] y[2]))
(ask (x[3] y[3]))
(ask (x[4] y[4]))
(ask (x[5] y[5]))
(ask (x[6] y[6]))
(ask (x[7] y[7]))
(ask (x[8] y[8]))
(ask (x[9] y[9]))
;; filtering
(ask (x[10] y[10]))
;;prediction
(ask (x[11] y[11]))
```

```
(ask (x[12] y[12]))
(ask (x[13] y[13]))
(ask (x[14] y[14]))
(ask (x[15] y[15]))
```

The fictitious data and the inferred values are shown in Figure 13.5: The filled dots are the data points, the inner small circles are the smoothed, filtered and predicted values, and the radius of the outer circles around them is the standard deviation.

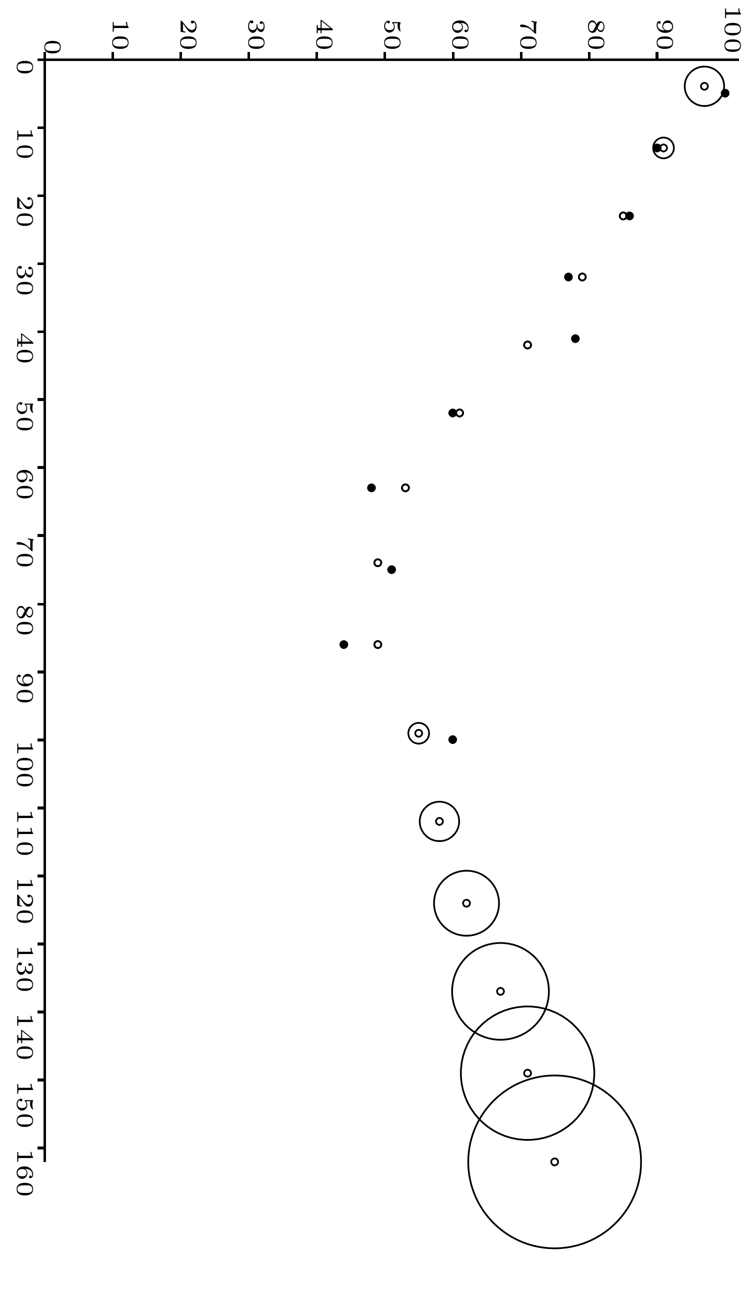


FIGURE 13.5: Tracking an object moving in the plane

Part V

Conclusion

14

Synopsis and Discussion

Chapter Outline

In this chapter, a synopsis and a discussion of this thesis are given. In Section 14.1, the theoretical results of this thesis are reconsidered. In Section 14.2, the computational aspects are discussed. Finally, in Section 14.3, open questions for future work are briefly summarised.

14.1 Theoretical Considerations

Abstracting from deterministic knowledge and general Gaussian linear systems, the exposition of this thesis is summarised in Figure 14.1: The white, the bright and

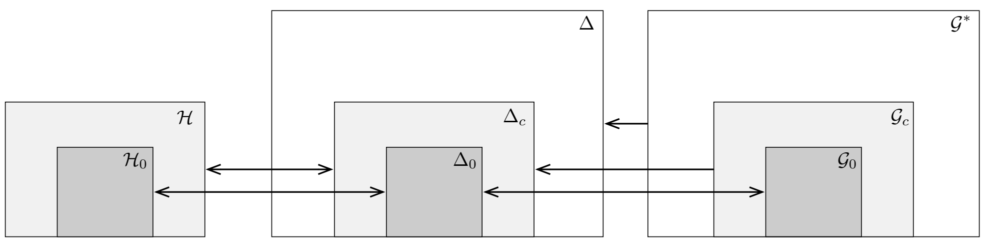


FIGURE 14.1: Gaussian hints, symmetric Gaussian potentials and separative extension of Gaussian potentials

dark grey-shaded areas denote corresponding subsets of Gaussian hints, symmetric Gaussian potentials and the separative extension of Gaussian potentials. An arrow indicates that the corresponding valuation (sub-)algebra is embedded in the one pointed to.

1. *Precise Gaussian hints $\mathcal{H}_0$, Gaussian potentials $\mathcal{G}$ and symmetric Gaussian potentials Δ_0 with positive definite pseudo-covariance matrix are isomorphic.* The isomorphicity of precise Gaussian hints and Gaussian potentials was established in Section 6.7; since $\mathcal{G}_0$ is the image of Gaussian potentials in the

separative extension, this establishes the isomorphicity of $\mathcal{H}_0$ and $\mathcal{G}_0$. Furthermore, since $\mathcal{G}^*$ can be embedded in Δ (Theorem 9.22) with Δ_0 being the image of $\mathcal{G}_0$ in Δ and since the elements of $\mathcal{G}_0$ are fully marginalisable, it also follows that $\mathcal{G}_0$ and Δ_0 are isomorphic valuation algebras with full marginalisation.

2. *Gaussian hints $\mathcal{H}$ and conditional symmetric Gaussian potentials Δ_c are isomorphic and extend conditional Gaussian potentials $\mathcal{G}_c$ with respect to marginalisation.* Gaussian hints and conditional Gaussian potentials correspond to symmetric Gaussian potentials with non-negative definite pseudo-covariance matrix: The one-to-one correspondence of Gaussian hints and conditional Gaussian potentials was established via the intermediate of conditional Gaussian densities in Section 7.3; the one-to-one correspondence of conditional Gaussian potentials and their symmetric counterparts was established in Chapter 9. Furthermore, since Gaussian hints and conditional symmetric Gaussian potentials are both fully marginalisable, they are isomorphic. However, the elimination of vacuous variables has not been explained within the algebraic theory of separative valuation algebras.

3. *Symmetric Gaussian potentials Δ extend the separative extension of Gaussian potentials $\mathcal{G}^*$ with respect to marginalisation.* The one-to-one correspondence of Gaussian quotients and symmetric Gaussian potentials was established in Section 9.1. Further, these valuation algebras both have inverses, which is not the case for Gaussian hints (except in the case of neutral Gaussian hints, which are idempotent).

It is remarkable that the algebraic approach of separative valuation algebras reflects the geometric approach of Gaussian hints regarding equivalence and combination. However, marginalisation is weaker in the algebraic approach; more precisely, the elimination of vacuous variables is not explained. Symmetric Gaussian potentials are therefore a generalisation of both Gaussian hints and conditional Gaussian potentials, since they embody a counterpart for every element in the separative extension and since marginalisation of conditional Gaussian potentials is fully defined. *Symmetric Gaussian potentials are thus the most general of all three representations.* In particular, conditional symmetric Gaussian potentials are fully marginalisable, and they also have inverses (although marginalisation of general symmetric Gaussian potentials is only partially defined). The latter property can be exploited in the Lauritzen-Spiegelhalter architecture. Furthermore, the representation of Gaussian linear information by a symmetric Gaussian potential is unique, which is not the case for Gaussian hints and Gaussian quotients.

Because of the aforementioned correspondences, it is argued that conditional Gaussian potentials contain the full information of a corresponding Gaussian linear system up to equivalence. This is remarkable since, in general, there are multiple functional models and hints inducing the same parametric distributional model. However, in the Gaussian linear case,

- the focal sets form a partition and

- these partitioning elements and singleton hypotheses are in one-to-one correspondence.

Therefore, in this particular case, the plausibility density values of the singleton hypotheses determine the full plausibility function.

