

https://dspace.org/

Wishart laws on convex cones

Mamane, Salha

2017

http://hdl.handle.net/10539/22737

Mamane, Salha (2017) Wishart laws on convex cones, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/22737>

Downloaded from a DSpace repository. https://dspace.org/

WISHART LAWS ON CONVEX CONES

A PhD thesis

by

Salha Mamane

Supervisor: Piotr Graczyk Co-supervisor: Montaz Ali





School of Statistics and Actuarial Science
University of the Witwatersrand
Johannesburg, South Africa
and
Laboratoire Angevin de REcherche en MAthématiques (LAREMA)
Université d'Angers, France

January 25, 2017

Presentation of the work

This work is done as part of the requirements of a joint thesis at the University of Angers (France) and at the University of the Witwatersrand, Johannesburg (South Africa). It has resulted in:

1. the article (Graczyk et al., 2016b) written in collaboration with Piotr Graczyk and Hideyuki Ishi;

2. the presentation "Variance function of Wishart exponential families in graphical models " at the *Summer School on Mathematical Methods of Statistics* in Angers in June 2016;

3. the article (Graczyk et al., 2016c) written in collaboration with Piotr Graczyk, Hideyuki Ishi and Hiroyuki Ochiai;

4. the presentation "Wishart exponential families in path graphs models " at the *Séminaire triangulaire Probabilités et Statistique* in Le Mans (France) in June 2015;

5. the article (Graczyk and Mamane, 2015) published in collaboration with Piotr Graczyk;

6. the presentation "Fisher Information and Exponential Families Parametrised by a Segment of Means" at the *Séminaire tournant Probabilités et Statistique*, in Poitiers (France) in January 2014.

Summary

In high dimensional settings, Wishart distributions defined within the framework of graphical models are of particular importance.

A graphical model for a random vector $(X_1, \ldots, X_n)$ is a family $\{f_\theta : \theta \in \Theta\}$ of probability distributions such that each f_θ satisfies a set of conditional independence relations encoded in graph G: each random variable is represented by a node and the absence of an edge between two nodes symbolizes conditional independence of the corresponding random variables given the remaining random variables. For a Gaussian graphical model, with covariance matrix Σ and concentration matrix $K = \Sigma^{-1}$, the conditional independence constraints are equivalent to $K_{ij} = 0$ for all non-adjacent vertices i and j. This implies that the entries of Σ corresponding to non-adjacent vertices are not free parameters and the model can be alternatively parametrized by an incomplete matrix with entries corresponding to non-adjacent vertices left out. In the framework of Gaussian graphical models, Wishart distributions are therefore defined on two alternative restrictions of the cone of symmetric positive definite matrices: the cone P_G of symmetric positive definite matrices x satisfying $x_{ij} = 0$ for all non-adjacent vertices i and j and its dual cone Q_G. The existing construction of Wishart exponential families on Q_G and P_G used two different approaches for homogeneous and non-homogeneous cones and as a result, provides two different presentations of the distributions on the two classes of cones. Also, it does not specify the set of parameters of these families for non-homogeneous graphs. Only a conjecture was made about this parameter set.

In this thesis, we propose some parameter parsimonious models which are of great importance in high dimensional data analysis. We first present a background of classical Wishart distributions and multiparameter Wishart distributions in terms of the canonical measures of cones. Then, we provide a harmonious construction of Wishart exponential families in nearest neighbour interaction graphical models, in other terms governed by the graph A_n : $\overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$. Our simple method is based on analysis on convex cones compared to existent work which relies more on graph theory. The focus is on nearest neighbours interactions graphical models which have the advantage of being relatively simple while including all particular cases of interest such as the univariate case, a symmetric cone case, a non-symmetric homogeneous cone case and an infinite number of non-homogeneous cone cases. We derive the Laplace transforms of the Riesz generating measures. Next, we define the Wishart distributions and explicitly determine their classical objects such as the Wishart densities, the Laplace transforms and the mean functions. The Wishart distributions on Q_{A_n} are constructed as the exponential family generated from the gamma function on Q_{A_n}. The Wishart distributions on P_{A_n} are then constructed as the Diaconis-Ylvisaker conjugate family for the exponential family of Wishart distributions on Q_{A_n}. For Wishart distributions on Q_{A_n}, explicit formulas for the inverse mean map and the variance function are derived. Later, the methods of construction of Wishart laws introduced in this thesis are used to solve the Letac-Massam Conjecture on the set of parameters of type I Wishart distributions on Q_{A_n}. Finally, we introduce and study exponential families of distributions parametrized by a segment of means with an emphasis on their Fisher information. This class of models will be useful in high-dimensional data analysis, particularly when one is hesitating between two parameter values. We derive the mean function, the variance function and the Fisher information of the model. We also propose some estimators and explore their properties. The particular cases of Gaussian and Wishart exponential families parametrized by a segment of means are examined.

Résumé

En analyse multivariée de données de grande dimension, les lois de Wishart définies dans le contexte des modèles graphiques revêtent une importance particulière. Un modèle graphique pour un vecteur aléatoire $(X_1, \ldots, X_n)$ est une famille $\{f_\theta : \theta \in \Theta\}$ de lois de probabilité satisfaisant chacune un ensemble de relations d'indépendances conditionnelles représentées par un graphe G : chaque variable aléatoire est représentée par un sommet et l'absence d'une arête entre deux sommets symbolise l'indépendance conditionnelle des variables correspondantes étant données les autres variables. Ce mariage entre la théorie de la probabilité et la théorie des graphes assure une représentation modulaire et parcimonieuse en paramètres de la loi jointe des variables du modèle, permettant ainsi l'estimation des paramètres avec une taille d'échantillon raisonnable et un calcul plus efficient des lois marginales a posteriori. Pour un modèle graphique Gaussien avec une matrice de covariance Σ et une matrice de précision $K = \Sigma^{-1}$, les relations d'indépendances conditionnelles sont équivalentes à $K_{ij} = 0$, pour tous sommets non adjacents i et j. Cela implique que les éléments de la matrice Σ, correspondant à une paire de sommets non adjacents, ne sont pas des paramètres libres. Le modèle peut donc de manière alternative être paramétré par une matrice incomplète dont les éléments correspondant à une paire de sommets non adjacents sont omis. Dans le contexte des modèles graphiques Gaussiens, les lois de Wishart sont par conséquent définies sur des restrictions du cône des matrices symétriques définies positives : le cône P_G des matrices symétriques définies positives x satisfaisant $x_{ij} = 0$, pour tous sommets i et j non adjacents, et son cône dual Q_G. Lorsque

ces cônes sont non-homogènes, la construction existante de lois de Wishart sur les cônes Q_G et P_G utilise deux méthodes différentes pour les graphes homogènes et les graphes non homogènes résultant ainsi à deux formulations différentes des lois de Wishart sur les deux classes de cônes. De plus, elle ne spécifie pas entièrement l'ensemble des valeurs possibles pour les paramètres. Seule une conjecture sur cet ensemble est fournie.

Dans cette thèse nous proposons des modèles parcimonieux en paramètres qui sont de grande utilité en analyse de données de grande dimension. Nous rappelons d'abord les lois de Wishart classiques et multiparamètres présentées du point de vue des mesures canoniques des cônes. Puis, nous présentons une construction harmonieuse de familles exponentielles de lois de Wishart sur les cônes P_G et Q_G. Elle se focalise sur les modèles graphiques d'interactions des plus proches voisins qui sont régis par le graphe A_n : $\overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$ et qui présentent l'avantage d'être relativement simples tout en incluant des exemples de tous les cas particuliers intéressants : le cas univarié, un cas d'un cône symétrique, un cas d'un cône homogène non symétrique, et une infinité de cas de cônes non-homogènes. Notre méthode, simple, se fonde sur l'analyse sur les cônes convexes en contraste avec les travaux précédents qui se basent surtout sur la théorie des graphes. Les lois de Wishart sur Q_{A_n} sont définies à travers la fonction gamma sur Q_{A_n} et les lois de Wishart sur P_{A_n} sont definies comme la famille de Diaconis-Ylvisaker conjuguée à la famille des lois de Wishart sur Q_{A_n}. Les objets classiques associés, tels que les mesures génératrices de ces familles exponentielles, les densités, les transformées de Laplace et les fonctions moyennes, sont déterminés. De plus, pour les lois de Wishart sur Q_{A_n}, les formules de la fonction réciproque de la moyenne et la fonction variance sont établies. Ensuite, les méthodes développées sont utilisées pour résoudre la conjecture de Letac-Massam sur l'ensemble des paramètres de la loi de Wishart sur Q_{A_n}.

Cette thèse étudie aussi les sous-modèles paramétrés par un segment $[m_1, m_2]$ dans $\mathcal{M}$, lorsque $(Q_m)_{m\in\mathcal{M}}$ est une famille exponentielle paramétrée par le domaine des moyennes

$\mathcal{M}$. Ces sous-modèles présentent l'avantage d'être parcimonieux en paramètres dans les cas multidimensionnels et sont particulièrement utiles lorsque l'on hésite entre deux possibles valeurs d'un paramètre. L'accent est mis sur les modèles paramétrés par des matrices. La fonction moyenne, la fonction variance, l'information de Fisher et les estimateurs sont déterminés et les cas particuliers des familles exponentielles Gaussiennes et Wishart sont examinés.

Acknowledgements

I am grateful to my supervisor Prof Piotr Graczyk. Merci pour votre disponibilité tout au long du travail de cette thèse.

I am also grateful to Prof Montaz Ali, the co-supervisor of this thesis.

I would like also to thank Prof Coenraad Labuschagne for accepting to be the co-supervisor of this thesis until he moved from Wits to the University of Johannesburg.

A special thanks to Prof Hideyuki Ishi for his collaboration on two of the papers written as part of this thesis.

I wish also to thank Prof Hiroyuki Ochiai for his collaboration on one of the papers written as part of this thesis.

Mes remerciements vont aussi aux Professeurs Gérard Letac, Lioudmila Vostrikova et Loïc Chaumont membres du comité de suivi de cette thèse.

Finally, I would like to gratefully acknowledge the financial support of the French Ministry of Foreign Affairs through the French Embassy in South Africa.

Contents

Presentation of the work i

Summary ii

Résumé iv

Acknowledgements vii

Notations xi

Introduction 1

1.1 Introduction 1

1.2 Objectives of the research 3

1.3 Organization of the work 4

2 Literature Review 5

3 Background 12

3.1 Convex cones 12

3.2 Graphical Models 13

3.2.1 Undirected graphs 13

3.2.2 Conditional independence and graphs 15

3.2.3 Graphical Gaussian models 16

3.3 Exponential families of distributions 18

3.4 The Wishart distribution 19

3.4.1 The gamma and chi-square distributions . . . 19
3.4.2 The Wishart distribution . . . 20
3.4.3 The classical Wishart distribution . . . 21
3.4.4 The multiparameter Wishart distribution . . . 22

4 Wishart exponential families on cones related to nearest neighbours interactions graphs 25

4.1 Introduction . . . 25
4.2 Preliminaries on A_n graphs and related cones . . . 27
4.2.1 Perfect elimination Orders . . . 29
4.2.2 Generalized power functions . . . 31
4.3 Recursive construction of the cones P_{A_n} and Q_{A_n} and changes of variables 33
4.4 Laplace transform of generalized power functions on Q_{A_n} and P_{A_n} . . . 40
4.5 Wishart exponential families on Q_{A_n} . . . 46
4.5.1 Mean and covariance of the Wishart distributions on Q_{A_n} . . . 47
4.5.2 Inverse mean map . . . 48
4.5.3 Variance function . . . 54
4.6 Wishart exponential families on the cone P_{A_n}. . . . 63
4.6.1 Mean and covariance . . . 64
4.7 Relations with the type I and type II Wishart distributions of Letac and Massam (2007) . . . 65
4.8 Appendix . . . 68

5 On the Letac-Massam conjecture 70

5.1 Introduction . . . 70
5.2 Letac-Massam conjecture on Q_{A_n} . . . 71
5.2.1 Letac-Massam conjecture in terms of power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$ . . . 72

5.3 Proof 73
5.4 Generalized Letac-Massam conjecture 80
5.5 Discussions on (Ben-David and Rajaratnam, 2014) 80

6 Fisher Information and Exponential Families Parametrized by a Segment of Means 83

6.1 Introduction 83
6.2 Preliminaries 85
6.3 Fisher information of Gaussian and Wishart families parametrized by a segment of means 88
6.3.1 Exponential families of Gaussian distributions 90
6.3.2 Exponential families of Wishart distributions 96
6.3.3 Applications to estimation of the mean in exponential families parametrized by a segment of means 101

Conclusion 105

References 107

Notations

In this work, unless otherwise stated,

S_n	is the space of symmetric positive definite $n \times n$ matrices;
S_n^+	is the cone of positive definite $n \times n$ matrices;
$\overline{S_n^+}$	is the cone of positive semidefinite $n \times n$ matrices;
$\overline{E}$	is the closure of the set E;
$A > 0$	means the matrix A is positive definite;
$\mathrm{Tr}(A)$	is the trace of the matrix A;
${}^{\mathsf{t}}A$ or A^T	denotes the transpose of the matrix A;
$\langle x, y \rangle$	is the inner product of x and y;
$x \otimes y$	is the Kronecker product of x and y;
$\det(A)$ or $\lvert A \rvert$	is the determinant of matrix A;
$f(x)$	is the density function of the random variable X;
$f(x\lvert y)$	is the conditional density of X given Y;
G	is a graph with set of vertices V with cardinality n and set of edges $\mathcal{E}$;
$\mathcal{C}$	is the set of cliques of G;
$\mathcal{S}$	is the set of minimal separators of G;
A_n	is the graph $1 - 2 - \cdots - n$;
Z_G	is the space of symmetric matrices y such that $y_{ij} = 0$ for all non-adjacent vertices (i, j) in the graph G;

Z_G^* is the space of incomplete symmetric matrices x with entries x_{ij} missing for all non-adjacent vertices (i, j) in the graph G;

P_G is the convex cone $Z_G \cap S_n^+$;

Q_G is the dual of the cone P_G;

π is the projection of S_n on Z_G^*;

φ is the bijective function P_G to Q_G^* defined by $\varphi(y) = \pi(y^{-1})$;

y_A is the submatrix of y obtained by extracting from y the rows and columns indexed by A;

$\hat{x}$ is the unique positive definite completion of $x \in Q_G$ such that $\hat{x}^{-1} \in P_G$;

$(z_A)^0$ is the matrix obtained from z_A by filling up the entries corresponding to $V \backslash A$ with zero entries: $(z_A)_{ij}^0 = \begin{cases} z_{ij} & \text{if} \quad i, j \in A \subset V, \\ 0 & \text{otherwise.} \end{cases}$

Introduction

1.1 Introduction

The classical Wishart distribution, was first derived by Wishart (1928) as the distribution of the maximum likelihood estimator of the covariance matrix of the multivariate normal distribution. It is a matrix variate generalization of the gamma distribution. In high dimensional settings, Wishart distributions defined within the framework of graphical models are of particular importance.