Moreover, in the Gaussian linear case, conditional Gaussian densities are related to the same Gaussian hint (if they are equal up to a positive constant factor). Consider the example in the 2-dimensional xy-plane in Figure 14.2: The same focal

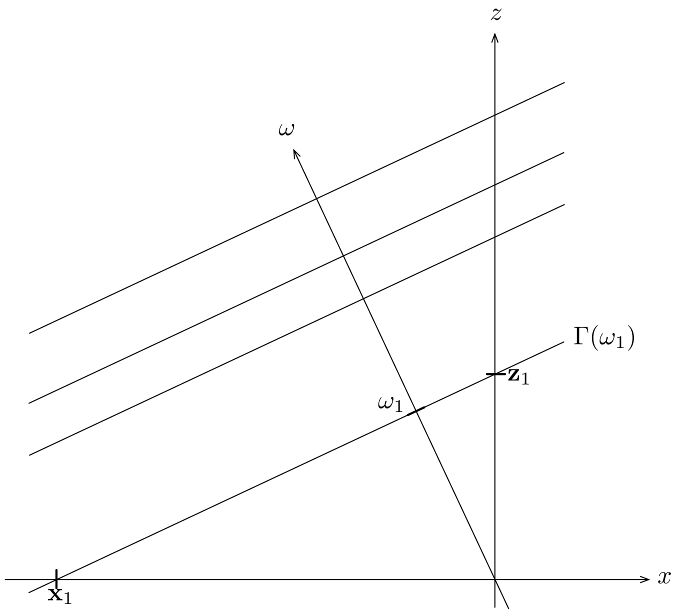


Figure 14.2: Indexing focal sets in different ways

sets are the parallel straight lines which could be indexed by the ω-, the x- and the z-axis; for instance, the indices ω_1, $\mathbf{x}_1$ and $\mathbf{z}_1$ on the corresponding axis could be used to index the straight line $\Gamma(\omega_1)$. In general, a conditional Gaussian density can be seen as a distribution over the focal sets (which form a partition); the constant positive factor only depends on the tail variables or the ω-axis, respectively. This stems from the fact that the focal sets of Gaussian hints contain points of the same conditional Gaussian density (with respect to the same fixed set of tail variables).

14.2 Computational Aspects

The representation of Gaussian linear information by symmetric Gaussian potentials is computationally attractive.

- Combination is only addition of the pseudo-mean vector and the pseudo-concentration matrix. Marginalisation by iterative variable elimination is fast and allows to easily detect and reduce vacuous variables.

- Furthermore, it is easy to derive the symmetric Gaussian potential from a Gaussian linear system, whereas the algorithms for the inference of a Gaussian hint are more expensive (Monney, 2003; Eichenberger, 2004).

- However, if the Gaussian distributions in the model are given in terms of covariance matrices instead of concentration matrices, these matrices have first to be inverted in order to get a Gaussian linear system as defined in this thesis. This may be worthwhile if the model can initially be split into "independent" factors with small domains by using the *GroupEquations* Algorithm 3 of Section 12.3.

14.3 Future Work

The following issues were not covered in detail in this thesis and could be analysed in future work.

- *Discrete variables*: The Gaussian linear information may depend on discrete variables. The resulting mixed distribution can be approximated by a CG-potential as discussed in (Lauritzen and Wermuth, 1984; Lauritzen, 1992; Cowell et al., 1999; Lauritzen and Jensen, 2001).

- *Numerical stability*: The numerical stability of the algorithms has not been investigated.

- *Interactive environment*: The implementation is not truly interactive: When answering a new query set, the whole current knowledge base is passed on to the Gauss solver, and a new join tree is built. Instead, information could be retracted from the knowledge base on the join tree by *updating* the join tree, using the techniques from (Schneuwly, 2007). Furthermore, (Schneuwly, 2007) also proposes tree modification algorithms for queries which are not covered by the current tree. These extensions would make the environment truly interactive.

- *Hypothesis evaluation*: No algorithms have been developed for the numerical evaluation of hypotheses.

Appendices

A
Some Results from Matrix Algebra

In this chapter, some basic results on regular and symmetric positive definite matrices are loosely collected for reference from the text. Integer- and variable-indexed matrices are used at the author's convenience.

LEMMA **A.1**. *Let $A, B \in \mathbb{R}(m,n)$. Then, for any matrix $C \in \mathbb{R}(r,m)$ of full column rank $\mathrm{r}(C) = m$ and any matrix $D \in \mathbb{R}(n,p)$ of full row rank $\mathrm{r}(D) = n$,*

(1) *$CA = CB$ implies $A = B$,*

(2) *$AD = BD$ implies $A = B$, and*

(3) *$CAD = CBD$ implies $A = B$.* ⊘

PROOF. (1) Let C_i denote the ith column of C ($i = 1,\ldots,m$), let E_j denote the jth column of $CA = CB$, let $a_{ij} = A(i,j)$, and $b_{ij} = B(i,j)$ ($i \in \{1,\ldots,m\}, j \in \{1,\ldots,n\}$). Then, every column of E is a linear combination of the columns of C,

$$E_j = \sum_{i=1}^{m} a_{ij} C_i = \sum_{i=1}^{m} b_{ij} C_i, \qquad j \in \{1,\ldots,n\}.$$

Since the columns of C are linearly independent, it follows by Lemma 4.3.5 of (Harville, 1997; p.34) that $a_{ij} = b_{ij}$ for all i and for all j, hence $A = B$.

(2) D having full row rank p implies that D' has full row rank p. Notice that $AD = BD \iff (AD)' = (BD)' \iff D'A' = D'B'$. Then, by (1), $A' = B'$, hence $A' = A'' = B'' = B$.

(3) From $(CA)D = (CB)D$, it follows by (2) that $CA = CB$ and then by (1) that $A = B$. □

LEMMA **A.2**. (1) *Let $A \in \mathbb{R}(m,n)$. Then, the set*

$$\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : \ A\mathbf{x} = 0\} \tag{A.1}$$

is a linear subspace of dimension $\dim(\mathcal{N}(A)) = n - \mathrm{r}(A)$, called null space.

(2) *Let* $A \in \mathbb{R}(m,n)$ *and* $\mathbf{z} \in \mathbb{R}^m$. *Let*

$$\Gamma = \{\mathbf{x} \in \mathbb{R}^n : \ A\mathbf{x} = \mathbf{z}\}$$

and

$$\Gamma_{\mathbf{p}} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{p} + \mathbf{x}^*, \mathbf{x}^* \in \mathcal{N}(A)\}$$

for $\mathbf{p} \in \mathbb{R}^n$. *Then,* $\Gamma = \Gamma_{\mathbf{p}}$ *for all* $\mathbf{p} \in \mathbb{R}^n$ *such that* $A\mathbf{p} = \mathbf{z}$.

(3) *Let* $A_1 \in \mathbb{R}(m_1, n)$, $A_2 \in \mathbb{R}(m_2, n)$, $z_1 \in \mathbb{R}^{m_1}$ *and* $z_2 \in \mathbb{R}^{m_2}$ *such that*

$$\Gamma_1 = \{\mathbf{x} \in \mathbb{R}^n : A_1\mathbf{x} = z_1\} = \{\mathbf{x} \in \mathbb{R}^n : A_2\mathbf{x} = z_2\} = \Gamma_2.$$

Then, $\mathcal{R}(A_1) = \mathcal{R}(A_2)$.

(4) *Let* $A_1, A_2 \in \mathbb{R}(m,n)$ *be matrices of full row rank* m *and let* $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^m$ *and let*

$$\Gamma_1 = \{\mathbf{x} : \ A_1\mathbf{x} = \mathbf{z}_1\}, \quad \Gamma_2 = \{\mathbf{x} : \ A_2\mathbf{x} = \mathbf{z}_2\}.$$

Then,

$$\Gamma_1 = \Gamma_2$$

if and only if

$$A_2 = TA_1, \quad \mathbf{z}_2 = T\mathbf{z}_1$$

for some regular matrix $T \in \mathbb{R}(m,m)$. ⊘

Proof. (1) See Lemma 11.4.1 of (Harville, 1997; p.143f.).

(2) On the one hand, assume $\mathbf{p} \in \mathbb{R}^n$ such that $A\mathbf{p} = \mathbf{z}$ and $\mathbf{x}^* \in \mathcal{N}(A)$, then $A(\mathbf{p} + \mathbf{x}^*) = A\mathbf{p} + 0_m = \mathbf{z}$ shows that $\Gamma \supseteq \Gamma_{\mathbf{p}}$. On the other hand, if $\mathbf{x} \in \Gamma$ and $\mathbf{p} \in \mathbb{R}^n$ such that $A\mathbf{p} = \mathbf{z}$, i.e. $A\mathbf{x} = \mathbf{z}$, then $A(\mathbf{x} - \mathbf{p}) = A\mathbf{x} - A\mathbf{p} = \mathbf{z} - \mathbf{z} = 0_m$ shows that $\mathbf{x}^* = (\mathbf{x} - \mathbf{p}) \in \mathcal{N}(A)$ and $\mathbf{x} = \mathbf{p} + \mathbf{x}^*$, hence also $\Gamma \subseteq \Gamma_{\mathbf{p}}$.

(3) Let $\mathbf{p} \in \Gamma_1 = \Gamma_2$. Then, according to (2),

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{p} + \mathbf{x}^*, \mathbf{x}^* \in \mathcal{N}(A_1)\} = \Gamma_1 = \Gamma_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{p} + \mathbf{x}^*, \mathbf{x}^* \in \mathcal{N}(A_2)\}.$$

Hence, $\mathcal{N}(A_1) = \mathcal{N}(A_2)$ by (1). Assume $x \in \mathcal{R}(A_1)$ and $x \notin \mathcal{R}(A_2)$. Then, $A_1x = 0$ and $A_2x \neq 0$ implies that $\mathcal{N}(A_1) \neq \mathcal{N}(A_2)$. This shows that $\mathcal{R}(A_1) = \mathcal{R}(A_2)$.

(4) On the one hand, if there is a regular matrix T such that $A_2 = TA_1$ and $\mathbf{z}_2 = T\mathbf{z}_1$, then $\mathbf{x} \in \Gamma_1$ implies that $\mathbf{z}_2 = T\mathbf{z}_1 = T(A_1\mathbf{x}) = A_2\mathbf{x}$, i.e. $\mathbf{x} \in \Gamma_2$, hence $\Gamma_1 \subseteq \Gamma_2$; the converse implication follows since T^{-1} is regular and $A_1 = T^{-1}TA_1 = T^{-1}A_2$ and $\mathbf{z}_1 = T^{-1}T\mathbf{z}_1 = \mathbf{z}_2$.
On the other hand, assume that $\Gamma_1 = \Gamma_2$. Then, $\mathcal{R}(A_1) = \mathcal{R}(A_2)$ by (3). Hence, there is a matrix $T \in \mathbb{R}(m,m)$ such that $TA_1 = A_2$. Furthermore, since $\mathrm{r}(T) \leq n$ and $n = \mathrm{r}(A_2) = \mathrm{r}(TA_1) \leq \mathrm{r}(T)$, it follows that T is regular. □

Lemma **A.3**. *Let $K_1, K_2 \in \mathrm{I\!R}(m,m)$ be symmetric and positive definite matrices. Then, there is then a regular matrix $T \in \mathrm{I\!R}(m,m)$ such that $K_2 = T'K_1T$.* ⊘

Proof. In light of Corollary 14.3.13 of (Harville, 1997; p.219), there are regular matrices $P_1, P_2 \in \mathrm{I\!R}(m,m)$ such that

$$K_1 = P_1'P_1, \quad K_2 = P_2'P_2.$$

Since P_1 and P_2 are regular, $\mathcal{C}(P_1) = \mathrm{I\!R}^m = \mathcal{C}(P_1)$, hence there is a matrix $T \in \mathrm{I\!R}(m,m)$ such that $P_2 = P_1T$; since $m = \mathrm{r}(P_2) = \mathrm{r}(P_1T) \leq \mathrm{r}(T) \leq m$, it follows that $\mathrm{r}(T) = m$, i.e. T is regular and

$$K_2 = P_2'P_2 = T'P_1'P_1T = T'K_1T.$$ □

Lemma **A.4**. *Let $K_1, K_2 \in \mathrm{I\!R}(m,m)$ be symmetric matrices. Then*

$$\mathbf{x}'K_1\mathbf{x} = \mathbf{x}'K_2\mathbf{x}$$

for all $\mathbf{x} \in \mathrm{I\!R}^m$ implies

$$K_1 = K_2.$$ ⊘

Proof. Assume $\mathbf{x}'K_1\mathbf{x} = \mathbf{x}'K_2\mathbf{x}$ for all $\mathbf{x} \in \mathrm{I\!R}^m$. Then, for all $\mathbf{x}, \mathbf{y} \in \mathrm{I\!R}^m$,

$$\begin{aligned}\mathbf{x}'K_1\mathbf{x} + \mathbf{y}'K_1\mathbf{y} + 2\mathbf{x}'K_1\mathbf{y} =& (\mathbf{x}+\mathbf{y})'K_1(\mathbf{x}+\mathbf{y})\\ =& (\mathbf{x}+\mathbf{y})'K_2(\mathbf{x}+\mathbf{y})\\ =& \mathbf{x}'K_2\mathbf{x} + \mathbf{y}'K_2\mathbf{y} + 2\mathbf{x}'K_2\mathbf{y}\\ =& \mathbf{x}'K_1\mathbf{x} + \mathbf{y}'K_1\mathbf{y} + 2\mathbf{x}'K_2\mathbf{y},\end{aligned}$$

hence indeed $\mathbf{x}'K_1\mathbf{y} = \mathbf{x}'K_2\mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathrm{I\!R}^m$. Since $K_1(i,j) = e_i'K_1e_j = e_i'K_2e_j = K_2(i,j)$ for $\mathbf{x} = e_i$ (the ith column of I_m) and e_j (the jth column of I_m) ($i = 1, \ldots, m$; $j = 1, \ldots, n$), indeed $K_1 = K_2$. □

Lemma **A.5**. *Let $A : x \times x \to \mathrm{I\!R}$ and $B : y \times y \to \mathrm{I\!R}$, $x, y \in D$ be symmetric non-negative definite matrices. Then, $A = A^{\uparrow x \cup y} + B^{\uparrow x \cup y}$ is symmetric non-negative definite. Furhermore, if A or B is positive definite, then A is positive definite.* ⊘

Proof. Let $u = x \cup y$. For every vector $\mathbf{u} \in \mathrm{I\!R}^u$, $\mathbf{u} \neq 0_u$, $(\mathbf{u}^{\downarrow x})'A\mathbf{u}^{\downarrow x} \geq 0$ and $(\mathbf{u}^{\downarrow y})'B\mathbf{u}^{\downarrow y} \geq 0$, and hence

$$\mathbf{u}'A\mathbf{u} = \mathbf{u}'(A^{\uparrow u} + B^{\uparrow u})\mathbf{u} = (\mathbf{u}^{\downarrow x})'A\mathbf{u}^{\downarrow x} + (\mathbf{u}^{\downarrow y})'B\mathbf{u}^{\downarrow y} \geq 0.$$

A similar argument shows that if A or B is positive definite, then A is positive definite. □

LEMMA **A.6**. *The inverse of a symmetric positive definite matrix $K \in \mathbb{R}(m,m)$, partitioned*

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

is

$$K^{-1} = \begin{pmatrix} c_{11} & -c_{11}K_{12}{K_{22}}^{-1} \\ -c_{22}K_{21}{K_{11}}^{-1} & c_{22} \end{pmatrix} \tag{A.2}$$

where

$$c_{11} =(K_{11} - K_{12}{K_{22}}^{-1}K_{21})^{-1} \tag{A.3}$$

and

$$c_{22} =(K_{22} - K_{21}{K_{11}}^{-1}K_{12})^{-1}.$$

Furthermore,

$$c_{11}K_{12}{K_{22}}^{-1} =(c_{22}K_{21}{K_{11}}^{-1})'. \tag{A.4}$$

⊘

PROOF. According to Corollary 14.2.11 of (Harville, 1997; p.214), a symmetric positive definite matrix is invertible, and its inverse is symmetric positive definite. Therefore, according to Theorem 8.5.11 of (Harville, 1997; p.99), the Schur complements c_{11} and c_{22} exist. Then,

$$\begin{aligned} &\begin{pmatrix} c_{11} & -c_{11}K_{12}{K_{22}}^{-1} \\ -c_{22}K_{21}{K_{11}}^{-1} & c_{22} \end{pmatrix} \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} \\ =&\begin{pmatrix} c_{11}K_{11} - c_{11}K_{12}{K_{22}}^{-1}K_{21} & c_{11}K_{12} - c_{11}K_{12}{K_{22}}^{-1}K_{22} \\ -c_{22}K_{21}{K_{11}}^{-1}K_{11} + c_{22}K_{21} & -c_{22}K_{21}{K_{11}}^{-1}K_{12} + c_{22}K_{22} \end{pmatrix} \\ =&\begin{pmatrix} c_{11}{c_{11}}^{-1} & 0 \\ 0 & c_{22}{c_{22}}^{-1} \end{pmatrix} = I_m, \end{aligned}$$

which proves (A.2). According to Corollary 14.2.11 of (Harville, 1997; p.214), the inverse of a symmetric positive definite matrix exists and is symmetric positive definite, hence K^{-1} is symmetric, thus equation (A.4) holds as well. □

LEMMA **A.7**. *Let $K : x \times x \to \mathbb{R}$ be a matrix, $x \in D$ finite, $x = x_1 \cup x_2$, $x_1 \cap x_2 = \emptyset$,*

$$K = \begin{pmatrix} K^{\downarrow x_1} & K^{\downarrow x_1,x_2} \\ K^{\downarrow x_2,x_1} & K^{\downarrow x_2} \end{pmatrix},$$

such that the principal submatrix $K^{\downarrow x_1}$ is symmetric and positive definite. Then, K is symmetric and positive definite if and only if $K^{\downarrow x_2} - K^{\downarrow x_2,x_1}(K^{\downarrow x_1})^{-1}K^{\downarrow x_1,x_2}$ is symmetric positive definite.