A graphical model for a random vector $X = (X_1, \ldots, X_n)$ is a family $\{f_\theta : \theta \in \Theta\}$ of probability distributions such that each f_θ satisfies a set of conditional independence relations encoded in graph: each random variable is represented by a node and the absence of an edge between two nodes represents conditional independence between the corresponding random variables given the remaining random variables. This marriage between Probability Theory and Graph Theory provides a parameter parsimonious and modular representation of the joint distribution of the random variables of the model, thereby allowing estimation of model parameters with a reasonable amount of data and a more effective computation of marginal posterior distributions.

Graphical models encompass as particular cases well-known statistical models such as naive Bayes models, state-space models, Markov and hidden Markov models and some particular hierarchical log-linear models. They are also applied to regression analysis (Dobra et al., 2010), longitudinal data analysis (Smith et al., 2009), spatial statistics (Irvine and Gitelman, 2011) or time series analysis (Avventi et al., 2013), (Dahlhaus, 2000), (Songsiri

et al., 2010) or (Eichler, 2012). It should, however, be noted that not all models that use a graph representation with some nodes and some edges are graphical models. Examples of such models include neural networks although these models can sometimes be reformulated as graphical models as explained by Jordan et al. (1999).

Graphical models find real-world applications to complex stochastic systems where they provide a powerful tool for modelling high dimensional multivariate distributions by only specifying the direct interactions between variables but succeeding in capturing all the complexity of the system. Graphical models are intuitive and easy to interpret which facilitates communications between subject-area experts and statisticians. This explains the wide range of applications which go from genetics (Lauritzen and Sheehan, 2003) to computer vision for self-driving cars (Oliver and Pentland, 2000) through finance (Abdelwahab et al., 2008), (Carvalho and West, 2007) or (Sewart and Whittaker, 1998), social science (Berrington et al., 2008), medical science (Caputo et al., 2003) or (Gather et al., 2002), image processing (Murphy et al., 2003), climate science (Callies et al., 2003) and environmental science (Irvine and Gitelman, 2011).

Graphical models are therefore extensively used in statistics, machine learning and artificial intelligence and the theory is developed by both statisticians and computer scientists. According to Koller and Friedman (2009), the rich development of the field is ensured by this synergy between statisticians and computer scientists, and the close and continuous interaction between theory and practice.

For a Gaussian graphical model, with covariance matrix Σ and concentration matrix $K = \Sigma^{-1}$, the conditional independence constraints are equivalent to $K_{ij} = 0$ for all non-adjacent vertices i and j. This implies that the entries of Σ corresponding to non-adjacent vertices are not free parameters and the model can be alternatively parametrized by an incomplete matrix with entries corresponding to non-adjacent vertices left out. In the framework of Gaussian graphical models, Wishart distributions are therefore defined on two alternative restrictions of the cone of symmetric positive definite matrices: the

cone P_G of symmetric positive definite matrices x satisfying $x_{ij} = 0$ for all non-adjacent vertices i and j and its dual cone Q_G.

1.2 Objectives of the research

This research will address two main problems:

1. The existing construction of Wishart exponential families on Q_G and P_G does not fully specify the set of parameters of these families for non-homogeneous graphs. Also, the construction used two different techniques for homogeneous and non-homogeneous graphs and as a result provides two different presentations of the parameter range for homogeneous and non-homogeneous graphs.

   The first objective is to provide an alternative construction of exponential families of Wishart distributions on the cones P_G and Q_G, a construction that fully specifies the shape parameters set and gives a unique description of this set for both homogeneous and non-homogeneous graphs. The focus will be on nearest neighbour interaction graphical models, in other terms governed by the graph A_n : $\overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$, which have the advantage of being relatively simple while including all particular cases of interest such as the univariate case, a symmetric cone case, a non-symmetric homogeneous cone case and an infinite number of non-homogeneous cones cases. The conditional independence relations encoded in such graph are of the form: $X_i \perp X_j | (X_k)_{k \neq i,j}$, for all $|i - j| > 1$.

2. Let $(Q_m)_{m \in \mathcal{M}}$ be a natural or general exponential family on $\mathbb{R}^d$ parametrized by the means domain $\mathcal{M}$. Let m_1 and m_2 be two points in $\mathcal{M}$.

   The submodel $(Q_{\theta m_1 + (1-\theta) m_2})_{\theta \in [0,1]}$ parametrized by the segment $[m_1, m_2]$ in $\mathcal{M}$ presents the advantage of being parameter parsimonious in high dimensional settings. This model will be particularly useful in practical situations when hesitating

between two parameter estimates m_1 and m_2 or in sequential data collection, when an updated estimate of a parameter largely differs from the previous estimate.

The second objective is the study of the submodel $(Q_{\theta m_1+(1-\theta)m_2})_{\theta\in[0,1]}$. The emphasis will be on models with a matrix parameter. The mean function, the variance function, the Fisher information and estimators will be derived, and the particular cases of Gaussian and Wishart exponential families parametrized by a segment of means will be examined.

1.3 Organization of the work

In Chapter 2, a literature review on graphical models and Wishart distributions is presented. Chapter 3 introduces some important background concepts and results on convex cones, graphical models, exponential families of distributions and Wishart distributions. Chapter 4 presents a novel construction of exponential families of Wishart distributions on P_{A_n} and Q_{A_n}. Chapter 5 answers the Letac-Massam conjecture on the set of parameters of type I Wishart distributions on Q_{A_n}. Chapter 6 is devoted to exponential families of distributions parametrized by a segment of means with a strong emphasis on their Fisher information. The work is concluded with a discussion on the work done and perspectives of future extensions.

Chapter 2

LITERATURE REVIEW

The origins of graphical models are explained in Koller and Friedman (2009), Edwards (2000) or Lauritzen (1996). These origins are traced back to Gibbs (1902) in the area of statistical physics, Wright (1921, 1934) in the area of genetics and to Wold (1954) and Blalock (1971) in economical and social science. But modern statistical graphical models genuinely started with Darroch et al. (1980) who, building on the work of Goodman (1970) on the analysis of contingency tables, introduced undirected graphical models for contingency tables as a special subsclass of hierarchical log-linear models with a more efficient parameter estimation and an intuitive interpretation in terms of conditional independence. It should however be noted that Dempster (1972) introduced a model which is essentially the graphical Gaussian model although it does not explicitly use a graphical representation. It was a Gaussian model with prescribed zeros in the concentration matrix for which he derived the maximum likelihood estimator $\hat{\Sigma}$ of the covariance matrix Σ. $\hat{\Sigma}$ is the positive definite matrix whose inverse has the same pattern of zeros as Σ^{-1} and agrees with the empirical covariance matrix for all pairs of indices corresponding to non-zero elements of the concentration matrix. He also proposed an iterative method for model selection and parameter estimation. Graphical Gaussian models are also known as covariance selection models (Dempster, 1972). They are also called concentration graph models in contrast to covariance graph models (Khare and Rajaratnam, 2011)) which ex-

ploit the pattern of zeros in the covariance matrix, thus reflecting marginal independence instead of conditional independence. Wermuth (1976) made the analogy between models for contingency tables and covariance selection models. She showed that both models are based on the definition of the pairwise independence structure and proposed log-likelihood ratio test statistics for model selection. But it was Speed and Kiiveri (1986) who formally associated an undirected graph to a covariance selection model.

Graphical models are broadly classified into two main groups: directed graphical models, also referred to as Bayesian networks or belief networks, which use directed edges between the nodes in the graph representing the statistical model, and undirected graphical models, also referred to as Markov networks or Markov random fields, which use undirected edges between the nodes of the graph representing the statistical model. Graphical chain models (Whittaker, 1990), (Cox and Wermuth, 1996), (Edwards, 2000) unify undirected and directed graph models. They are extensions of graphical models that allow for partially ordered data such as panel data. From subject-matter knowledge, the variables are partitioned into an ordered list of blocks; dependence relationships between variables within the same box are represented by undirected edges while dependence relationships between variables in different boxes are represented by directed edges.

The most natural class of graphical models to use in practice depends on whether the relationships of variables are symmetric like with spatial data (in which case undirected graphical models are more natural) or assymmetric (in which case directed graphical models are more natural). Very often, undirected models can be also equivalently represented as a directed model and conversely. But there are sets of conditional independence relations that can be encoded with either a directed or an undirected graph but not with the other. Conditional independence is easier to check on undirected graphs while model parameters are easier to interpret in directed graphical models. Directed graph models have also the potential of causal interpretation. Graphical models can be used with discrete or continuous variables or a mix of discrete and continuous variables.

The graphical modelling process consists of two steps: the selection of the graph structure which can be built from an expert opinion or learnt from data and the estimation of parameter values. Graphical model selection methods are virtually similar to model selection methods in regression models; including or not an explanatory variable is replaced by including or not an edge between two nodes. The methods, therefore include log-likelihood ratio tests, Akaike information criterion (AIC), Bayesian information criterion (BIC), graphical lasso (Friedman et al., 2008),(Banerjee et al., 2008). Inference methods in graphical methods include exact methods like the junction tree algorithm and approximation methods such as Markov chain Monte Carlo and variational methods (Wainwright and Jordan, 2008).

Missing data in graphical models were dealt with by Lauritzen (1995). Graphical models for mixed discrete and continuous variables were introduced by Lauritzen and Wermuth (1989). Decomposable graphical models are particularly important as they yield closed form maximum likelihood estimators.

Various theoretical aspects of graphical models have been extensively studied. More details can be found in monographs dedicated to the subject which include (Lauritzen, 1996), (Edwards, 2000), (Whittaker, 1990) and (Højsgaard et al., 2012). Other books have an extensive treatment of the subject; these include (Bishop, 2006), (Koller and Friedman, 2009), (Hastie et al., 2009), (Murphy, 2012).

The scope of applications of graphical models is very wide. Applications of graphical models in Artificial Intelligence started with Lauritzen and Spiegelhalter (1988).

Graphical models for time series are sometimes called dynamic graphical models. Avventi et al. (2013) applied graphical models to a zero-mean stationary Gaussian vector stochastic process.

Application of graphical models to regression analysis is illustrated by Dobra et al. (2010) who proposed a method of variables selection in regression analysis using undirected graphical Gaussian models and applied the method to the prediction of macroeco-

nomic growth.

In social science, Berrington et al. (2008) used a graphical chain model to study women's gender role attitudes and changes in their participation in labour force. The graphical modelling allowed the authors to go beyond previous research in being able to simultaneously studying impacts of gender role attitudes on changes in participation in labour force and also impacts of changes in participation in labour force on gender role attitudes.

In medical science, Caputo et al. (2003) used a graphical chain model to investigate the interactions between the determinants of undernutrition in Benin. Mohamed et al. (1998) used a chain graph model to investigate infant mortality and its determinants in Malaysia.

In environmental science, Irvine and Gitelman (2011) applied a directed graphical Gaussian model to stream health data in a study of the effects of urban land use on terrestrial life stages of insects.

Applications of graphical models in Finance are diverse. Carvalho and West (2007) applied a dynamic graphical model to the study of the interdependence of some financial markets and to stock portfolio selection. They constructed a dynamic matrix variate graphical model by adding a graphical model component to the state-space structure of matrix-variate dynamic linear models. The model captures the dependence structure between the time series but also the change of this dependence structure over time.

Sewart and Whittaker (1998) and Hand et al. (1997) used a graphical model for credit-scoring.

Carvalho and West (2007) and Carvalho et al. (2007) explored international currency portfolio selection and exchange rates prediction using a graphical model.

The classical Wishart distribution, was first derived by Wishart (1928) as the distribution of the maximum likelihood estimator of the covariance matrix of the multivariate normal distribution. It can therefore be viewed as a generalization of the gamma distribution

defined on the set of positive real numbers to a distribution defined on the set of positive definite matrices. A good treatment on the classical Wishart distribution can be found in (Muirhead, 2005) or (Eaton, 2007).

One important characterization of these classical Wishart distributions, defined on the cone S_n^+ of symmetric positive definite matrices, is as the natural exponential family generated by measures μ_p with a Laplace transform defined on S_n^+ by $\mathcal{L}_{\mu_p}(\theta) = (\det(\theta))^{-p}$ for p belonging to the Gindikin set $\{\frac{1}{2}, 1, \frac{3}{2}, \ldots, \frac{n-1}{2}\} \cup]\frac{n-1}{2}, \infty[$. The measures μ_p are called Riesz measures.

The non-central Wishart distribution is a natural generalization of Wishart distribution; it is defined as the distribution of $W = Y_1 Y_1^T + \ldots + Y_n Y_n^T$, when the random vector $(Y_1, \ldots, Y_n)$ follows a multivariate normal distribution with a non-zero mean. The construction of the exponential family of non-central Wishart distributions is presented in (Letac and Massam, 2008).

Multi-parameter Riesz measures are obtained by generalizing the real power of the determinant of θ to a product of powers of the principal minors of θ. Multi-parameter Wishart distributions are then obtained as the exponential family of distributions generated by these multi-parameter Riesz measures. Multi-parameter Wishart distributions are also called Riesz distributions by Hassairi and Lajmi (2001) and Boutouria and Hassairi (2009), who studied these distributions on symmetric and homogeneous cones, respectively.