⊘

Proof. First, the "if" part is proved. By Corollary 14.2.11 of (Harville, 1997; p.214), any symmetric positive definite matrix is invertible and the inverse is symmetric positive definite, so K^{-1} must be symmetric positive definite. By Corollary 14.2.12 (Harville, 1997; p.214), every principal submatrix of a symmetric positive definite matrix is symmetric positive definite, so $(K^{-1})^{\downarrow x_2}$ is symmetric positive definite, and so is its inverse $((K^{-1})^{\downarrow x_2})^{-1}$. At the same time, by Lemma A.6,

$$((K^{-1})^{\downarrow x_2})^{-1} = K^{\downarrow x_2} - K^{\downarrow x_2,x_1}(K^{\downarrow x_1})^{-1} K^{\downarrow x_1,x_2}.$$

This proves the "if" part.
In order to prove the "only if" part, the fact is needed that a matrix $\Sigma \in \mathbb{R}(x,x)$ is symmetric positive definite if and only if there exists a regular matrix $P \in \mathbb{R}(x,x)$ such that $\Sigma = P'P$ (Corollary 14.3.13 of (Harville, 1997; p.219)). In light of this and the necessity of the condition in the assertion of the theorem, there are regular matrices $A \in \mathbb{R}(x,x)$ and $B \in \mathbb{R}(x,x)$ such that

$$\begin{aligned} K^{\downarrow x_1} &= A'A, \\ K^{\downarrow x_2} - K^{\downarrow x_2,x_1}(K^{\downarrow x_1})^{-1} K^{\downarrow x_1,x_2} &= B'B. \end{aligned}$$

Define

$$C = \begin{pmatrix} A & A'^{-1} K^{\downarrow x_1,x_2} \\ 0_{x_2,x_1} & B \end{pmatrix}.$$

Since A and B are regular, C is regular by Lemma 8.5.4 (Harville, 1997; p.90). Since, in light of result (8.2.8) of (Harville, 1997; p.82),

$$A^{-1}A'^{-1} = (A'A)^{-1} = (K^{\downarrow x_1})^{-1},$$

also $K = C'C$. Therefore, by Corollary 14.3.13 of (Harville, 1997; p.219), K is symmetric and positive definite. This concludes the proof of the "only if" part. □

Corollary **A.8**. *Let $K_{11} : x_1 \times x_1 \to \mathbb{R}$ be a symmetric positive definite matrix, $x_1 \in D$ finite, and $K_{12} : x_1 \times x_2 \to \mathbb{R}$, $x_2 \in D$ finite such that $x_1 \cap x_2 = \emptyset$. Let $x = x_1 \cup x_2$. Then, there are symmetric positive definite matrices $K \in \mathbb{R}(x,x)$ and $K_{22} \in \mathbb{R}(x_2,x_2)$ such that*

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{12}' & K_{22}. \end{pmatrix}.$$ ⊘

Proof. Define

$$K_{22} = I_{x_2} + K_{12}' {K_{11}}^{-1} K_{12}.$$

By Lemma A.7, it is sufficient and necessary for K to be positive definite that

$$K_{22} - K_{12}' {K_{11}}^{-1} K_{12} = I_{x_2}$$

is symmetric positive definite, which is the case. □

LEMMA **A.9**. *Let $K : x \times x \to \mathrm{I\!R}$ be a symmetric positive definite, $x \in D$ finite, $x = x_1 \cup x_2$, $x_1 \cap x_2 = \emptyset$,*

$$K = \begin{pmatrix} K^{\downarrow x_1} & K^{\downarrow x_1, x_2} \\ K^{\downarrow x_2, x_1} & K^{\downarrow x_2} \end{pmatrix}.$$

Then

$$K = \begin{pmatrix} K^{\downarrow x_1} & K^{\downarrow x_1, x_2} \\ K^{\downarrow x_2, x_1} & K^{\downarrow x_2, x_1} \left(K^{\downarrow x_1}\right)^{-1} K^{\downarrow x_1, x_2} \end{pmatrix} \qquad \oslash$$

is a symmetric non-negative definite matrix of rank $|x_1|$.

PROOF. By Corollary 14.2.12 of (Harville, 1997; p.214), every principal submatrix of a symmetric positive definite matrix is symmetric positive definite, so $(K^{-1})^{\downarrow x_1}$ is symmetric and positive definite. By Corollary 14.2.12 of (Harville, 1997; p.214), any symmetric positive definite matrix is invertible and the inverse is symmetric positive definite, so $\left(K^{\downarrow x_1}\right)^{-1}$ is symmetric and positive definite. Then, $\left(K^{\downarrow x_1}\right)^{-1}$ being positive definite implies that for every vector $y \in \mathrm{I\!R}^x$,

$$y' K^{\downarrow x_2, x_1} \left(K^{\downarrow x_1}\right)^{-1} K^{\downarrow x_1, x_2} y \quad = \quad (K^{\downarrow x_1, x_2} y)' \left(K^{\downarrow x_1}\right)^{-1} (K^{\downarrow x_1, x_2} y) \quad \geq \quad 0,$$

thus $K^{\downarrow x_2, x_1} \left(K^{\downarrow x_1}\right)^{-1} K^{\downarrow x_1, x_2}$ is symmetric and non-negative definite. Furthermore, since $\left(K^{\downarrow x_1}\right)^{-1}$ is symmetric positive definite, in light of Corollary 14.3.13 of (Harville, 1997; p.219), there is a regular matrix $A \in \mathrm{I\!R}(x_1, x_1)$ such that

$$K^{\downarrow x_1} = A'A.$$

Then, by result (8.2.8) and (8.2.4) of (Harville, 1997; p.82),

$$\left(K^{\downarrow x_1}\right)^{-1} = A^{-1}(A^{-1})'.$$

Define $C \in \mathrm{I\!R}(x_1, x)$,

$$C = \left(A, (A^{-1})' K^{\downarrow x_1, x_2}\right).$$

Since A is regular, C has full row rank $|x_1|$. Then, by results (8.2.8) and (8.2.4) of (Harville, 1997; p.82), $K = C'C$. For K to be symmetric non-negative definite, by Theorem 14.3.7 of (Harville, 1997; p.218), it is necessary and sufficient that there exists a matrix $P \in \mathrm{I\!R}(r, x)$ such that $K = P'P$ and $r = \mathrm{r}(K)$. $\square$

LEMMA **A.10**. *Let $K \in \mathrm{I\!R}(x \cup z, x \cup z)$ be a symmetric matrix of rank $\mathrm{r}(K) = |x| = r$, $x \cap z = \emptyset$, such that $K^{\downarrow x}$ is symmetric and positive definite. Then, K is symmetric and non-negative definite if and only if $K^{\downarrow z} = K^{\downarrow z, x} {K^{\downarrow x}}^{-1} K^{\downarrow x, z}$.* $\oslash$

Proof. On the one hand, assume that K is symmetric and non-negative definite. Then, in light of Theorem 14.3.7 of (Harville, 1997; p.218), there is a matrix $A \in \mathbb{R}(x \cup z, r)$ of full column rank r such that

$$K = A'A.$$

Then,

$$K^{\downarrow x} = A^{\downarrow x,r\prime} A^{\downarrow x,r}.$$

Since $K^{\downarrow x}$ is symmetric and positive definite, it is regular of rank $\mathrm{r}(K^{\downarrow x}) = r$. Hence, $r \geq \mathrm{r}(A^{\downarrow x,r}) \geq r$ shows that $A^{\downarrow x,r}$ is regular as well. Therefore, the other columns of A are linear combinations of the columns of $A_1 = A^{\downarrow x,r}$, i.e. there is a matrix $\Lambda \in \mathbb{R}(x, y)$ such that

$$A = \begin{pmatrix} A_1 & A_1\Lambda \end{pmatrix}.$$

Then,

$$K = \begin{pmatrix} A_1'A_1 & A_1'A_1\Lambda \\ \Lambda'A_1'A_1 & \Lambda'A_1'A_1\Lambda' \end{pmatrix}.$$

Hence, indeed

$$K^{\downarrow z,x} K^{\downarrow x^{-1}} K^{\downarrow x,z} = \Lambda'A_1'A_1(A_1'A_1)^{-1}A_1'A_1\Lambda = \Lambda'A_1'A_1\Lambda = K^{\downarrow z}.$$

On the other hand, assume $K^{\downarrow z} = K^{\downarrow z,x} K^{\downarrow x^{-1}} K^{\downarrow x,z}$. Then, since $K^{\downarrow x^{-1}}$ is symmetric and positive definite,

$$\mathbf{z}'K^{\downarrow z}\mathbf{z} = (K^{\downarrow x,z}\mathbf{z})'K^{\downarrow x^{-1}}K^{\downarrow x,z}\mathbf{z} \geq 0$$

for all $\mathbf{z} \in \mathbb{R}^z$. Hence, $K^{\downarrow z}$ is symmetric and non-negative definite. Furthermore, since $K^{\downarrow x}$ is symmetric and positive definite, it follows by Corollary 14.3.13 of (Harville, 1997; p.219) that there is a regular matrix $A_1 \in \mathbb{R}(r, x)$ such that

$$K^{\downarrow x} = A_1'A_1.$$

Then,

$$\begin{aligned} K &= \begin{pmatrix} A_1'A_1 & K^{\downarrow x,z} \\ K^{\downarrow z,x} & K^{\downarrow z,x}A_1'^{-1}A_1^{-1}K^{\downarrow x,z} \end{pmatrix} \\ &= \begin{pmatrix} A_1' \\ K^{\downarrow z,x}A_1^{-1} \end{pmatrix} \begin{pmatrix} A_1 & A_1'^{-1}K^{\downarrow z,x} \end{pmatrix}. \end{aligned}$$

The matrix

$$A = \begin{pmatrix} A_1 & A_1'^{-1}K^{\downarrow z,x} \end{pmatrix}$$

has rank $\mathrm{r}(A) = \mathrm{r}(A_1) = r$. Since $K = A'A$, Theorem 14.3.7 of (Harville, 1997; p.218) shows that K is indeed a symmetric and non-negative definite matrix of rank r. □

LEMMA **A.11**. *Let $K \in \mathbb{R}(p,p)$ be a symmetric non-negative definite matrix of rank r. Then, there is a subset $x_1 \subseteq p$ of cardinality $|x_1| = r$ such that $K^{\downarrow x_1}$ is symmetric positive definite. Furthermore, given any such subset $x_1 \subseteq p$ such that $K^{\downarrow x_1}$ is symmetric positive definite of rank r,*

$$K = \begin{pmatrix} K^{\downarrow x_1} & K^{\downarrow x_1,x_2} \\ K^{\downarrow x_2,x_1} & K^{\downarrow x_2,x_1}\left(K^{\downarrow x_1}\right)^{-1}K^{\downarrow x_1,x_2} \end{pmatrix}$$

for $x_2 = p - x_1$. ⊘

PROOF. By Theorem 14.3.7 of (Harville, 1997; p.218), for K to be symmetric non-negative definite, it is necessary and sufficient that there exists a matrix $P \in \mathbb{R}(r,p)$ of rank r such that $K = P'P$.

Then, there is a subset $x_1 \subseteq p$ of cardinality $|x_1| = r$ such that the submatrix $P_1 = P^{\downarrow r,x_1}$ is regular. The matrix

$$P_1'P_1 = K^{\downarrow x_1}$$

is symmetric positive definite by Corollary 14.3.13 of (Harville, 1997; p.219). This proves the first assertion of the lemma.

Furthermore, let $x_1 \subseteq p$ such that $K^{\downarrow x_1}$ is symmetric positive definite of rank r. Then, $K^{\downarrow x_1} = \left(P^{\downarrow r,x_1}\right)'P^{\downarrow r,x_1}$. Further, in light of Corollary 8.3.2 and Lemma 4.4.3 of (Harville, 1997; p.83;p.37), $r \leq \mathrm{r}(P^{\downarrow r,x_1}) \leq r$. Hence, $\mathrm{r}(P^{\downarrow r,x_1}) = r$. Define $P_1 = P^{\downarrow r,x_1}$ and $P_2 = P^{\downarrow r,x_2}$. Since P_1 is regular, $\mathcal{C}({P_1}^{-1})' = \mathcal{C}(P_1) \supseteq \mathcal{C}(P_2)$, i.e. the columns of P_2 are linear combinations of the columns of $({P_1}^{-1})'$. Therefore, there is a matrix $K_{12} \in \mathbb{R}(x_1, x_2)$ such that $P_2 = ({P_1}^{-1})'K_{12}$, and

$$P = \begin{pmatrix} P_1 & ({P_1}^{-1})'K_{12} \end{pmatrix}.$$

Then, in light of results (8.2.8) and (8.2.4) of (Harville, 1997; p.82),

$$K = P'P = \begin{pmatrix} P_1'P_1 & K_{12} \\ K_{12}' & K_{12}'(P_1'P_1)^{-1}K_{12} \end{pmatrix}.$$

This concludes the proof that K is of the claimed form. □

B
Gaussian Densities

In this chapter, a brief review of Gaussian densities is given.

B.1 The Gaussian Distribution as Large Quincunx

Consider the quincunx or bean machine in Figure B.1:[1] Assume that a ball (grey dot) dropped at the top funnel has equal probability 0.5 of falling left or right when hitting a pin (black dots). If there are n junction levels, the probability of a ball falling into box $k \in \{0, \dots, n\}$ is equal to the probability of falling right k times out of n. If the outcome x_i of the decision at level i is 0 (if the ball flips to the left) or 1 (if it flips to the right), then the number of times falling right is equal to the sum $S_n = \sum_{i=0}^{n} x_i$ of these outcomes, and its distribution is binomial,

$$\Pr(S_n = k) = \binom{n}{k} \cdot 0.5^k \cdot 0.5^{n-k}.$$

Since most of the time, a ball will flip right and left about the same number of times, most experiments will end in the ball landing somewhere in the middle; most paths from top to bottom end somewhere in the middle. For $n = 5$, in the long run, $\frac{20}{32}$ will end in the middle boxes 2 or 3, as shown by the grey dots at the bottom of Figure B.1.

In this quincunx setup, the decisions of falling left or right are identical and independent of each other:

- (*identical*): It is always the same decision. The ball either flips one position to the left or the right, never further away.
- (*independent* of each other): The probability of these decisions is always the same. A decision does not affect the probabilities of other decisions. Each decision has expected value $\mu = \sum_{i=0}^{1} \frac{1}{2} i = \frac{1}{2}$ and variance $\sigma^2 = \sum_{i=0}^{1} \frac{1}{2}(i - \mu)^2 = \frac{1}{4}$.

[1]The device is alleged to have been invented by Sir Francis Galton, see `http://en.wikipedia.org/wiki/Francis_Galton`, accessed 2008/6/9.

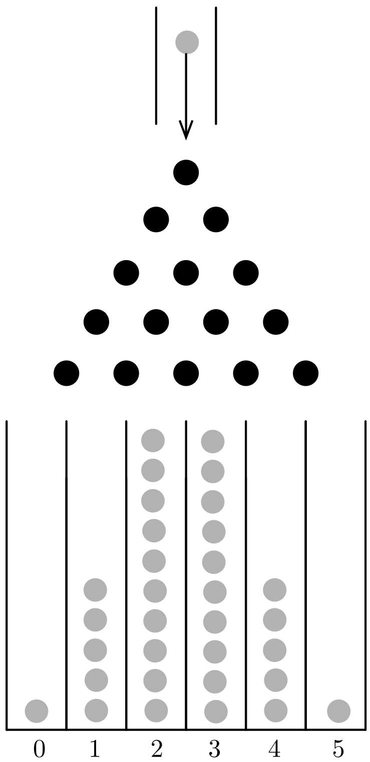


FIGURE B.1: Galton's Quincunx

Under these two assumptions, the central limit theorem says that, for large n, these probabilities can be approximated by the standard normal distribution,

$$\lim_{n\to\infty} \Pr\left(\frac{\sqrt{n}(S_n - n\mu)}{\sigma} \leq z\right) = \Phi(z).$$

As shown in Figure B.2[2], the "68–95–99.7% rule" of the normal distribution says

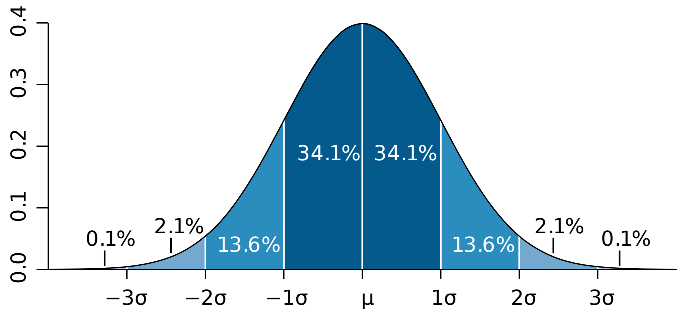


FIGURE B.2: Standard deviation

that, in the long run, 68% of outcomes lie within one standard deviation σ from the mean μ (dark blue), 95% of outcomes within 2σ (medium and dark blue), and 99.7% within 3σ (light, medium, and dark blue). Extreme values, deviating from the mean, are negligible.

If the decisions are not independent (for instance, if they aggregate or clot), the Gaussian distribution is not suited.