Another important generalization of the Wishart distribution in connection with graphical models was introduced by Dawid and Lauritzen (1993). They showed that the unique distribution p that is Markov over a graph G and has consistent clique marginals p_C, $C \in \mathcal{C}$ has the form $(\prod_{C \in \mathcal{C}} p_C)/(\prod_{S \in \mathcal{S}} p_S)$, where $\mathcal{C}$ is the set of cliques of G, and $\mathcal{S}$ is the set of minimal separators of G. They introduced the hyper Markov property and showed that the maximum likelihood estimator of the parameter of an exponential family of distributions Markov over G is hyper Markov. They also proved that hyper Markov laws with respect

to a graph G are conjugate priors for the sampling family of distributions Markov over the graph G. Dawid and Lauritzen (1993) considered the example of zero mean graphical Gaussian models where they called the distribution of the maximum likelihood estimator of the covariance matrix the hyper Wishart distribution. They also introduced the hyper inverse Wishart distribution as the unique hyper Markov law corresponding to a consistent specification of (classical) inverse Wishart distributions for the cliques marginals and showed that it is the conjugate prior of the covariance matrix in the Bayesian analysis of graphical Gaussian models.

Roverato (2000) derived the distribution of the concentration matrix $K = \Sigma^{-1}$ when $\pi(\Sigma)$ follows a hyper inverse Wishart distribution and called it the G-conditional Wishart distribution.

Andersson and Wojnar (2004) generalized the Wishart distribution from the cone of positive definite matrices to a general homogeneous cone.

In the framework of graphical Gaussian models, Letac and Massam (2007) constructed two classes of multi-parameter Wishart distributions on the cones Q_G and P_G associated to a decomposable graph G and called them type I and type II Wishart distributions. They showed that type I and type II Wishart distributions generalize the hyper Wishart distribution and the G-conditional Wishart distribution respectively. Type I and type II Wishart distributions were constructed as exponential families of distributions generated by measures μ_G and ν_G which are some kind of Markov combination of the measure $|x|^p|x|^{-(n+1)/2}1_{S_n^+(x)}dx$ that generates the exponential family of classical Wishart distributions. They introduced the inverse type I Wishart distribution as the distribution of the 'inverse' $Y = \hat{X}^{-1}$ of a random variable X following the type I Wishart distribution and derived its density function. They also introduced the inverse type II Wishart distribution as the distribution of the 'inverse' $X = \pi(Y^{-1})$ of a random variable Y following the type II Wishart distribution and derived its density function. When the cones Q_G and P_G are homogeneous the parameter set of type I and type II Wishart distributions is fully speci-

fied but for non-homogeneous cones Q_G and P_G, only a subset of the parameter set of the distributions is specified. Letac and Massam (2007) made a conjecture about this parameter set but recent work by Ben-David and Rajaratnam (2014) suggests that this conjecture may not hold for some decomposable non-homogeneous graphs.

Andersson and Klein (2010) proposed another construction of Wishart distributions on decomposable graphs that generalizes type I Wishart distributions. The construction relies on the representation of a decomposable undirected graph as an acyclic mixed graph.

Graczyk and Ishi (2014) showed how Wishart distributions can be constructed on convex cones via quadratic maps as in the classical case. They defined Riesz measures associated to a quadratic map as the image of the Lebesgue measure by that quadratic map and derived the Wishart distributions as the exponential families generated by these Riesz measures.

The main application of the classical Wishart distribution is as a model for covariance matrices, thus its pervasive use in multivariate stochastic volatility models. For example, Philipov and Glickman (2006) used a model in which asset returns follow a multivariate normal distribution with a time dependent concentration matrix Σ_t^{-1} which follows a Wishart distribution with a time dependent scale parameter S_{t-1}.

The Wishart distribution also occurs in Wishart processes as the distribution of the time-marginals. Wishart processes, first introduced by Bru (1989, 1991), are solutions of the matrix stochastic differential equation $dX_t = X_0 + \sqrt{X_t}dB_t + dB_t^T\sqrt{X_t} + \alpha I dt$, where $\alpha > 0$ and B_t is a Brownian matrix.

When a graphical model is used to exploit sparsity in a multivariate stochastic model, the hyper Wishart and G-Wishart distributions come in naturally. This is illustrated in (Carvalho and West, 2007), where a hyper inverse Wishart distribution is used as the distribution of the covariance matrix in a matrix dynamic linear model combined with a graphical model that exploits the sparsity in the cross-sectional concentration matrix.

Chapter 3

BACKGROUND

3.1 Convex cones

In this section, some important concepts and results on convex cones are recalled. More details can be found in (Faraut and Korányi, 1994).

Consider a Euclidean space H.

A subset Ω of H is said to be a convex cone if for all $x, y \in \Omega$ and $\lambda_1, \lambda_2 > 0$, we have $\lambda_1 x + \lambda_2 y \in \Omega$.

The (open) dual cone of an open convex cone Ω is defined as

$$\Omega^* = \{y \in H : \langle x, y\rangle > 0, \forall x \in \overline{\Omega}\backslash\{0\}\}. \tag{3.1}$$

The cone Ω is said to be self-dual if $\Omega = \Omega^*$.

Let $GL(H)$ be the general linear group of H, that is, the group of bijective linear maps on H. The automorphism group $G(\Omega)$ of an open convex cone is defined by

$$G(\Omega) = \{g \in GL(H) : g\Omega = \Omega\}. \tag{3.2}$$

The cone Ω is said to be homogeneous if $G(\Omega)$ acts transitively on Ω, that is, for all $x, y \in \Omega$ there exists $g \in G(\Omega)$ such that $y = g(x)$.

The cone Ω is said to be symmetric if it is homogeneous and self-dual. For example, S_n^+, the set of positive definite symmetric $n \times n$ matrices is a symmetric cone of the space S_n

of symmetric $n \times n$ matrices. $G(S_n^+)$ is the set of linear transformations on S_n^+ of the form

$$\rho(A) : S_n^+ \longrightarrow S_n^+, \quad x \mapsto \rho(A)x = AxA^T, \quad \text{where } A \text{ is an invertible matrix.}$$

The cone S_n^+ is homogeneous. Indeed for any $x, y \in S_n^+$, we can write $x = aa^T$ and $y = bb^T$ where a and b are invertible lower triangular matrices and let $A = ba^{-1}$; then $\rho(A)(x) = ba^{-1}aa^T(ba^{-1})^T = y$.

The characteristic function φ_Ω of a cone Ω is defined as

$$\varphi_\Omega(x) = \int_{\Omega^*} e^{-\langle x,y\rangle} dy, \quad \text{for all } x \in \Omega. \tag{3.3}$$

The measure $\varphi_\Omega(x)dx$ is called the canonical measure of the cone Ω. It is $G(\Omega)$ invariant; this means that for any measurable function $f : \Omega \to \Omega$ and $g \in G(\Omega)$,

$$\int f \circ g(x)\varphi_\Omega(x)dx = \int f(x)\varphi_\Omega(x)dx. \tag{3.4}$$

3.2 Graphical Models

In this section, some important concepts and results on undirected graphical models are recalled. More details can be found in (Lauritzen, 1996),(Letac, 2014), (Edwards, 2000) and (Koller and Friedman, 2009). In the thesis, it becomes clear that using the "cliques-separators" approach in the theory of Riesz measures and Wishart laws on graphical cones is not the natural one.

3.2.1 Undirected graphs

An undirected graph is a pair of sets $G = (V, \mathcal{E})$, where V is a finite set and $\mathcal{E}$ is a subset of $\mathcal{P}_2(V)$, the set of all subsets of V with cardinality two. The elements of V are called nodes or vertices and the elements of $\mathcal{E}$ are called edges. If $v_1 = v_2$ or $\{v_1, v_2\} \in \mathcal{E}$, then v_1 and v_2 are said to be adjacent and this is noted $v_1 \sim v_2$. The set of neighbours of a vertex i is defined as $Ne(i) = \{j \in V \backslash \{i\} : j \sim i\}$ and the closure of i is defined as

$Cl(i) = Ne(i) \cup \{i\}$. Graphs are visualized by representing each node by a point and each edge $\{v_1, v_2\}$ by a line with the nodes $\{v_1, v_2\}$ as endpoints.

A graph $G_0 = (V_0, \mathcal{E}_0)$ is said to be a subgraph of a graph $G = (V, \mathcal{E})$ if $V_0 \subset V$ and $\mathcal{E}_0 \subset \mathcal{E}$.

Let U be a subset of V and define $\mathcal{E}_U = \{\{v_1, v_2\} \in \mathcal{E} : v_1 \in U \text{ and } v_2 \in U\}$. The graph $G_U = (U, \mathcal{E}_U)$ is called the subgraph of G induced by U.

A subset $A \subset V$ is said to be complete if all pairs of vertices in A are adjacent. A graph $G = (V, \mathcal{E})$ is said to be complete if V is complete. A complete subset C of V is said to be a clique if it is not strictly contained in another complete subset of V. The set of all cliques is denoted by $\mathcal{C}$.

A path of length n between two vertices α and β is a subgraph $\alpha = v_0 \sim \ldots \sim v_n = \beta$ of G. A graph is said to be connected if there is a path between every pair of vertices. A cycle of length n is a path $\alpha = v_0 \sim \ldots \sim v_n = \alpha$. A cycle $\alpha = v_0 \sim \ldots \sim v_n = \alpha$ is said to have a chord if there exists $0 < i < n$ and $j \notin \{i-1, i+1\}$ such that $v_i \sim v_j$. A tree is a connected graph with no cycles.

A subset S of V is said to separate a subset A of V from a subset B of V if every path between a vertex in A and a vertex in B contains a vertex in S. A subset S of V is said to be a minimal separator of A and B if it is a separator of A and B and no subset of it is a separator of A and B.

Definition 3.2.1. Consider an ordering $C'_1 < \ldots < C'_k$ of the cliques of G. Let $H'_1 = C'_1$ and for all $2 \leqslant j \leqslant k$, let $H'_j = H'_{j-1} \cup C'_j$ and $S'_j = H'_{j-1} \cap C'_j$.

The ordering $C'_1 < \ldots < C'_k$ is a perfect order of cliques if for all $2 \leqslant j \leqslant k$, there exists $i \leqslant j-1$ such that $S'_j \subset C'_i$.

The S'_j are minimal separators of G. The number of j such that $S = S'_j$ is called the multiplicity of the separator S and will be denoted by $\lambda(S)$.

Definitions 3.2.2.

1. A graph with no chordless cycle of length greater than three is called a decomposable graph.

2. A graph G is said to be homogeneous if it is decomposable and does not contain the graph A_4 as an induced subgraph.

The nearest neighbours interactions graph $A_n : \overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$, is complete for $n \leqslant 2$, homogeneous for $n \leqslant 3$, and non-homogeneous for $n \geqslant 4$. Its sets of cliques and minimal separators are respectively $\mathcal{C} = \{\{i, i+1\} : 1 \leqslant i \leqslant n-1\}$ and $\mathcal{S} = \{\{i\} : 2 \leqslant i \leqslant n-1\}$.

3.2.2 Conditional independence and graphs

Definition 3.2.3. Two random variables X and Y are said to be conditionally independent given a random variable Z if the conditional density function of (X, Y) given Z factorizes as $f(x, y|z) = f(x|z)f(y|z)$. This will be noted $X \perp Y|Z$.

Consider a graph $G = (V, \mathcal{E})$. Let $(X_v)_{v\in V}$ be a collection of random variables and let A, B, C be three subsets of V. If X_A and X_B are conditionally independent given X_C, we write $X_A \perp X_B|X_C$.

Definition 3.2.4 (Markov properties)**.**

1. $(X_v)_{v\in V}$ is pairwise Markov with respect to G if for all $\alpha, \beta \in V$,
   $\alpha \not\sim \beta \implies X_\alpha \perp X_\beta|X_{V\setminus\{\alpha,\beta\}}$.

2. $(X_v)_{v\in V}$ is local Markov with respect to G if for all $v \in V$, $X_v \perp X_{V\setminus Cl(v)}|X_{Ne(v)}$.

3. $(X_v)_{v\in V}$ is global Markov with respect to G if for all A, B, C non-empty disjoint subsets of V such that C separates A and B, we have $X_A \perp X_B|X_C$.

Definition 3.2.5 (Factorization)**.** A joint density function f of the $(X_v)_{v\in V}$ is said to factorize with respect to G if

$$f(x) = \prod_{i=1}^{k} \psi_i(x_{C_i}),$$

where for all $1 \leqslant i \leqslant k$, the function ψ_i depends on x only through x_{C_i} and $\{C_1 \dots C_k\}$ is the set of cliques of G.

Theorem 3.2.6. *If f factorizes with respect to G, then $(X_v)_{v\in V}$ is global Markov with respect to G.*

Further, if $f(x) > 0$ for all x, then

$$\begin{aligned} f \text{ factorizes with respect to } G &\iff (X_v)_{v\in V} \text{ is global Markov with respect to } G; \\ &\iff (X_v)_{v\in V} \text{ is local Markov with respect to } G; \\ &\iff (X_v)_{v\in V} \text{ is pairwise Markov with respect to } G. \end{aligned}$$

Theorem 3.2.7. *Let $\mathcal{C}$ and $\mathcal{S}$ be the sets of cliques and minimal separators of a graph G. Let $\lambda(S)$ be the multiplicity of a minimal separator S as defined in Definition 3.2.1.*

$$f(x) = \frac{\prod_{C\in\mathcal{C}} f_c(x_C)}{\prod_{S\in\mathcal{S}} \lambda(S) f_S(x_S)}$$

is the unique distribution Markov over G that has the given consistent distributions f_C, $C \in \mathcal{C}$ as its clique marginals.

3.2.3 Graphical Gaussian models

In this paragraph, we recall some results from (Letac and Massam, 2007; Andersson and Klein, 2010).

Consider an n-dimensional Gaussian model $N(0, \Sigma)$ which is Markov over a graph $G = (V, \mathcal{E})$. Let $\mathcal{C}$ be the set of cliques of G. Let Z_G be the space of symmetric matrices y such that $y_{ij} = 0$ for all non-adjacent vertices (i, j).

Let Z_G^* be the space of incomplete symmetric matrices x with entries x_{ij} missing for all non-adjacent vertices (i, j).

Let $\pi : S_n \to Z_G^*$ be the projection of S_n on Z_G^*.

The conditional independence constraints are equivalent to

$$K = \Sigma^{-1} \in P_G := Z_G \cap S_n^+. \tag{3.5}$$

They are also equivalent to $\Sigma_{ij} = \Sigma_{i,V\setminus\{i,j\}}\Sigma^{-1}_{V\setminus\{i,j\},V\setminus\{i,j\}}\Sigma_{V\setminus\{i,j\},j}$ for all non-adjacent vertices i and j, which in turn is equivalent to

$$\pi(\Sigma) \in \{x \in Z_G^* : x_C > 0, \forall\, C \in \mathcal{C}\}. \tag{3.6}$$

$\{x \in Z_G^* : x_C > 0, \forall\, C \in \mathcal{C}\}$ is actually the dual cone of P_G,

$$Q_G = \{x \in Z_G^* : \mathrm{Tr}(xy) > 0, \forall\, y \in \overline{P_G}\setminus\{0\}\}.$$

Indeed, an adaptation for A_n graphs of the general proof given in Letac and Massam (2007) is as follows:

- Proof of $\{x \in Z_G^* : x_C > 0, \forall\, C \in \mathcal{C}\} \subset Q_{A_n}$:

  Let $x \in Q_{A_n}$ and let $y \in \overline{P_{A_n}}\setminus\{0\}$. Since the mapping $P_G^{-1} \to Q_G$, $x \mapsto \pi(x)$ is a bijection, there exists $\hat{x}$ positive definite such that $\hat{x}$ is a completion of x.

  We have $\hat{x}^{1/2}y\hat{x}^{1/2} > 0$, thus $\mathrm{Tr}(xy) = \mathrm{Tr}(\hat{x}y) = \mathrm{Tr}(\hat{x}^{1/2}y\hat{x}^{1/2}) > 0$.

  Therefore, $x \in Q_{A_n}$.

- Proof of $Q_{A_n} \subset \{x \in Z_G^* : x_C > 0, \forall\, C \in \mathcal{C}\}$:

  Let $x \in Q_{A_n}$. Let $i \in \{1, \ldots, n-1\}$ and let $\alpha \in \mathbb{R}^2$. Consider $v \in \mathbb{R}^n$ defined by

  $$v_j = \begin{cases} \alpha_1 & \text{if} \quad j = i \\ \alpha_2 & \text{if} \quad j = i+1 \\ 0 & \text{else} \end{cases}$$

  We have $vv^T \in \overline{P_{A_n}}\setminus\{0\}$. Thus, $v^T x v = \mathrm{Tr}(x\, vv^T) > 0$. But we have $v^T x v = \alpha^T x_{\{i,i+1\}}\alpha$. Therefore, $\alpha^T x_{\{i,i+1\}}\alpha > 0$ and $x \in Q_{A_n}$.

P_G and Q_G are homogeneous (as defined in Section 3.1) if the graph G is homogeneous (as defined in Definition 3.2.2) (Letac and Massam, 2007). Therefore, the cones

P_{A_3} and Q_{A_3} are homogeneous and for all $n \geqslant 4$, the cones P_{A_n} and Q_{A_n} are non-homogeneous.

Theorem 3.2.8. *For a decomposable graph G, for all $x \in Q_G$, there exists a unique completed matrix $\hat{x} \in S_n^+$ such that $\pi(\hat{x}) = x$ and $\hat{x}^{-1} \in P_G$.*

Lauritzen's formula gives

$$\hat{x}^{-1} = \sum_{C \in \mathcal{C}} \left((x_C)^{-1}\right)^0 - \sum_{S \in \mathcal{S}} \lambda(S) \left((x_S)^{-1}\right)^0, \tag{3.7}$$

where $\lambda(S)$ is the multiplicity of the separator S as defined in Definition 3.2.1 and $(z_A)^0$ denotes the matrix obtained from z_A by filling up the entries corresponding to $V \backslash A$ with zero entries.

φ will denote the bijective map from P_G to Q_G such that $\varphi(y) = \pi(y^{-1})$.

3.3 Exponential families of distributions

In this section, some important concepts and results on exponential families of distribution are introduced. Exponential families of distributions are extensively used in statistics and intensively studied. More details can be found in (Barndorff-Nielsen, 1978), (Brown, 1986), (Lehmann and Casella, 1998) or (Lehmann and Romano, 2005). The presentation given here essentially follows (Letac and Casalis, 2000).

Consider a real vector space E and its dual space E^* (the space of linear forms on E). Let $\langle\ \rangle : E^* \times E \longrightarrow \mathbb{R}$, $(s, x) \longmapsto \langle s, x \rangle$ be the canonical bilinear form on $E^* \times E$. Let μ be a positive measure on E.

The moment generating function of μ is the map $\mathcal{M}_\mu : E^* \to [0, \infty]$ defined by $\mathcal{M}_\mu(s) = \int_E e^{\langle s, x \rangle} \mu(dx)$. Let $K(s) = \ln(\mathcal{M}_\mu(s))$.

The natural exponential family generated by μ is the family of probability distributions defined by $\{P_s(dx; \mu) = e^{\langle s, x \rangle - K(s)} \mu(dx) : s \in S\}$, where S is the interior of the set $\{s : K(s) < \infty\}$, assumed to be non-empty.

Consider a σ-finite measurable space $(\Omega, \mathcal{A}, \nu)$. The general exponential family generated by the measure ν and the map $T : \Omega \to E$ is the family

$$\{P_s(T, \nu) = \exp\{\langle s\,,\, T\rangle - K(s)\}d\nu :\ \ s \in S\}. \tag{3.8}$$

If μ is the image of the measure ν by T on E, the family $P_s(\mu)$ is called the natural exponential family associated with the general exponential family $P_s(T, \nu)$. The mean function of the exponential family is the map m defined by $m(s) = \mathbb{E}_s(T) = K'(s)$ which is an analytic diffeomorphism from S^0 (the interior of S) to the open set $\mathcal{M} = m(S^0)$. $\mathcal{M}$ is called the domain of the means of the family.

The map $\psi : \mathcal{M} \to S^0$, $m \mapsto \psi(m) = (K')^{-1}(m)$ is called the inverse mean map. The exponential family, parametrized by the domain of the means M is given by

$$Q_m(\mu)(dx) = e^{\langle \psi(m)\,,\, x\rangle - K(\psi(m))}\mu(dx), \quad m \in \mathcal{M}. \tag{3.9}$$

The variance function of the family is defined by $V(m) = K''(\psi(m)) := v(\psi(m))$. The variance function is very important as it characterizes the exponential family. Indeed, if the variance functions V_{F_1} and V_{F_2} of two natural exponential families F_1 and F_2 coincide on a non-empty subset of the intersection $\mathcal{M}_{F_1} \cap \mathcal{M}_{F_2}$ of the domains of means, then $F_1 = F_2$ (Letac, 1989).

3.4 The Wishart distribution

In this section, we introduce the definition and some important properties of the classical Wishart distribution. More details can be found in Eaton (2007), Faraut and Korányi (1994) and Muirhead (2005). The presentation here is based on the characteristic function and the canonical measure of a cone.

3.4.1 The gamma and chi-square distributions

The gamma function is defined for $s > 0$ as

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx. \tag{3.10}$$

Observing that the characteristic function of the cone $(0, \infty)$ is

$$\varphi_{\mathbb{R}^+}(x) = \int_0^\infty e^{-xy} dy = x^{-1},$$

the gamma function can be rewritten as

$$\Gamma(s) = \int_0^\infty e^{-x} x^s \varphi_{\mathbb{R}^+}(x) dx. \tag{3.11}$$

Using the invariance (formula (3.4)) of the canonical measure $\varphi_{\mathbb{R}^+}(x)dx$, we obtain

$$\int_0^\infty e^{-xy} (xy)^s \varphi_{\mathbb{R}^+}(x) dx = \Gamma(s). \tag{3.12}$$

Therefore,

$$\gamma(s, y; x)dx = \frac{1}{\Gamma(s)} e^{-xy} (xy)^s \varphi_{\mathbb{R}^+}(x) 1_{\mathbb{R}^+}(x) dx \tag{3.13}$$

is a probability density function. The corresponding probability distribution is called the Gamma distribution $G(s, y)$. The Laplace transform of $G(s, y)$ is, for all $\theta > 0$, given by

$$\mathcal{L}_{G(s,y)}(\theta) = \int_0^\infty \frac{1}{\Gamma(s)} e^{-x(\theta+y)} (xy)^s \varphi_{\mathbb{R}^+}(x) dx = \left(1 + \tfrac{\theta}{y}\right)^{-s}. \tag{3.14}$$

The mean and variance of the gamma distribution are respectively $\frac{s}{y}$ and $\frac{s}{y^2}$.

The probability distribution $G\left(\frac{k}{2}, \frac{1}{2}\right)$ is called the chi-square distribution. It is the distribution of $\sum_{i=1}^k X_i^2$ when the random variables X_i, $i = 1, \ldots, k$ are independent and follow a standard normal distribution.

3.4.2 The Wishart distribution

The gamma function on S_n^+ is defined for all $s > \frac{n-1}{2}$ as

$$\Gamma_{S_n^+}(s) := \int_{S_n^+} e^{-\operatorname{Tr}(x)} |x|^{s-\frac{n+1}{2}} dx = \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(s - \frac{i-1}{2}\right). \tag{3.15}$$

If Y is an $n \times n$ symmetric matrix and A an $n \times n$ invertible matrix, the Jacobian of the transformation $Z = AY\,{}^{\mathsf{t}}A$ is $|A|^{n+1}$ (Mathai et al., 2012). Therefore, the characteristic function of the cone S_n^+ is

$$\varphi_{S_n^+}(x) = \int_{S_n^+} e^{-\operatorname{Tr}(xy)} dy = \int_{S_n^+} e^{-\operatorname{Tr}(x^{1/2}yx^{1/2})} dy = |x|^{-\frac{n+1}{2}} \int_{S_n^+} e^{-\operatorname{Tr}(z)} dz = c|x|^{-\frac{n+1}{2}},$$

where c is a constant.

In the sequel, we will omit the constant and write $\varphi_{S_n^+}(x) = |x|^{-\frac{n+1}{2}}$. The gamma function on S_n^+ can thus be rewritten as

$$\Gamma_{S_n^+}(s) := \int_{S_n^+} e^{-\operatorname{Tr}(x)} |x|^s \varphi_{S_n^+}(x) dx.$$

Using the invariance (formula (3.4)) of the canonical measure $\varphi_{S_n^+}(x)dx$, we obtain

$$\int_{S_n^+} e^{-\operatorname{Tr}(xy)} |xy|^s \varphi_{S_n^+}(x) dx = \int_{S_n^+} e^{-\operatorname{Tr}(y^{1/2}xy^{1/2})} |y^{1/2}xy^{1/2}|^s \varphi_{S_n^+}(x) dx = \Gamma_{S_n^+}(s).$$

Therefore,

$$\gamma_{S_n^+}(s, y; x)dx = \frac{1}{\Gamma_{S_n^+}(s)} e^{-\operatorname{Tr}(xy)} |xy|^s \varphi_{S_n^+}(x) 1_{S_n^+}(x) dx$$

is a probability density function; the corresponding probability distribution is the Wishart distribution $W_n(s, y)$. It is sometimes also called the matrix-variate gamma distribution (Mathai et al., 2012).

The Laplace transform of $W_n(s, y)$ is $\mathcal{L}_{W_n(s,y)}$ given by

$$\mathcal{L}_{W_n(s,y)}(\theta) = |I_n + y^{-1}\theta|^{-s}. \tag{3.16}$$

Rewriting the density function as

$$\gamma_{S_n^+}(s,y;x)dx = e^{\langle -x,y\rangle - \ln|y|^{-s}}\nu(dx),$$

where $\nu(dx) = \frac{|x|^s}{\Gamma_{S_n^+}(s)}\varphi_{S_n^+}(x)1_{S_n^+}dx$, we note that for a fixed $s > 0$, $W_n(s,y)$ is an exponential family of distributions generated by the measure ν with a Laplace transform $\mathcal{L}_\nu(y) = |y|^{-s}$, for all $y \in S_n^+$.

The mean and variance of $W_n(s,y)$ are respectively sy^{-1} and $s(y^{-1}\otimes y^{-1})$.

3.4.3 The classical Wishart distribution

The reparametrization $W_n^c(s,y) = W_n\left(\frac{s}{2}, \frac{1}{2}y^{-1}\right)$ is similar to the chi-square distribution. It is called the classical Wishart distribution. If $X_1,\ldots,X_d$ is a random sample from an n-dimensional normal distribution with mean zero and covariance Σ, then $W = X_1X_1^T + \ldots + X_dX_d^T$ follows the classical Wishart distribution $W_n^c(d,\Sigma)$. The maximum likelihood estimator of Σ follows the classical Wishart distribution $W_n^c(d, \frac{1}{d}\Sigma)$.

Defining the classical Wishart distribution $W_n^c(s,y)$ by its Laplace transform $\mathcal{L}_{W_n^c(s,y)}(\theta) = \det(I_n + 2\,\Sigma\theta)^{-s/2}$ extends the set of possible values of the parameter s to the so-called Gindikin's set $\{1,2,3,\ldots,n-1\}\cup]n-1,\infty[$. The mean and covariance of $W_n^c(s,\Sigma)$ are respectively $s\,\Sigma$ and $2s\,\Sigma\otimes\Sigma$. If $s > n-1$, $W_n^c(s,\Sigma)$ has a density given by

$$f(x) = \frac{1}{\Gamma_{S_n^+}(s)}\frac{|x|^{\frac{s-n-1}{2}}}{2^{\frac{ns}{2}}|\Sigma|^{\frac{s}{2}}}e^{-\frac{1}{2}\operatorname{Tr}(\Sigma^{-1}x)}. \tag{3.17}$$

For $s \in \{1,2,3,\ldots,n-1\}$, the classical Wishart distribution is singular; it has no density function with respect to the Lebesgue measure and is concentrated on a subspace of positive semidefinite matrices of rank less than n.