B.2 Relocating and Scaling the Standard Gaussian Density

The univariate standard density function of Φ is

$$\phi_{0,1}(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mathbf{x}^2}, \tag{B.1}$$

whose graph is bell-shaped. It is a probability density function since it is a non-negative real-valued function such that

$$\int_{\mathbb{R}} \phi(\mathbf{x}) d\mathbf{x} = 1.$$

[2]Source: `http://commons.wikimedia.org/wiki/Image:Standard_deviation_diagram.svg`, accessed 2008/6/9, licensed under Creative Commons Attribution 2.5 by Petter Strandmark

Notice that

$$\begin{aligned}(\int_{\mathbb{R}} \phi(\mathbf{x})d\mathbf{x})^2 &= \int_{\mathbb{R}} \phi(\mathbf{x})d\mathbf{x} \cdot \int_{\mathbb{R}} \phi(\mathbf{y})d\mathbf{y} \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{x}^2}{2}} d\mathbf{x} \cdot \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{y}^2}{2}} d\mathbf{y} \\ &= \int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-\frac{\mathbf{x}^2+\mathbf{y}^2}{2}} d\mathbf{x}d\mathbf{y} \\ &= \frac{1}{2\pi} \int_0^\infty \left(\int_0^{2\pi} e^{-\frac{r^2(\cos^2(\alpha)+\sin^2(\alpha))}{2}} \cdot |J_{(x,y),(r,\alpha)}| d\alpha \right) dr \\ &= \frac{1}{2\pi} \int_0^\infty \left(\int_0^{2\pi} e^{-\frac{r^2}{2}} r d\alpha \right) dr \\ &= \int_0^\infty e^{-\frac{r^2}{2}} r dr = [-e^{-\frac{r^2}{2}}]_{r=0}^\infty = 0 - (-1) = 1\end{aligned}$$

for

$$(\mathbf{x}, \mathbf{y}) = T^{-1}(r, \alpha) = (r \cdot \cos(\alpha), r \cdot \sin(\alpha))$$

since the Jacobian of the transformation

$$T : \mathbb{R} \times \mathbb{R} \to [0, \infty[\times[0, 2\pi[$$

to polar coordinates

$$(r, \alpha) = T(\mathbf{x}, \mathbf{y}) = (\sqrt{\mathbf{x}^2 + \mathbf{y}^2}, \arccos(\frac{x}{r}))$$

is

$$\begin{aligned}J_{(x,y),(r,\alpha)} &= \det(\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \alpha} & \frac{\partial y}{\partial \alpha} \end{pmatrix}) = \det(\begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -r \cdot \sin(\alpha) & r \cdot \cos(\alpha) \end{pmatrix}) \\ &= r \cdot \cos^2(\alpha) + r \cdot \sin^2(\alpha) = r.\end{aligned}$$

Hence, since ϕ is non-negative and therefore also $\int_{\mathbb{R}} \phi(\mathbf{x})d\mathbf{x}$, this implies that indeed $\int_{\mathbb{R}} \phi(\mathbf{x})d\mathbf{x} = 1$. The transformation from Euclidean to polar coordinates is shown in Figure B.3. Notice that

$$\cos(\alpha) = \frac{\mathbf{x}}{r}, \qquad \sin(\alpha) = \frac{\mathbf{y}}{r}.$$

The multivariate standard density function is the product

$$\begin{aligned}\phi_{0,I_n}(\mathbf{x}_1, \ldots, \mathbf{x}_n) &= \phi_{0,1}(\mathbf{x}_1) \cdot \ldots \cdot \phi_{0,1}(\mathbf{x}_n) \\ &= \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\sum_{i=1}^n \mathbf{x}_i^2} \\ &= \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\mathbf{x}'\mathbf{x}} = \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\mathbf{x}' I_n \mathbf{x}},\end{aligned} \tag{B.2}$$

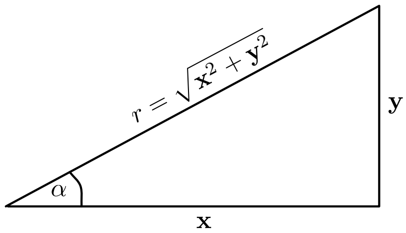


FIGURE B.3: Transformation of Euclidean to polar coordinates

whose points of equal density c form circles centered around the origin of the coordinate system since

$$\sum_{i=1}^{n} \mathbf{x}_i^2 = -2 \cdot \ln(c \cdot \sqrt{(2\pi)^n}).$$

This density ϕ_{0,I_n} can be expressed with respect to a different basis $B \in \mathbb{R}(n,n)$ of $\mathbb{R}^n$. This means transforming x to $\xi = Tx$ by $T = B^{-1}$ or substituting x for $x = T^{-1}\xi$. Since the Jacobian of this substitution is

$$J_{x,Tx} = \det(T^{-1}),$$

it holds that

$$\begin{aligned}
\phi_{0,I_n}(\mathbf{x})d\mathbf{x} =& |J_{x,Tx}| \cdot \phi_{0,I_n}(\mathbf{x})d(T\mathbf{x}) \\
=& |J_{x,Tx}| \cdot \phi_{0,I_n}(T^{-1}\xi)d\xi \\
=& |\det(T^{-1})| \cdot \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}(T^{-1}\xi)'(T^{-1}\xi)} d\xi \\
=& |\det(T^{-1})| \cdot \frac{1}{\sqrt{(2\pi)^n}} e^{-\frac{1}{2}\xi'(T^{-1\prime}T^{-1})\xi} d\xi \\
=& \sqrt{\frac{|\det(K)|}{(2\pi)^n}} \cdot e^{-\frac{1}{2}\mathbf{x}'K\mathbf{x}} d\xi = \phi_{0,K}(\xi)d\xi
\end{aligned}$$

for

$$K = T^{-1\prime}T^{-1}$$

since

$$\begin{aligned}
\sqrt{|\det(K)|} =& \sqrt{|\det(T^{-1\prime}T^{-1})|} = \sqrt{|\det(T^{-1\prime})\det(T^{-1})|} \\
=& \sqrt{|\det(T^{-1})\det(T^{-1})|} = |\det(T^{-1})|
\end{aligned}$$

in light of Theorem 13.3.4 and Lemma 13.2.1 of (Harville, 1997; p.187;p.181). Notice that the matrix K is symmetric and positive definite since T^{-1} being regular implies that for all $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}'K\mathbf{x} = \mathbf{x}'T^{-1\prime}T^{-1}\mathbf{x} = (T^{-1}\mathbf{x})'(T^{-1}\mathbf{x})$$

equals zero if and only if $T\mathbf{x} = 0_n$. The points of equal density in the graph of $\phi_{0,K}$ lie on ellipsoids around the origin of the coordinate system.

Furthermore, the density can be expressed with respect to a new origin $\mu \in \mathbb{R}^n$ of coordinates with respect to the basis B. This means transforming x to $\xi = Tx+\mu$ or substituting x for $x = T^{-1}(\xi - \mu)$. Here, again $J_{x,Tx} = \det(T^{-1})$ and

$$\begin{aligned}\phi_{0,I_n}(\mathbf{x})d\mathbf{x} =& |J_{x,Tx}| \cdot \phi_{0,I_n}(\mathbf{x})d(T\mathbf{x}+\mu)\\ =& |J_{x,Tx}| \cdot \phi_{0,I_n}(T^{-1}(\xi-\mu))d\xi\\ =& \sqrt{\frac{|\det(K)|}{(2\pi)^n}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mu)'K(\mathbf{x}-\mu)}d\xi = \phi_{\mu,K}(\xi)d\xi\end{aligned}$$

for $K = T^{-1\prime}T^{-1}$.

Conversely, in light of Theorem 14.3.7 of (Harville, 1997; p.218), every symmetric and positive definite matrix $K \in \mathbb{R}^n$ is of the form $T'T$ for some regular matrix $T \in \mathbb{R}(n,n)$. Therefore, every Gaussian density

$$\phi_{\mu,K}(\mathbf{x}) = \sqrt{\frac{|\det(K)|}{(2\pi)^n}} \cdot e^{-\frac{1}{2}(\mathbf{x}-\mu)'K(\mathbf{x}-\mu)} \tag{B.3}$$

for some $\mu \in \mathbb{R}^n$, $K \in \mathbb{R}(n,n)$ symmetric and positive definite is the result of *re-locating* by μ and *scaling* by T a random vector of n independent standard normal variables with density ϕ_{0,I_n}. Notice that T needs not be unique.

B.3 Marginalising a Multivariate Gaussian Density

Any random vector x with values in $\mathbb{R}^n$ with Gaussian density $\phi_{\mu,K}$ is the product of re-locating and scaling a random vector y of density ϕ_{0,I_n} by $\mu \in \mathbb{R}^n$ and some regular $T \in \mathbb{R}(n,n)$ such that $(TT')^{-1} = T^{-1}T = K$. Let $T_1 \in \mathbb{R}(n_1,n)$, $T_2 \in \mathbb{R}(n_2,n)$ such that

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}$$

and $n_1 + n_2 = n$, i.e.