3.4.4 The multiparameter Wishart distribution

For all $y \in S_n^+$ and $1 \leqslant i \leqslant n$, the matrix $y_{\{1:i\}}$ is the upper left submatrix of y of size $i \times i$, and $y_{\{i:n\}}$ is the lower right submatrix of size $(n-i+1) \times (n-i+1)$.

The generalized power functions $\Delta_{\underline{s}}$ and $\delta_{\underline{s}}$ on S_n^+ are defined by

$$\Delta_{\underline{s}}(y) = y_{11}^{s_1 - s_2} |y_{1:2}|^{s_2 - s_3} \dots |y_{1:n-1}|^{s_{n-1} - s_n} |y_{1:n}|^{s_n} \tag{3.18}$$

and

$$\delta_{\underline{s}}(y) = |y_{1:n}|^{s_1} |y_{2:n}|^{s_2 - s_1} \dots |y_{n-1:n}|^{s_{n-1} - s_{n-2}} y_{nn}^{s_n - s_{n-1}}. \tag{3.19}$$

The two power functions are known to be related by the property $\Delta_{-\underline{s}}(y) = \delta_{\underline{s}}(y^{-1})$ (Faraut and Korányi, 1994, Proposition VII.1.5), where the notation $\Delta_{-\underline{s}}$ means $-\underline{s}$ replaces $\underline{s}$.

The multiparameter Gamma function on S_n^+ is defined, for $\underline{s} \in \mathbb{R}^n$ such that $s_i > \frac{i-1}{2}$, as

$$\Gamma_{S_n^+}(\underline{s}) := \int_{S_n^+} e^{-\operatorname{Tr}(x)} \Delta_{\underline{s}}(x) \varphi_{S_n^+}(x) dx = \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^{n} \Gamma\left(s_i - \frac{i-1}{2}\right). \tag{3.20}$$

Using the invariance of the measure $\varphi_{S_n^+}(x)dx$ by linear automorphisms of S_n^+ and writing $y \in S_n^+$ as $y = {}^t bb$ and $z = bx\,{}^t b$ with b a lower triangular matrix with positive diagonal elements, we get

$x = b^{-1} z\, {}^t b^{-1}$, $\quad \operatorname{Tr}(xy) = \operatorname{Tr}(x\, {}^t bb) = \operatorname{Tr}(bx\, {}^t b)$,

$\Delta_{\underline{s}}(x) = \Delta_{\underline{s}}(b^{-1} z\, {}^t b^{-1}) = \Delta_{\underline{s}}(b^{-1}\, {}^t b^{-1}) \Delta_{\underline{s}}(z) = \Delta_{\underline{s}}(y^{-1}) \Delta_{\underline{s}}(bx\, {}^t b)$ (Faraut and Korányi, 1994, Proposition VI.3.10) and

$$\int_{S_n^+} e^{-\operatorname{Tr}(xy)} \Delta_{\underline{s}}(x) \varphi_{S_n^+}(x) dx = \Delta_{\underline{s}}(y^{-1}) \int_{S_n^+} e^{-\operatorname{Tr}(bxb^T)} \Delta_{\underline{s}}(bxb^T) \varphi_{S_n^+}(x) dx = \delta_{-\underline{s}}(y) \Gamma_{S_n^+}(\underline{s}).$$

Therefore,

$$\omega(\underline{s}, y; x) dx = \frac{1}{\Gamma_{S_n^+}(\underline{s})} e^{-\operatorname{Tr}(xy)} \Delta_{\underline{s}}(x) \delta_{\underline{s}}(y) \varphi_{S_n^+}(x) dx \tag{3.21}$$

is a probability density function on S_n^+; the corresponding probability distribution is called the multiparameter Wishart distribution $W_n(\underline{s}, y)$.

Moreover, for a fixed $\underline{s}$, $W_n(\underline{s}, y)$ is an exponential family of distributions generated by the measure $R_{\underline{s}}(dx) = \frac{1}{\Gamma_{S_n^+}(\underline{s})}\Delta_{\underline{s}}(x)\varphi_{S_n^+}(x)dx$ which has a Laplace transform given by $\mathcal{L}_{R_{\underline{s}}}(y) = \delta_{-\underline{s}}(y)$.

The Laplace transform of $W_n(\underline{s}, y)$ is $\mathcal{L}_{W_n(\underline{s},y)}(\theta) = \frac{\delta_{-\underline{s}}(y+\theta)}{\delta_{-\underline{s}}(y)}$.

Remark 3.4.1. *Consider $y \in S_n^+$ and $\underline{s} \in \mathbb{R}^n$.*

Let us define $y^* = Ry\,{}^{\mathsf{t}}R$ and $\underline{s}^*$ by $y_{ij}^* = y_{n-i+1,n-j+1}$ and $s_i^* = s_{n-i+1}$.

Then, it is easy to see that $\mathrm{Tr}(y^*) = \mathrm{Tr}(y)$, $\delta_{\underline{s}^*}(y^*) = \Delta_{\underline{s}}(y)$, $\varphi_{S_n^+}(y^*) = \varphi_{S_n^+}(y)$ and $dy^* = dy$.

Remark 3.4.2. *We could have alternatively defined the multi-parameter gamma function as*

$$\tilde{\Gamma}_{S_n^+}(\underline{s}) := \int_{S_n^+} e^{-\mathrm{Tr}(x)}\delta_{\underline{s}}(x)\varphi_{S_n^+}(x)dx, \tag{3.22}$$

Using Remark 3.4.1, we obtain

$$\tilde{\Gamma}_{S_n^+}(\underline{s}) = \int_{S_n^+} e^{-\mathrm{Tr}(x^*)}\Delta_{\underline{s}^*}(x^*)\varphi_{S_n^+}(x^*)dx^* = \Gamma_{S_n^+}(\underline{s}^*).$$

A similar reasoning as above gives, for fixed $\underline{s}$, the exponential family of multiparameter Wishart distributions $\tilde{W}_n(\underline{s}, y)$ with density function

$$\tilde{\omega}(\underline{s}, y; x)dx = \frac{1}{\tilde{\Gamma}_{S_n^+}(\underline{s})}e^{-\mathrm{Tr}(xy)}\delta_{\underline{s}}(x)\Delta_{\underline{s}}(y)\varphi_{S_n^+}(x)dx. \tag{3.23}$$

$\tilde{W}_n(\underline{s}, y)$ is generated by the measure $\tilde{R}_{\underline{s}}(dx) = \frac{1}{\tilde{\Gamma}_{S_n^+}(\underline{s})}\delta_{\underline{s}}(x)\varphi_{S_n^+}(x)dx$ which has a Laplace transform given by $\mathcal{L}_{\tilde{R}_{\underline{s}}}(y) = \Delta_{-\underline{s}}(y)$.

The Laplace transform of $\tilde{W}_n(\underline{s}, y)$ is $\mathcal{L}_{\tilde{W}_n(\underline{s},y)}(\theta) = \frac{\Delta_{-\underline{s}}(y+\theta)}{\Delta_{-\underline{s}}(y)} = \frac{\delta_{-\underline{s}^*}(y^*+\theta^*)}{\delta_{-\underline{s}^*}(y^*)}$.

Therefore, $W_n(\underline{s}, y) = \tilde{W}_n(\underline{s}^*, y^*)$.

Chapter 4

WISHART EXPONENTIAL FAMILIES ON CONES RELATED TO NEAREST NEIGHBOURS INTERACTIONS GRAPHS

4.1 Introduction

The classical Wishart distribution was first derived by Wishart (1928) as the distribution of the maximum likelihood estimator of the covariance matrix of the multivariate normal distribution. Applications in estimation and other practical aspects of Wishart distributions are intensely studied, cf. Sugiura and Konno (1988); Tsukuma and Konno (2006); Konno (2007, 2009); Kuriki and Numata (2010).

In the framework of graphical Gaussian models, the distribution of the maximum likelihood estimator of $\pi(\Sigma)$, where π denotes the canonical projection onto Q_G, was derived by Dawid and Lauritzen (1993), who called it the hyper Wishart distribution. Dawid and Lauritzen (1993) also considered the hyper inverse Wishart distribution which is defined on Q_G as the Diaconis-Ylvisaker conjugate prior distribution for $\pi(\Sigma)$, and Roverato (2000) derived the so-called G-Wishart distribution on P_G, that is, the distribution of the concentration matrix $K = \Sigma^{-1}$ when $\pi(\Sigma)$ follows the hyper inverse Wishart distribution. Letac and Massam (2007) constructed two classes of multi-parameter Wishart distributions on

the cones Q_G and P_G associated to a decomposable graph G and called them type I and type II Wishart distributions, respectively. They are more flexible because they have multiple shape parameters. In fact, the type I and type II Wishart distributions generalize the hyper Wishart distribution and the G-Wishart distribution respectively.

The Wishart exponential families introduced and studied in this thesis include the type I and type II Wishart distributions constructed by Letac and Massam (2007) on the cones Q_G and P_G associated to nearest neighbours interactions graphs. Our methods, which are new and different from methods of articles cited above, simplify in a significant way the Wishart theory for graphical models. This chapter makes it clear that using the "cliques-separators" approach in the theory of Riesz measures and Wishart laws on graphical cones is not the natural one. Our approach allows the derivation of results which are technically challenging until now.

The methods introduced in this chapter allow to solve the Letac-Massam Conjecture on the cones Q_{A_n} in Chapter 5. Together with the results presented in this chapter we achieve in this way the complete study of all classical objects of an exponential family for the Wishart natural exponential families on the cones Q_{A_n}.

Some of the results of our research may be extended to cones related to all decomposable graphs (work in progress). Many of them are however specific for the cones Q_{A_n} and P_{A_n} (indexation of Riesz and Wishart measures by $M = 1, \ldots, n$, Letac-Massam Conjecture, Inverse Mean Map, Variance function).

This chapter is orgarnized as follows. Sections 4.2, 4.3 and 4.4 provide the main tools in order to define and to study the Wishart natural exponential families on the cones Q_{A_n} and P_{A_n}. In Section 4.2, useful notions of perfect elimination orders $\prec$ on A_n and of generalized power functions $\delta_{\underline{s}}^{\prec}$ and $\Delta_{\underline{s}}^{\prec}$, $\underline{s} \in \mathbb{R}^n$ will be introduced on the cones Q_{A_n} and P_{A_n} respectively. In Theorem 4.2.9, a classical relation between the power functions $\delta_{\underline{s}}^{\prec}$ and $\Delta_{-\underline{s}}^{\prec}$ is proved as well as the dependence of $\delta_{\underline{s}}^{\prec}$ and $\Delta_{\underline{s}}^{\prec}$ on the maximal element M

of $\prec$ only. Thus, in the sequel, only generalized power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$ appear. Next important tool of analysis of Wishart exponential families are recursive construction of the cones P_G and Q_G and corresponding changes of variables. They are introduced and studied in Section 4.3, and are immediately applied in Section 4.4 in order to compute the Laplace transform of generalized power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$ (Theorems 4.4.1 and 4.4.2).

In Section 4.5, Wishart natural exponential families on the cones Q_{A_n} are defined, and all their classical objects are explicitly determined, beginning with the Riesz generating measures, Wishart densities, Laplace transform, mean and covariance. In Theorem 4.5.4 and Corollary 4.5.7, an explicit formula for the inverse mean map is proved. It provides an infinite number of versions of Lauritzen formulas for bijections between the cones Q_G and P_G. In Section 4.5.3, two explicit formulas are given for the variance function of a Wishart family. The formula of Theorem 4.5.15 is surprisingly simple and similar to the case of the symmetric cone S_n^+.

Section 4.6 is on Wishart natural exponential families on the cones P_{A_n} and follows a similar scheme as Section 4.5, however the inverse mean map and variance function are not available on the cones P_{A_n}. The analysis on these cones is more difficult.

Finally, in Section 4.7, we establish the relations between the Wishart natural exponential families defined and studied in this chapter and the type I and type II Wishart distributions from Letac and Massam (2007). Our methods give a simple proof of the formulas for Laplace transforms of type I and type II Wishart distributions from Letac and Massam (2007).

4.2 Preliminaries on A_n graphs and related cones

In this section, we study properties of nearest neighbours interactions graphs that will be important in the theory of Riesz measures and Wishart distributions on the cones related

to these graphs. In particular, we characterize all the eliminating orders of vertices and we introduce generalized power functions related to such orders. We show that they only depend on the maximal element $M \in \{1, \dots, n\}$ of the order.

Recall that an undirected graph is a pair $G = (V, \mathcal{E})$, where V is a finite set and $\mathcal{E}$ is a subset of $\mathcal{P}_2(V)$, the set of all subsets of $\mathcal{E}$ with cardinality two. For convenience, we introduce a subset $E \subset V \times V$ defined by $E := \{(v, v') : v \sim v'\} \cup \{(v, v) : v \in V\}$. The graph with $V = \{v_1, v_2, \dots, v_n\}$ and $\mathcal{E} = \{\{v_j,\ v_{j+1}\} : 1 \leqslant j \leqslant n-1\}$ is denoted by A_n and represented as $\overset{1}{\bullet} - \overset{2}{\bullet} - \cdots - \overset{n}{\bullet}$. An n-dimensional Gaussian model $(X_v)_{v \in V}$ is said to be Markov with respect to a graph G if for any $(v,\ v') \notin E$, the random variables X_v and $X_{v'}$ are conditionally independent given all the other variables. The conditional independence relations encoded in A_n graph are of the form: $X_{v_i} \perp X_{v_j} | (X_{v_k})_{k \neq i,j}$, for all $|i - j| > 1$. Thus, A_n graphs correspond to nearest neighbours interactions models. In what follows, we often denote the vertex v_i by i.

For a graph G, let $Z_G \subset S_n$ be the vector space consisting of $y \in S_n$ such that $y_{ij} = 0$ if $(i\ ,\ j) \notin E$. Let $I_G = Z_G^*$ be the dual vector space with respect to the scalar product $\langle y, \eta \rangle = \mathrm{Tr}(y\eta) = \sum_{(i,j) \in E} y_{ij}\eta_{ij}$, $y \in Z_G$, $\eta \in I_G$. In the statistical literature, the vector space I_G is commonly realized as the space of $n \times n$ symmetric matrices η, in which only the elements η_{ij}, $(i, j) \in E$, are given. We adopt this realisation of I_G in this thesis.

If $I \subset V$, we denote by y_I the submatrix of $y \in Z_G$ obtained by extracting from y the lines and the columns indexed by I. The same notation is used for $\eta \in I_G$. Let P_G be the cone defined by $P_G = \{y \in Z_G : y > 0\}$, and $Q_G \subset I_G$ the dual cone of P_G, that is, $Q_G = \{\eta \in I_G :\ \forall y \in \overline{P_G} \backslash \{0\},\ \langle y, \eta \rangle > 0\}$. A Gaussian vector model $(X_v)_{v \in V}$ is Markov with respect to G if and only if the concentration matrix $K = \Sigma^{-1}$ belongs to P_G.

When $G = A_n$, the cone Q_G is described as $Q_G = \{\eta \in I_G : \eta_{\{i,i+1\}} > 0,\ i = 1, \dots, n-1\}$. Let π be the projection of S_n onto I_G, $x \mapsto \eta$ such that $\eta_{ij} = x_{ij}$ if $(i, j) \in E$. Then it is known (cf. Letac and Massam (2007); Andersson and Klein (2010))

that the mapping $P_G \longrightarrow Q_G, y \longmapsto \pi(y^{-1})$ is a bijection.

4.2.1 Perfect elimination Orders

Different orders of vertices $v_1, v_2, \ldots, v_n$ should be considered in order to have a harmonious theory of Riesz and Wishart distributions on the cones related to A_n graphs. The orders that will be important in this work are called *perfect elimination orders of vertices* and will be presented now.

Definition 4.2.1. Consider a graph $G = (V, \mathcal{E})$ and a total strict order $\prec$ of the vertices of G. The set of future neighbours of a vertex v is defined as $v^+ = \{w \in V : v \prec w \text{ and } v \sim w\}$. The set of all predecessors of a vertex $v \in V$ with respect to $\prec$ is defined as $v^- = \{u \in V : u \prec v\}$.

Definition 4.2.2. A total strict order $\prec$ of the vertices of a graph G is said to be a perfect elimination order if v^+ is complete for all $v \in V$.

Example 4.2.3. *For the graph* $A_3 : 1-2-3$*, the orders* $1 \prec 2 \prec 3$, $1 \prec 3 \prec 2$, $3 \prec 2 \prec 1$ *and* $3 \prec 1 \prec 2$ *are perfect elimination orders while* $2 \prec 1 \prec 3$ *and* $2 \prec 3 \prec 1$ *are not.*

Theorem 4.2.4 (Grone et al. (1984); Paulsen et al. (1989); Roverato (2000))**.**

There exists a perfect vertex elimination order of the vertices of the graph $G = (V, E)$ *if and only if* G *is decomposable.*

Also, provided a perfect vertex elimination order of vertices is used, the upper triangular matrix in the Choleski decomposition of the concentration matrix in a graphical model has the same pattern of zeros as the concentration matrix.

Next, we present a characterization of the perfect elimination orders in the case of the graph A_n. An algorithm that generates all perfect elimination orders for a general graph is given by (Chandran et al., 2003).

Proposition 4.2.5. *Consider the graph $A_n : 1 - 2 - 3 - \cdots - n$. A total strict order $\prec$ is a perfect elimination order if and only if there exits $M \in \{1, \ldots, n\}$ such that $\prec$ is an intertwining of the two sequences $1 \prec \ldots \prec M$ and $n \prec \ldots \prec M$.*

In particular M is the maximal element, $\prec$ is the ordinary order if $M = n$ and the reversed one if $M = 1$. Finally, there are 2^{n-1} possible perfect elimination orders on the graph A_n.

Proof. Consider a perfect elimination order $\prec$ on A_n. The only vertices of A_n having only one neighbour are the two exterior vertices. If a vertex v with two neighbours $v - 1$ and $v + 1$ were minimal for $\prec$, then the set v^+ would contain these vertices and would not be complete. Thus, the minimal element of $\prec$ is one of the exterior vertices 1 or n of the graph. Without loss of generality, let us say the order starts with 1. It follows from Definition 4.2.2 that a perfect elimination order without its minimal element forms again a perfect elimination order on the graph A_{n-1} obtained from A_n by suppressing 1 or n. The element following 1 may be 2 or n. This recursive argument proves that in a perfect elimination order the sequences $1 \prec 2 \ldots \prec M$ and $n \prec n - 1 \prec \ldots \prec M$ must appear intertwined. We also see that we construct in this way 2^{n-1} different orders.

Conversely, if an order $\prec$ on A_n is obtained by intertwining of the sequences $1 \prec 2 \ldots \prec M$ and $n \prec n - 1 \prec \ldots \prec M$, it follows that the sets v^+ of future neighbours of v are singletons or empty (for $v = M$). Thus $\prec$ is a perfect elimination order. □

Example 4.2.6. *Consider $n = 4$ and $M = 3$. By intertwining of the sequences $1 \prec 2 \prec 3$ and $4 \prec 3$ we obtain the perfect elimination orders*

$$4 \prec 1 \prec 2 \prec 3; \;\; 1 \prec 4 \prec 2 \prec 3; \;\; 1 \prec 2 \prec 4 \prec 3.$$

Similarly, for $M = 1$ we get the perfect elimination order $4 \prec 3 \prec 2 \prec 1$; for $M = 2$ we get three perfect elimination orders $4 \prec 3 \prec 1 \prec 2$, $4 \prec 1 \prec 3 \prec 2$, $1 \prec 4 \prec 3 \prec 2$, and for $M = 4$ we have the usual order $1 < 2 < 3 < 4$.

Thus there are 8 perfect elimination orders and 16 non-perfect elimination orders of the four vertices of A_4 graph.

4.2.2 Generalized power functions

In this section, we define and study generalized power functions on the cones P_G and Q_G. First let us introduce some useful notations. For $1 \leqslant i \leqslant j \leqslant n$, let $\{i : j\} \subset V$ be the set of $a \in V$ for which $i \leqslant a \leqslant j$. Then, for $y \in Z_G$ and $1 \leqslant i \leqslant n$, the matrix $y_{\{1:i\}}$ is the upper left submatrix of y of size $(i \times i$, and $y_{\{i:n\}}$ is the lower right submatrix of size $(n-i+1) \times (n-i+1)$. Recall that on the cone S_n^+, the generalized power functions are $\Delta_{\underline{s}}(y) = \prod_{i=1}^n |y_{\{1:i\}}|^{s_i - s_{i+1}}$ and $\delta_{\underline{s}}(y) = \prod_{i=1}^n |y_{\{i:n\}}|^{s_i - s_{i-1}}$, with $s_0 = s_{n+1} = 0$.

Definition 4.2.7. For $\underline{s} \in \mathbb{R}^n$, setting $\det y_\varnothing = 1 = \det \eta_\varnothing$, we define

$$\Delta_{\underline{s}}^{\prec}(y) := \prod_{v \in V} \Big(\frac{\det y_{\{v\} \cup v^-}}{\det y_{v^-}} \Big)^{s_v} \qquad (y \in P_G), \tag{4.1}$$

$$\delta_{\underline{s}}^{\prec}(\eta) := \prod_{v \in V} \Big(\frac{\det \eta_{\{v\} \cup v^+}}{\det \eta_{v^+}} \Big)^{s_v} \qquad (\eta \in Q_G). \tag{4.2}$$

Note that Definition 4.2.7 applied to the complete graph with the usual order $1 < \ldots < n$ gives $\Delta_{\underline{s}}$ and $\delta_{\underline{s}}$. For any $\underline{s}$ the following formula $\delta_{\underline{s}}(y^{-1}) = \Delta_{-\underline{s}}(y)$ holds (Faraut and Korányi, 1994). In Theorem 4.2.9 we find an analogous formula in the case of the cones P_G and Q_G.

We will see in Theorem 4.2.9 that on the cones related to the graphs A_n, different order-depending power functions $\Delta_{\underline{s}}^{\prec}$ and $\delta_{\underline{s}}^{\prec}$ defined in Definition 4.2.7 may be expressed in terms of explicit "M-power functions" $\Delta_{\underline{s}}^{(M)}$ and $\delta_{\underline{s}}^{(M)}$ that will be defined below. They depend only on the choice of $M \in V$.

Definition 4.2.8. Let $M \in V$, $y \in P_{A_n}$ and $\eta \in Q_{A_n}$. We define the M-power functions $\Delta_{\underline{s}}^{(M)}(y)$ on P_{A_n} and $\delta_{\underline{s}}^{(M)}(x)$ on Q_{A_n} by the following formulas:

$$\Delta^{(M)}_{\underline{s}}(y) = \left(\prod_{i=1}^{M-1} |y_{\{1:i\}}|^{s_i - s_{i+1}}\right) |y|^{s_M} \left(\prod_{i=M+1}^{n} |y_{\{i:n\}}|^{s_i - s_{i-1}}\right), \tag{4.3}$$

$$\delta^{(M)}_{\underline{s}}(\eta) = \frac{\left(\prod_{i=1}^{M-1} |\eta_{\{i:i+1\}}|^{s_i}\right) \left(\prod_{i=M+1}^{n} |\eta_{\{i-1:i\}}|^{s_i}\right)}{\left(\prod_{i=2}^{M-1} \eta_{ii}^{s_{i-1}}\right) \eta_{MM}^{s_{M-1} - s_M + s_{M+1}} \left(\prod_{i=M+1}^{n-1} \eta_{ii}^{s_{i+1}}\right)}. \tag{4.4}$$

Observe that for $M \in \{1, n\}$ there are $n-1$ factors in the denominator of (4.4), and for $M \in \{2, \dots n-1\}$ there are $n-2$ factors (powers of $\eta_{22} \dots \eta_{n-1,n-1}$).

The main result of this section is the following theorem.

Theorem 4.2.9. *Consider a graph A_n with a perfect elimination order $\prec$. Let M be the maximal element with respect to $\prec$. Then for all $y \in P_{A_n}$, we have*

$$\delta^{\prec}_{\underline{s}}(\pi(y^{-1})) = \Delta^{\prec}_{-\underline{s}}(y) = \Delta^{(M)}_{-\underline{s}}(y). \tag{4.5}$$

The proof of Theorem 4.2.9 is preceded by a series of elementary lemmas.

Lemma 4.2.10. *Let $y \in P_{A_n}$ and $i < j < j+1 < k < m$. The determinant of the submatrix $y_{\{i:j\} \cup \{k:m\}}$ can be factorized as $|y_{\{i:j\} \cup \{k:m\}}| = |y_{\{i:j\}}||y_{\{k:m\}}|$.*

Lemma 4.2.11. *Let $y \in P_{A_n}$ and $\eta = \pi(y^{-1})$. Then for all $i, i+1 \in V$, we have*

$$\left|\eta_{\{i,i+1\}}\right| = |y|^{-1} |y_{V \setminus \{i,i+1\}}|.$$

Proof. We repeatedly use the cofactor formula for an inverse matrix. We use $\eta_{ii} = |y|^{-1}|y_{V\setminus\{i\}}|$ and show that $\eta_{i,i+1} = -y_{i,i+1}|y|^{-1}|y_{V\setminus\{i,i+1\}}|$. It follows that $\left|\eta_{\{i,i+1\}}\right| = |y|^{-2}|y_{V\setminus\{i,i+1\}}| \left[|y_{\{i+1:n\}}||y_{\{1:i\}}| - y^2_{i,i+1}|y_{\{1:i-1\}}||y_{\{i+2:n\}}|\right]$. The last factor in brackets equals $|y|$. □

Proof. (of Theorem 4.2.9)

Part 1: $\delta_{\underline{s}}^{\prec}(\pi(y^{-1})) = \Delta_{-\underline{s}}^{(M)}(y)$. From Proposition 4.2.5, we have

$$i^+ = \begin{cases} \{i+1\} & \text{if } i \leqslant M-1, \\ \varnothing & \text{if } i = M, \\ \{i-1\} & \text{if } i \geqslant M+1. \end{cases}$$

Using $\eta_{ii} = |y|^{-1}|y_{V\setminus\{i\}}|$ with $\eta = \pi(y^{-1})$ and Lemmas 4.2.10 and 4.2.11, we get $\delta_{\underline{s}}^{\prec}(\pi(y^{-1})) = \Delta_{-\underline{s}}^{(M)}(y)$.

Part 2: $\Delta_{\underline{s}}^{\prec}(y) = \Delta_{\underline{s}}^{(M)}(y)$. Let us first consider the perfect elimination order $\prec_M$ given by

$$1 \prec_M 2 \prec_M \ldots \prec_M M-1 \prec_M n \prec_M n-1 \prec_M \ldots \prec_M M+1 \prec_M M. \tag{4.6}$$

Using $\eta_{ii} = |y|^{-1}|y_{V\setminus\{i\}}|$, Lemmas 4.2.10 and 4.2.11 again, we get $\Delta_{\underline{s}}^{\prec_M}(y) = \Delta_{\underline{s}}^{(M)}(y)$. It is easy to see using Proposition 4.2.5 and the factorization from Lemma 4.2.10 that for any other perfect elimination order $\prec$, the factors of $\Delta_{\underline{s}}^{\prec}(y)$ under the powers s_i are exactly the same as for $\prec_M$. Indeed, if $i \leqslant M-1$, let $n-j$ be the largest vertex greater than M such that $n-j \prec i$. Then, the factor under the power s_i is

$$\frac{|y_{\{i\}\cup i^-|}}{|y_{i^-|}} = \frac{|y_{\{1:i\}}||y_{\{n-j:n\}}|}{|y_{\{1:i-1\}}||y_{\{n-j:n\}}|} = \frac{|y_{\{1:i\}}|}{|y_{\{1:i-1\}}|}.$$

A similar argument shows that this is also true for $i = M$ and for $i > M$. □

Corollary 4.2.12. *Let $\prec_1$ and $\prec_2$ be two perfect elimination orders on A_n such that $\max_{\prec_1} V = \max_{\prec_2} V$. Then $\delta_{\underline{s}}^{\prec_1}(\eta) = \delta_{\underline{s}}^{\prec_2}(\eta)$ for all $\eta \in Q_{A_n}$. If $\max_{\prec} V = M$ then we have $\delta_{\underline{s}}^{\prec}(\eta) = \delta_{\underline{s}}^{(M)}(\eta)$.*

4.3 Recursive construction of the cones P_{A_n} and Q_{A_n} and changes of variables

In this section, we introduce very useful recursive constructions of the cones P_{A_n} and Q_{A_n} from the cones $P_{A_{n-1}}$ and $Q_{A_{n-1}}$. There are two variants of them for $A_{n-1} : 2 - \cdots - n$

and $A_{n-1} : 1 - \cdots - (n-1)$. Corresponding changes of variables for integration on P_{A_n} and Q_{A_n} are introduced.