$$x = T(y-\mu) = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix}(y-\mu) = \begin{pmatrix} T_1(y-\mu) \\ T_2(y-\mu) \end{pmatrix}.$$

Since T is regular, T_1 has full row rank n_1 and T_2 full row rank n_2. In light of Lemma 11.3.1 of (Harville, 1997; p.142), the null space

$$\mathcal{N}(T_1) = \{\mathbf{x} \in \mathbb{R}^n : \ T_1\mathbf{x} = 0\}$$

has dimension $n - \mathrm{r}(T_1) = n - n_1 = n_2$. Let $\tilde{T}_2 \in \mathbb{R}(n_2,n)$ such that the columns of $\tilde{T}_2'$ form a basis of $\mathcal{N}(T_1)$, i.e. N_2 has full row rank n_2. Then, the rows of

$$\tilde{T} = \begin{pmatrix} T_1 \\ \tilde{T}_2 \end{pmatrix}$$

are linearly independent and $\tilde{T}$ is thus regular. Then,

$$\tilde{\Sigma} = \tilde{T}\tilde{T}' = \begin{pmatrix} T_1T_1' & T_1\tilde{T}_2' \\ \tilde{T}_2T_1' & \tilde{T}_2\tilde{T}_2' \end{pmatrix} = \begin{pmatrix} T_1T_1' & 0_{n_1,n_2} \\ 0_{n_2,n_1} & \tilde{T}_2\tilde{T}_2' \end{pmatrix},$$

and hence, since the block-diagonal matrix $\tilde{\Sigma}$ is regular, T_1T_1' and $\tilde{T}_2\tilde{T}_2'$ must be regular in light of Lemma 8.5.1 of (Harville, 1997; p.88), and

$$\tilde{K} = (\tilde{T}\tilde{T}')^{-1} = \begin{pmatrix} (T_1T_1')^{-1} & 0_{n_1,n_2} \\ 0_{n_2,n_1} & (\tilde{T}_2\tilde{T}_2')^{-1} \end{pmatrix}.$$

Then,

$$\tilde{T}(y-\mu) = \begin{pmatrix} T_1(y-\mu) \\ \tilde{T}_2(y-\mu) \end{pmatrix}$$

has density

$$\phi_{\tilde{T}\mu,\tilde{K}} = \phi_{T_1\mu,(T_1T_1')^{-1}} \cdot \phi_{\tilde{T}_2\mu,(\tilde{T}_2\tilde{T}_2')^{-1}},$$

which shows that the marginal density of $x_1 = T_1(y-\mu)$ is

$$\begin{aligned} \int_{\mathbf{x}_2\in\mathrm{I\!R}^{n_2}} \phi_{\tilde{T}\mu,\tilde{K}}(\mathbf{x}_1,\mathbf{x}_2)d\mathbf{x}_2 &= \int_{\mathbf{x}_2\in\mathrm{I\!R}^{n_2}} \phi_{T_1\mu,(T_1T_1')^{-1}}(\mathbf{x}_1)\cdot\phi_{\tilde{T}_1\mu,(\tilde{T}_1\tilde{T}_1')^{-1}}(\mathbf{x}_2)d\mathbf{x}_2 \\ &= \phi_{T_1\mu,(T_1T_1')^{-1}}(\mathbf{x}_1)\cdot\int_{\mathbf{x}_2\in\mathrm{I\!R}^{n_2}} \phi_{\tilde{T}_1\mu,(\tilde{T}_1\tilde{T}_1')^{-1}}(\mathbf{x}_2)d\mathbf{x}_2 \\ &= \phi_{T_1\mu,(T_1T_1')^{-1}}(\mathbf{x}_1). \end{aligned} \tag{B.4}$$

Notice that

$$(T_1T_1')^{-1} = \Sigma_{11}{}^{-1}$$

in

$$K^{-1} = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = TT' = \begin{pmatrix} T_1T_1' & T_1T_2' \\ T_2T_1' & T_2T_{2'} \end{pmatrix}$$

even if $T_1T_2' \neq 0_{n_1,n_2}$, and in light of Lemma A.6

$$(T_1T_1')^{-1} = K_{11} - K_{12}K_{22}{}^{-1}K_{21}$$

for

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix},$$

partitioned according to n_1 and n_2 rows and columns, respectively.

In summary, the marginal of a Gaussian density $\phi_{\mu,K}$ with respect to x is given by

$$\phi_{\mu^{\downarrow x},((K^{-1})^{\downarrow x})^{-1}}. \tag{B.5}$$

B.4 Conditioning a Multivariate Gaussian Density

Let $\phi_{\mu,K}$ be a Gaussian density on $\mathbb{R}^n$ for some $\mu \in \mathbb{R}^n$, $K \in \mathbb{R}(n,n)$ symmetric and positive definite. Partition

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

and

$$K = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix}$$

according to the first n_1 and $n_2 = n - n_1$ rows and columns, respectively. Then, for $\mathbf{x}_1 \in \mathbb{R}^{n_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{n_2}$:

$$\begin{aligned}
&(\mathbf{x}_1' - \mu_1', \mathbf{x}_2' - \mu_2{}')K \begin{pmatrix} \mathbf{x}_1 - \mu_1 \\ \mathbf{x}_2 - \mu_2 \end{pmatrix} && \text{(B.6)} \\
=&(\mathbf{x}_1' - \mu_1')K_{11}(\mathbf{x}_1 - \mu_1) \\
&+ (\mathbf{x}_2' - \mu_2{}')K_{21}(\mathbf{x}_1 - \mu_1) + (\mathbf{x}_1' - \mu_1')K_{12}(\mathbf{x}_2 - \mu_2) \\
&+ (\mathbf{x}_2' - \mu_2{}')K_{22}(\mathbf{x}_2 - \mu_2) \\
=&(\mathbf{x}_1' - \mu_1' + (\mathbf{x}_2' - \mu_2{}')K_{21}K_{11}{}^{-1})K_{11}(\mathbf{x}_1 - \mu_1 + K_{11}{}^{-1}K_{12}(\mathbf{x}_2 - \mu_2)) \\
&- (\mathbf{x}_2' - \mu_2{}')K_{21}K_{11}{}^{-1}K_{12}(\mathbf{x}_2 - \mu_2) + (\mathbf{x}_2' - \mu_2{}')K_{22}(\mathbf{x}_2 - \mu_2). \\
=&(\mathbf{x}_1' - \mu_1' + (\mathbf{x}_2' - \mu_2{}')K_{21}K_{11}{}^{-1})K_{11}(\mathbf{x}_1 - \mu_1 + K_{11}{}^{-1}K_{12}(\mathbf{x}_2 - \mu_2)) \\
&+ (\mathbf{x}_2' - \mu_2{}')(K_{22} - K_{21}K_{11}{}^{-1}K_{12})(\mathbf{x}_2 - \mu_2). && \text{(B.7)}
\end{aligned}$$

The last term does not depend on $\mathbf{x}_1$ and thus becomes a constant factor in equation (B.3). Furthermore, K_{11}, being a principal submatrix of the symmetric positive definite matrix K, is symmetric and positive definite by Corollary 14.2.12 of (Harville, 1997; p.214). Therefore, observing that $K_{11}{}^{-1} = K_{11}'{}^{-1}$, the conditional distribution of fixing $\mathbf{x}_2$ is Gaussian with mean

$$\mu_1 - K_{11}{}^{-1}K_{12}(\mathbf{x}_2 - \mu_2) \tag{B.8}$$

and concentration

$$K_{11}. \tag{B.9}$$

Alternatively, the conditional distribution can be given in terms of the variance-covariance matrix

$$\Sigma = K^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11} \in \mathbb{R}(n_1, n_1)$.

LEMMA **B.1**. *The conditional Gaussian mean and concentration are*

$$\mu_1 - K_{11}{}^{-1}K_{12}(\mathbf{x}_2 - \mu_2) = \mu_1 + \Sigma_{12}\Sigma_{22}{}^{-1}(\mathbf{x}_2 - \mu_2) \tag{B.10}$$

$$K_{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}{}^{-1}\Sigma_{21})^{-1}. \tag{B.11}$$

⊘

PROOF. Since $\Sigma = K^{-1}$, being the inverse of the symmetric and positive definite matrix K, is symmetric and positive definite, Lemma A.6 shows that

$$K_{11} = \left(\Sigma_{11} - \Sigma_{12}{\Sigma_{22}}^{-1}\Sigma_{21}\right)^{-1} \tag{B.12}$$

and

$$\begin{aligned} -{K_{11}}^{-1}K_{12} &= -{K_{11}}^{-1}(-K_{11}\Sigma_{12}{\Sigma_{22}}^{-1}) \\ &= \Sigma_{12}{\Sigma_{22}}^{-1}. \end{aligned} \tag{B.13}$$

□

References

B. Anrig, R. Haenni, J. Kohlas, and N. Lehmann. Assumption-based Modeling using ABEL. In D. Gabbay, R. Kruse, A. Nonnengart, and H. J. Ohlbach, editors, *First International Joint Conference on Qualitative and Quantitative Practical Reasoning; ECSQARU–FAPR'97*, volume 1244 of *Lecture Notes in Artifical Intelligence*, pages 171–182. Springer, 1997.

S. Arnborg, D. Corneil, and A. Proskurowski. Complexity of Finding Embedings in a k-Tree. *SIAM J. of Algebraic and Discrete Methods*, 8(2):277–284, 1987.

S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra.* Number 78 in Graduate Texts in Mathematics. Springer-Verlag, 1981.

A. H. Clifford and G. B. Preston. *Algebraic Theory of Semigroups.* American Mathematical Society, Providence, Rhode Island, 1967.

R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Information Sci. and Stats. Springer, New York, 1999.

A. P. Dempster. On the difficulties Inherent in Fisher's Fiducial Argument. *American Statistical Association Journal*, 59:56–66, 1964.

A. P. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals of Math. Stat.*, 38:325–339, 1967.

A. P. Dempster. Normal Belief Functions and the Kalman Filter. Technical report, Harvard University, USA, May 1990a.

A. P. Dempster. Construction and Local Computation Aspects of Network Belief Functions. In J. Q. Smith and R. M. Oliver, editors, *Influence Diagrams, Belief Nets and Decision Analysis.* Wiley, 1990b.

Ch. Eichenberger. Implementing Gaussian Hints. Master's thesis, Department of Informatics, University of Fribourg, 2004.