Proposition 4.3.1. *1. For $n \geqslant 2$, let $\Phi_n : \mathbb{R}^+ \times \mathbb{R} \times P_{A_{n-1}} \longrightarrow P_{A_n}$, $(a, b, z) \longmapsto y$ with*

$$y = A(b) \begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & z & \\ 0 & & & \end{pmatrix} {}^{\mathsf{t}}A(b), \qquad A(b) = \begin{pmatrix} 1 & & & \\ b & 1 & & \\ \vdots & & \ddots & \\ 0 & \ldots & 0 & 1 \end{pmatrix},$$

and let $\Psi_n : \mathbb{R}^+ \times \mathbb{R} \times Q_{A_{n-1}} \longrightarrow Q_{A_n}$, $(\alpha, \beta, x) \longmapsto \eta$ with

$$\eta = \pi \left({}^{\mathsf{t}}A(\beta) \begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & x & \\ 0 & & & \end{pmatrix} A(\beta) \right).$$

Then the maps Φ_n and Ψ_n are bijections.

2. Let $\tilde{\Phi}_n : \mathbb{R}^+ \times \mathbb{R} \times P_{A_{n-1}} \longrightarrow P_{A_n}$, $(a, b, z) \longmapsto \tilde{y}$ with

$$\tilde{y} = {}^{\mathsf{t}}B(b) \begin{pmatrix} & & & 0 \\ & z & & \vdots \\ & & & 0 \\ 0 & \cdots & 0 & a \end{pmatrix} B(b), \quad B(b) = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & \ldots & b & 1 \end{pmatrix},$$

and let $\tilde{\Psi}_n : \mathbb{R}^+ \times \mathbb{R} \times Q_{A_{n-1}} \longrightarrow Q_{A_n}$, $(\alpha, \beta, x) \longmapsto \tilde{\eta}$ with

$$\tilde{\eta} = \pi \left(B(\beta) \begin{pmatrix} & & & 0 \\ & x & & \vdots \\ & & & 0 \\ 0 & \cdots & 0 & \alpha \end{pmatrix} {}^{\mathsf{t}}B(\beta) \right).$$

Then the maps $\tilde{\Phi}_n$ and $\tilde{\Psi}_n$ are bijections.

3. *The Jacobians of the changes of variables* $y = \Phi_n(a, b, z)$ *and* $y = \tilde{\Phi}_n(a, b, z)$ *are given by*

$$J_{\Phi_n}(a, b, z) = a, \quad J_{\tilde{\Phi}_n}(a, b, z) = a. \tag{4.7}$$

The Jacobians of the changes of variables $\eta = \Psi_n(\alpha, \beta, x)$ *and* $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$ *are given by*

$$J_{\Psi_n}(\alpha, \beta, x) = x_{22}, \quad J_{\tilde{\Psi}_n}(\alpha, \beta, x) = x_{n-1,n-1}. \tag{4.8}$$

It should be noted that for $\Phi_n(a, b, z)$ *and* $\Psi_n(\alpha, \beta, x)$ *the rows and columns of* z *and* x *are numbered* $2, \ldots, n$ *while for* $\tilde{\Phi}_n(a, b, z)$ *and* $\tilde{\Psi}_n(\alpha, \beta, x)$ *they are numbered* $1, \ldots, n-1$.

Proof. 1. Let $y' = \begin{pmatrix} a & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & z & \\ 0 & & & \end{pmatrix}$ and $\eta' = \begin{pmatrix} \alpha & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & x & \\ 0 & & & \end{pmatrix}$. Then

$$y_{ij} = \begin{cases} ab & \text{if } (i,j) = (1,2) \text{ or } (i,j) = (2,1), \\ ab^2 + z_{22} & \text{if } i = j = 2, \\ y'_{ij} & \text{otherwise.} \end{cases} \tag{4.9}$$

Thus, on the one hand, if $(a, b, z) \in \mathbb{R}^+ \times \mathbb{R} \times P_{A_{n-1}}$, then $y \in Z_{A_n}$. And $z > 0$ implies $y' > 0$ as every principal minor of y' equals a times a principal minor of z. From $y = Ty'\,{}^{\mathrm{t}}T$ with $T = A(b)$, we get $y \in P_{A_n}$. On the other hand, if $y \in P_{A_n}$, we have $a = y_{11} > 0$, $b = \frac{y_{12}}{y_{11}}$, $z_{22} = y_{22} - \frac{y_{12}^2}{y_{11}}$ and $z_{ij} = y_{ij}$ for all $i \neq 2$ and $j \neq 2$. We use the notation $z = (z_{ij})_{2 \leqslant i,j \leqslant n}$. Now, let us show that $z \in P_{A_{n-1}}$. We have $y' = T^{-1}y\,{}^{\mathrm{t}}T^{-1} > 0$. Hence, we have also $z > 0$ since each principal minor of z equals $1/a$ times a principal minor of y'. Therefore, the map Φ_n is indeed a bijection from $\mathbb{R}^+ \times \mathbb{R} \times P_{A_{n-1}}$ onto P_{A_n}.

Let us turn to Ψ_n. The relation between η and η' is given by

$$\eta_{ij} = \begin{cases} \alpha + \beta^2 x_{22} & \text{if } i = j = 1, \\ \beta x_{22} & \text{if } (i,j) = (1,2) \text{ or } (i,j) = (2,1), \\ \eta'_{ij} & \text{otherwise.} \end{cases} \tag{4.10}$$

First we show that if $(\alpha, \beta, x) \in \mathbb{R}^+ \times \mathbb{R} \times Q_{A_{n-1}}$, then $\eta \in I_{A_n}$. Actually, since $x_{\{2,3\}} > 0$, we have $\alpha + \beta^2 x_{22} > 0$ and $\eta_{\{1,2\}} = \begin{pmatrix} \alpha + \beta^2 x_{22} & \beta x_{22} \\ \beta x_{22} & x_{22} \end{pmatrix} > 0$, where we recall that the indices $\{2,3\}$ and $\{1,2\}$ denote sets. On the other hand, if $\eta \in Q_{A_n}$, we have $x_{ij} = \eta_{ij}$ for all $i, j = 2, \ldots, n$. Thus, $\eta \in Q_{A_n}$ implies $x \in Q_{A_{n-1}}$.

2. Let $\tilde{y}' = \begin{pmatrix} & & & 0 \\ & z & & \vdots \\ & & & 0 \\ 0 & \ldots & 0 & a \end{pmatrix}$ and $\tilde{\eta}' = \begin{pmatrix} & & & 0 \\ & x & & \vdots \\ & & & 0 \\ 0 & \ldots & 0 & \alpha \end{pmatrix}$. Then we have

$$\tilde{y}_{ij} = \begin{cases} ab & \text{if } (i,j) = (n-1, n) \text{ or } (i,j) = (n, n-1), \\ ab^2 + z_{n-1,n-1} & \text{if } i = j = n-1, \\ \tilde{y}'_{ij} & \text{otherwise,} \end{cases} \tag{4.11}$$

and

$$\tilde{\eta}_{ij} = \begin{cases} \alpha + \beta^2 x_{n-1,n-1} & \text{if } i = j = n, \\ \beta x_{n-1,n-1} & \text{if } (i,j) = (n-1, n) \text{ or } (i,j) = (n, n-1), \\ \tilde{\eta}'_{ij} & \text{otherwise.} \end{cases} \tag{4.12}$$

Similar reasoning as above shows that $\tilde{\Phi}$ and $\tilde{\Psi}$ are indeed bijections.

3. From (4.9), we have $\frac{\partial y_{11}}{\partial a} = 1$; $\frac{\partial y_{11}}{\partial b} = 0$; and for all i, j, $\frac{\partial y_{11}}{\partial z_{ij}} = 0$;

$\frac{\partial y_{12}}{\partial a} = b$; $\frac{\partial y_{12}}{\partial b} = a$ and for all i, j, $\frac{\partial y_{12}}{\partial z_{ij}} = 0$;

$\frac{\partial y_{22}}{\partial a} = b^2$; $\frac{\partial y_{22}}{\partial b} = 2ab$; $\frac{\partial y_{22}}{\partial z_{22}} = 1$ and for all $(i,j) \neq (2,2)$, $\frac{\partial y_{22}}{\partial z_{ij}} = 0$;

for all $i, j \neq 1, 2$, we have $\frac{\partial y_{ij}}{\partial z_{ij}} = 1$ and $\frac{\partial y_{ij}}{\partial z_{kl}} = 0$ if $(i,j) \neq (k,l)$.

The Jacobian of the change of variable $y = \Phi_n(a, b, z)$ is therefore,

$$J = \begin{vmatrix} \frac{\partial y_{11}}{\partial a} & \frac{\partial y_{11}}{\partial b} & \frac{\partial y_{11}}{\partial z_{11}} & \dots & \frac{\partial y_{11}}{\partial z_{nn}} \\ \frac{\partial y_{12}}{\partial a} & \frac{\partial y_{12}}{\partial b} & \frac{\partial y_{12}}{\partial z_{11}} & \dots & \frac{\partial y_{12}}{\partial z_{nn}} \\ \frac{\partial y_{22}}{\partial a} & \frac{\partial y_{22}}{\partial b} & \frac{\partial y_{22}}{\partial z_{11}} & \dots & \frac{\partial y_{22}}{\partial z_{nn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_{nn}}{\partial a} & \frac{\partial y_{nn}}{\partial \beta} & \frac{\partial y_{nn}}{\partial z_{11}} & \dots & \frac{\partial y_{nn}}{\partial x_{nn}} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 & \dots & \dots & 0 \\ b & a & 0 & 0 & \dots & \dots & 0 \\ b^2 & 2ab & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ 0 & 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & & \dots & 0 & 1 \end{vmatrix} = a.$$

The proof of the second part is similar. □

Example 4.3.2. *In order to obtain the cone P_{A_4} associated to the graph $A_4 : 1-2-3-4$ we can go as follows from the cone P_{A_3} associated to the graph $A_3 : 2-3-4$:*

$$\Phi_4 : \mathbb{R}^+ \times \mathbb{R} \times P_{A_3} \longrightarrow P_{A_4},\ (a, b, z) \longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ b & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & z_{22} & z_{23} & 0 \\ 0 & z_{23} & z_{33} & z_{z34} \\ 0 & 0 & z_{34} & z_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ b & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}^T.$$

Hence, P_{A_4} is the set of matrices of the form $\begin{pmatrix} a & ab & 0 & 0 \\ ab & ab^2 + z_{22} & z_{23} & 0 \\ 0 & z_{23} & z_{33} & z_{34} \\ 0 & 0 & z_{34} & z_{44} \end{pmatrix}$ *with $a > 0$, $b \in \mathbb{R}$*

and $z = \begin{pmatrix} z_{22} & z_{23} & 0 \\ z_{23} & z_{33} & z_{34} \\ 0 & z_{34} & z_{44} \end{pmatrix} \in P_{A_3}$.

We can also obtain P_{A_4} by going from the cone P_{A_3} associated to the graph $1-2-3$; we proceed as follows:

$$\tilde{\Phi}_4 : \mathbb{R}^+ \times \mathbb{R} \times P_{A_3} \longrightarrow P_{A_4},\ (a, b, z) \longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b & 1 \end{pmatrix}^T \begin{pmatrix} z_{11} & z_{12} & 0 & 0 \\ z_{12} & z_{22} & z_{23} & 0 \\ 0 & z_{23} & z_{33} & 0 \\ 0 & 0 & 0 & a \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & b & 1 \end{pmatrix}.$$

Hence, P_{A_4} is the set of matrices of the form $\begin{pmatrix} z_{11} & z_{12} & 0 & 0 \\ z_{12} & z_{22} & z_{23} & 0 \\ 0 & z_{23} & ab^2 + z_{33} & ab \\ 0 & 0 & ab & a \end{pmatrix}$ *with* $a > 0$, $b \in \mathbb{R}$ *and* $z = \begin{pmatrix} z_{11} & z_{12} & 0 \\ z_{12} & z_{22} & z_{23} \\ 0 & z_{23} & z_{33} \end{pmatrix} \in P_{A_3}$.

Lemma 4.3.3.