D. A. S. Fraser. *The Structure of Inference.* Wiley, New York, 1968.

G. H. Golub and Ch. F. Van Loan. *Matrix Computations*, volume 3 of *Johns Hopkins Series in the Mathematical Sciences*. The Johns Hopkins University Press, second edition, 1989. Second edition.

J. H. Goodnight. A Tutorial on the SWEEP Operator. *The American Statistician*, 33(3):149–158, 1979.

R. Haenni, B. Anrig, R. Bissig, and N. Lehmann. ABEL homepage, 1998.

P. R. Halmos. *Measure Theory*. Van Nostrand, 1950.

D. A. Harville. *Matrix Algebra From a Statistician's Perspective*. Springer, 1997.

E. Hewitt and H. S. Zuckerman. The l_1 algebra of a commutative semigroup. *Amer. Math. Soc.*, 83:70–97, 1956.

F. V. Jensen, S. L. Lauritzen, and K. G. Olesen. Bayesian Updating in Causal Probabilistic Networks by Local Computation. *Computational Statistics Quarterly*, 4:269–282, 1990.

R. E. Kalman. A new Approach to Linear Filtering and Predictive Problems. *Transactions ASME, Journal of basic engineering*, 82:34–45, 1960.

J. Kohlas. *Information Algebras: Generic Structures for Inference*. Discrete Mathematics and Theoretical Computer Science. Springer-Verlag, London, Berlin, Heidelberg, 2003.

J. Kohlas and T. Hürlimann. LPL: A Structured Language for Linear Programming Modeling. *OR Spektrum*, 10:55–63, 1988.

J. Kohlas and P.-A. Monney. *Statistical Information. Assumption-Based Statistical Inference*, volume 3 of *Sigma Series in Stochastics*. Heldermann, Lemgo, Germany, 2008.

J. Kohlas and P.-A. Monney. *A Mathematical Theory of Hints. An Approach to the Dempster-Shafer Theory of Evidence*, volume 425 of *Lecture Notes in Economics and Mathematical Systems*. Springer, 1995.

J. Kohlas and P. P. Shenoy. Computation in Valuation Algebras. In J. Kohlas and S. Moral, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 5: Algorithms for Uncertainty and Defeasible Reasoning*, pages 5–39. Kluwer, Dordrecht, 2000.

J. Kohlas and R. F. Stärk. Information Algebras and Consequence Operators. Technical Report 96–14, Institute of Informatics, University of Fribourg, 1996.

F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor Graphs and the Sum-Product Algorithm. *IEEETIT: IEEE Transactions on Information Theory*, 47, 2001.

J. Langel. Local Computation with Models in Propositional Logic. Master's thesis, Department of Informatics, University of Fribourg, 2004.

S. L. Lauritzen. Propagation of Probabilities, Means, and Variances in Mixed Graphical Association Models. *Journal of the American Statistical Association*, Vol. 87(No. 420):1098–1108, December 1992.

S. L. Lauritzen and F. V. Jensen. Local Computation with Valuations from a Commutative Semigroup. *Annals of Mathematics and Artificial Intelligence*, 21 (1):51–70, 1997.

S. L. Lauritzen and F. V. Jensen. Stable local computation with conditional Gaussian distributions. *Statistics and Computing*, 11(2):191–203, 2001.

S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statis. Soc. B*, 50:157–224, 1988.

S. L. Lauritzen and N. Wermuth. Mixed Interaction Models. Technical Report R 84-8, Institute for Electronic Systems, Aalborg University, 1984.

N. Lehmann. *Argumentation System and Belief Functions.* PhD thesis, Department of Informatics, University of Fribourg, 2001.

N. Lehmann. ABEL 3.0. Talk at RUN Group, University of Berne, Switzerland, December 2005.

N. Lehmann, Ch. Eichenberger, and T. Hürlimann. Assumption-Based Reasoning with LPL. Technical Report 05-17, Department of Informatics, University of Fribourg, 2005.

L. Liu. A Theory of Gaussian Belief Functions. *International Journal of Approximate Reasoning*, Volume 14:95–126, February-April 1996a.

L. Liu. Propagation of Gaussian belief functions. In D. Fisher and H. Lenz, editors, *Learning Models from Data*, volume V of *Artificial Intelligence and Statistics*, chapter 8, pages 79–88. Springer-Verlag, New York, 1996b.

L. Liu. Local Computation of Gaussian Belief Functions. *International Journal of Approximate Reasoning*, Volume 22:217–248, December 1999.

L. Liu, C. Shenoy, and P. P. Shenoy. Knowledge Representation and Integration for Portfolio Evaluation Using Linear Belief Functions. Technical report, School of Business, University of Kansas, 2003a.

L. Liu, C. Shenoy, and P. P. Shenoy. A Linear Belief Function Approach to Portfolio Evaluation. In U. Kjaerulff and C. Meek, editors, *Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03)*, pages 370–377, San Francisco, CA, 2003b. Morgan Kaufmann Publishers.

D. MacKay. *Information Theory, Inference, and Learning Algorithms.* Cambridge University Press, September 2003.

D. Maier. *The Theory of Relational Databases.* Pitman, London, 1983.

K. Mellouli. *On the Propagation of Beliefs in Networks Using the Dempster-Shafer Theory of Evidence*. PhD thesis, School of Business, University of Kansas, 1988.

P.-A. Monney. *A Mathematical Theory of Arguments for Statistical Evidence*. Contributions to Statistics. Physica-Verlag, 2003.

J. Neveu. *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris, 1964.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.

M. Pouly. *A Generic Architecture for Local Computation*. PhD thesis, Department of Informatics, University of Fribourg, 2008.

M. Pouly. NENOK 1.1 User Guide. Technical Report 06-02, Department of Informatics, University of Fribourg, 2006.

M. Pouly. A generic Architecture for local Computation. Master's thesis, Department of Informatics, University of Fribourg, 2004.

A. Rényi. *Foundations of Probability*. Holden Day, 1970.

S. T. Roweis and Z. Ghahramani. A Unifying Review of Linear Gaussian Models. *Neural Computation*, 11(2):305–345, 1999.

S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, second edition edition, 2003.

C. Schneuwly. *Computing in Valuation Algebras*. PhD thesis, Department of Informatics, University of Fribourg, 2007.

C. Schneuwly, M. Pouly, and J. Kohlas. Local Computation in Covering Join Trees. Technical Report 04-16, Department of Informatics, University of Fribourg, 2004.

G. Shafer. Allocations of Probablity. *Ann. of Prob.*, 7:827–839, 1979.

G. Shafer. An Axiomatic Study of Computation in Hypertrees. Working Paper 232, School of Business, University of Kansas, 1991.

G. Shafer. *Probabilistic Expert Systems*. Number 67 in CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, PA, 1996.

P. P. Shenoy. Valuation-Based Systems: A Framework for Managing Uncertainty in Expert Systems. In L. A. Zadeh and J. Kacprzyk, editors, *Fuzzy Logic for the Management of Uncertainty*, pages 83–104. John Wiley & Sons, 1992.

P. P. Shenoy. Binary join trees for computing marginals in the Shenoy-Shafer architecture. *International Journal of Approximate Reasoning*, 17:239–263, 1997.

P. P. Shenoy and G. Shafer. Propagating Belief Functions using Local Computation. *IEEE Expert*, 1(3):43–52, 1986.

P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. S. Levitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 4*, volume 9 of *Machine intelligence and pattern recognition*, pages 169–198, Amsterdam, 1990. Elsevier.

R. P. Srivastava and L. Liu. Applications of Belief Functions in Business Decisions: A Review. *Information Systems Frontiers*, 5(4):359–378, 2003.

G. L. Steele. *Common Lisp – the Language.* Digital Press, 1990.

Index

Curriculum Vitae

Personalien

Vorname	:	Christian
Name	:	Eichenberger
Adresse	:	Birkenweg 7
Wohnort	:	5702 Niederlenz (Schweiz)
Geburtsdatum	:	6. Juli 1979
Geburtsort	:	Aarau
Telefon	:	+41 26 300 83 31
E-Mail	:	`christianmarkus.eichenberger@unifr.ch`
Heimatort	:	Lenzburg & Beinwil a. S.
Zivilstand	:	ledig

Schulbildung

1986 – 1991	:	Primarschule, Niederlenz
1991 – 1995	:	Bezirksschule, Möriken-Wildegg
1995 – 1999	:	Alte Kantonsschule Aarau (Schweiz) Abschluss: *Matura Typus A*
1999 – 2004	:	Studium der Informatik mit Nebenfach rätoromanische Sprache und Kultur (nur BSc), Universität Freiburg (Schweiz) Abschluss: *MSc in Informatik*

Sprachen

- Muttersprache Deutsch
- Gute Kenntnisse in Französisch
- Gute Kenntnisse in Englisch
- Gute Kenntnisse in Sursilvan

BERUFLICHE TÄTIGKEITEN

2000 – 2001	:	ABB Semiconductors, Lenzburg Ferienjob als Programmierer für Reportgenerierung
2003–2004	:	Universität Freiburg i.Ue., Departement für Informatik Unterassistent für das Erstjahresprojekt
2004–2009	:	Universität Freiburg i.Ue., Departement für Informatik Assistent