1. *Let $y = \Phi_n(a, b, z)$ and $\eta = \Psi_n(\alpha, \beta, x)$.*

   Then, for all $M \in \{2, \dots, n\}$, we have

$$\Delta_{\underline{s}}^{(M)}(y) = a^{s_1} \Delta_{(s_2,\dots,s_n)}^{(M)}(z), \tag{4.13}$$

$$\delta_{\underline{s}}^{(M)}(\eta) = \alpha^{s_1} \delta_{(s_2,\dots,s_n)}^{(M)}(x). \tag{4.14}$$

   Let $y = \tilde{\Phi}_n(a, b, z)$ and $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$. Then, for all $M = 1, \dots, n-1$,

$$\Delta_{\underline{s}}^{(M)}(y) = a^{s_n} \Delta_{(s_1,\dots,s_{n-1})}^{(M)}(z), \tag{4.15}$$

$$\delta_{\underline{s}}^{(M)}(\eta) = \alpha^{s_n} \delta_{(s_1,\dots,s_{n-1})}^{(M)}(x). \tag{4.16}$$

2. *Let us define $\varphi_{A_n} : Q_{A_n} \to \mathbb{R}_+$ by $\varphi_{A_1}(\eta) = \eta^{-1}$, and for $n \geqslant 2$*

$$\varphi_{A_n}(\eta) = \prod_{i=1}^{n-1} |\eta_{\{i,i+1\}}|^{-3/2} \prod_{i \neq 1,n} \eta_{ii}. \tag{4.17}$$

   Let $\eta = \Psi_n(\alpha, \beta, x)$ and $\tilde{\eta} = \tilde{\Psi}_n(\alpha, \beta, x)$. Then,

$$\varphi_{A_n}(\eta) = x_{22}^{-1/2} \alpha^{-3/2} \varphi_{A_{n-1}}(x) \tag{4.18}$$

   and

$$\varphi_{A_n}(\tilde{\eta}) = x_{n-1,n-1}^{-1/2} \alpha^{-3/2} \varphi_{A_{n-1}}(x). \tag{4.19}$$

3. *If* $y = \Phi_n(a, b, z)$ *and* $\eta = \Psi_n(\alpha, \beta, x)$*, then*

$$\mathrm{Tr}(y\eta) = a\alpha + ax_{22}(b+\beta)^2 + \mathrm{Tr}(zx). \tag{4.20}$$

If $y = \tilde{\Phi}_n(a, b, z)$ *and* $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$*, then*

$$\mathrm{Tr}(y\eta) = a\alpha + ax_{n-1,n-1}(b+\beta)^2 + \mathrm{Tr}(zx). \tag{4.21}$$

Proof. 1. For $M \geqslant 2$, we have

$$\frac{\Delta^{(M)}_{\underline{s}}(y)}{\Delta^{(M)}_{(s_2,\ldots,s_n)}(z)} = (y_{11})^{s_1-s_2}\left[\prod_{i=2}^{M-1}\left(\frac{|y_{\{1:i\}}|}{|z_{\{2:i\}}|}\right)^{s_i-s_{i+1}}\right]\left(\frac{|y|}{|z|}\right)^{s_M}.$$

Using Lemma 4.8.1 in the appendix, we have $|y_{\{1:i\}}| = a|z_{\{2:i\}}|$. Thus,

$$\frac{\Delta^{(M)}_{\underline{s}}(y)}{\Delta^{(M)}_{(s_2,\ldots,s_n)}(z)} = a^{s_1}.$$

Noting that $a = y_{nn}$, we have for $M = 1, \ldots, n-1$,

$$\begin{aligned}\Delta^{(M)}_{\underline{s}}(\tilde{y}) &= |\tilde{y}|^{s_1}\prod_{i=2}^{n}|\tilde{y}_{\{i:n\}}|^{s_i-s_{i-1}} = a^{s_1}|z|^{s_1}\prod_{i=2}^{n-1}\left(a\,|z_{\{i:n\}}|^{s_i-s_{i-1}}\right)a^{s_n-s_{n-1}}\\ &= a^{s_n}|z|^{s_1}\prod_{i=2}^{n-1}|z_{\{i:n\}}|^{s_i-s_{i-1}} = a^{s_n}\Delta^{(M)}_{(s_1,\ldots,s_{n-1})}(z).\end{aligned}$$

Similarly, we show that $\delta^{(M)}_{\underline{s}}(\eta) = \alpha^{s_1}\delta^{(M)}_{(s_2,\ldots,s_n)}(x)$ for $M \geqslant 2$ and that $\delta^{(M)}_{\underline{s}}(\eta) = \alpha^{s_n}\delta^{(M)}_{\underline{s}}(x)$ for all $M \leqslant n-1$.

2. Let $\eta = \Psi(\alpha, \beta, x)$ and $\tilde{\eta} = \tilde{\Psi}(\alpha, \beta, x)$. For $n = 2$, we have

$$\varphi_{A_2}(\eta) = |\eta_{\{1,2\}}|^{-3/2} = \begin{vmatrix}\alpha+\beta^2 x & \beta x\\ \beta x & x\end{vmatrix}^{-3/2} = \alpha^{-3/2}x^{-3/2}$$
$$= x^{-1/2}\alpha^{-3/2}\varphi_{A_1}(x).$$

For $n > 2$, using (4.10), we have

$$\varphi_{A_n}(\eta) \;=\; \eta_{22}\,|\eta_{\{1,2\}}|^{-3/2}\,\frac{\prod_{i=2}^{n-1}|\eta_{\{i,i+1\}}|^{-3/2}}{\prod_{i=3}^{n-1}\eta_{ii}^{-1}} = x_{22}^{-1/2}\alpha^{-3/2}\varphi_{A_{n-1}}(x).$$

The proof of the second part is analogous.

3. We have

$$\begin{aligned}\mathrm{Tr}(y\eta) = \sum_{i=1}^{n}\sum_{k=1}^{n} y_{ik}\eta_{ki} &= y_{11}\eta_{11} + y_{12}\eta_{21} + y_{21}\eta_{12} + \sum_{i=2}^{n}\sum_{k=2}^{n} y_{ik}\eta_{ki}\\ &= a(\alpha + \beta^2 x_{22}) + 2ab\beta x_{22} + \sum_{i=2}^{n}\sum_{k=2}^{n} y_{ik}\eta_{ki}.\end{aligned}$$

Now observing from (4.9) and (4.10) that

$$\sum_{i=2}^{n}\sum_{k=2}^{n} y_{ik}\eta_{ki} = y_{22}\eta_{22} + \mathrm{Tr}(zx) - z_{22}\eta_{22} = (ab^2 + z_{22})x_{22} + \mathrm{Tr}(zx) - z_{22}\eta_{22}$$

we get

$$\mathrm{Tr}(y\eta) = a\alpha + a\beta^2 x_{22} + 2ab\beta x_{22} + ab^2 x_{22} + \mathrm{Tr}(zx) = a\alpha + ax_{22}(b+\beta)^2 + \mathrm{Tr}(zx).$$

Formula (4.21) is proved similarly.

□

Lemma 4.3.4. *Consider $y \in P_{A_n}$.*

1. *If $y = \Phi_n(a, b, z)$, then $\varphi(z)_{jj} = \varphi(y)_{jj}$ for $j \geqslant 2$.*
2. *If $y = \tilde{\Phi}_n(a, b, z)$, then $\varphi(z)_{jj} = \varphi(y)_{jj}$ for $j \leqslant n-1$.*

Proof. 1. Note that $y = \Phi_n(a, b, z)$ is expressed in the form $T\begin{pmatrix} a & \\ & z \end{pmatrix}{}^{\mathrm{t}}T$, where $T = A(b)$ in Lemma 4.3.3. In general, let A, U, L be $n \times n$ matrices with U upper

triangular and L lower triangular. Then $(UAL)_{\{j:n\}} = U_{\{j:n\}}A_{\{j:n\}}L_{\{j:n\}}$. It follows that $(y^{-1})_{\{j:n\}} = (({}^{\mathsf{t}}T)^{-1})_{\{j:n\}}(z^{-1})_{\{j:n\}}(T^{-1})_{\{j:n\}} = (z^{-1})_{\{j:n\}}$ for $j \geqslant 2$ since $(T^{-1})_{\{j:n\}} = I_{\{j:n\}} = (({}^{\mathsf{t}}T)^{-1})_{\{j:n\}}$. In particular $(y^{-1})_{jj} = (z^{-1})_{jj}$.

2. Similar to the proof of the first part.

□

4.4 Laplace transform of generalized power functions on Q_{A_n} and P_{A_n}

Theorem 4.4.1. *For all $n \geqslant 1$, for all $1 \leqslant M \leqslant n$ and for all $y \in P_{A_n}$, the integral $\int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)}\delta_{\underline{s}}^{(M)}(\eta)\varphi_{A_n}(\eta)d\eta$ converges if and only if $s_i > \frac{1}{2}$ for all $i \neq M$, and $s_M > 0$. In this case, we have*

$$\int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)}\delta_{\underline{s}}^{(M)}(\eta)\varphi_{A_n}(\eta)d\eta = \pi^{(n-1)/2}\Big\{\prod_{i\neq M}\Gamma(s_i - \frac{1}{2})\Big\}\Gamma(s_M)\Delta_{-\underline{s}}^{(M)}(y). \tag{4.22}$$

Proof. We will proceed by induction on the number n of vertices. For $n = 1$, we have the gamma integral that converges if and only if $s > 0$, so that

$$\int_0^\infty e^{-y\eta}\delta_s^{(1)}(\eta)\varphi_{A_1}(\eta)d\eta = \int_0^\infty e^{-y\eta}\eta^{s-1}d\eta = \Gamma(s)y^{-s}.$$

Now assume that the assertion holds for a graph with $n-1$ vertices.

Case $M > 1$. Let $y = \Phi_n(a, b, z)$ and let us make the change of variable $\eta = \Psi_n(\alpha, \beta, x)$. The induction hypothesis gives

$$\begin{aligned}&\int_{Q_{A_{n-1}}} e^{-\operatorname{Tr}(zx)}\delta_{(s_2,\ldots,s_n)}^{(M)}(x)\varphi_{A_{n-1}}(x)dx \\ &\quad = \quad \pi^{(n-2)/2}\Big\{\prod_{i\neq 1,M}\Gamma(s_i - \frac{1}{2})\Big\}\Gamma(s_M)\Delta_{-(s_2,\ldots,s_n)}^{(M)}(z)\end{aligned} \tag{4.23}$$

if and only if $s_i > \frac{1}{2}$ for all $i \neq M$, and $s_M > 0$. By Lemma 3, the change of variable $\eta = \Psi_n(\alpha, \beta, x)$ gives $d\eta = x_{22}d\alpha d\beta dx$. Thus, we have

$$\int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)}\delta_{\underline{s}}^{(M)}(\eta)\varphi_{A_n}(\eta)d\eta$$

$$= \int_0^\infty \int_{-\infty}^\infty \int_{Q_{A_{n-1}}} e^{-(a\alpha + a x_{22}(b+\beta)^2 + \mathrm{Tr}(zx))} \alpha^{s_1 - 3/2} \delta^{(M)}_{(s_2,\dots,s_n)}(x) \varphi_{A_{n-1}}(x) x_{22}^{1/2} d\alpha d\beta dx,$$

where we used parts 3 and 1 of Lemma 4.3.3. Now, using the Gaussian integral

$$\int_{-\infty}^\infty e^{-a x_{22}(b+\beta)^2} d\beta = \pi^{1/2} a^{-1/2} x_{22}^{-1/2}$$

and the gamma integral

$$\int_0^\infty e^{-a\alpha} \alpha^{s_1 - 3/2} d\alpha = a^{-s_1 + 1/2} \Gamma(s_1 - \frac{1}{2}),$$

that is finite if and only if $s_1 > \frac{1}{2}$, we get

$$\begin{aligned} &\int_{Q_{A_n}} e^{-\mathrm{Tr}(y\eta)} \delta^{(M)}_{\underline{s}}(\eta) \varphi_{A_n}(\eta) d\eta \\ &= \pi^{1/2} a^{-s_1} \Gamma(s_1 - \frac{1}{2}) \int_{Q_{A_{n-1}}} e^{-\mathrm{Tr}(zx)} \delta^{(M)}_{(s_2,\dots,s_n)}(x) \varphi_{A_{n-1}}(x) dx. \end{aligned} \tag{4.24}$$

Finally, using Formulas (4.23) and (4.13) completes the proof in the case $M > 1$.

Case $M = 1$.

Let $y = \tilde{\Phi}_n(a, b, z)$ and let us make the change of variable $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$. The induction hypothesis gives

$$\int_{Q_{A_{n-1}}} e^{-\mathrm{Tr}(zx)} \delta^{(1)}_{(s_1,\dots,s_{n-1})}(x) \varphi_{A_{n-1}}(x) dx = \pi^{(n-2)/2} \Big\{ \prod_{i \neq n,1} \Gamma(s_i - \frac{1}{2}) \Big\} \Gamma(s_1) \Delta^{(1)}_{-(s_1,\dots,s_{n-1})}(z), \tag{4.25}$$

if and only if $s_i > \frac{1}{2}$, for all $i \neq M$ and $s_M > 0$.

By Lemma 3, the change of variable $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$ gives $d\eta = x_{n-1,n-1} d\alpha d\beta dx$. Thus, we have

$$\textstyle\int_{Q_{A_n}} e^{-\mathrm{Tr}(y\eta)} \delta^{(1)}_{\underline{s}}(\eta) \varphi_{A_n}(\eta) d\eta = \int_0^\infty \int_{-\infty}^\infty \int_{Q_{A_{n-1}}} g(a, \beta, x) d\alpha d\beta dx,$$

where

$$g(a,\beta,x) = e^{-(a\alpha + a x_{n-1,n-1}(b+\beta)^2 + \mathrm{Tr}(zx))}\alpha^{s_n-3/2}\delta^{(1)}_{(s_1,\ldots,s_{n-1})}(x)\varphi_{A_{n-1}}(x)x^{1/2}_{n-1,n-1}$$

and where we used Lemmas 3 and part 1 of Lemma 4.3.3.

Now, using the Gaussian integral

$$\int_{-\infty}^{\infty} e^{-a x_{n-1,n-1}(b+\beta)^2}d\beta = \pi^{1/2}a^{-1/2}x^{-1/2}_{n-1,n-1}$$

and the gamma integral

$$\int_0^{\infty} e^{-a\alpha}\alpha^{s_n-3/2}d\alpha = a^{-s_n+1/2}\Gamma(s_n - \frac{1}{2}),$$

that converges if and only if $s_n > \frac{1}{2}$, we get

$$\begin{aligned}&\int_{Q_{A_n}} e^{-\mathrm{Tr}(y\eta)}\delta^{(1)}_{\underline{s}}(\eta)\varphi_{A_n}(\eta)d\eta \\ &= \pi^{1/2}a^{-s_n}\Gamma(s_n - \frac{1}{2})\int_{Q_{A_{n-1}}} e^{-\mathrm{Tr}(zx)}\delta^{(1)}_{(s_1,\ldots,s_{n-1})}(x)\varphi_{A_{n-1}}(x)dx.\end{aligned} \tag{4.26}$$

Finally, using Formulae (4.25) and (4.15) completes the proof.

□

Theorem 4.4.2. *For all $n \geqslant 1$, for all $1 \leqslant M \leqslant n$ and for all $\eta \in Q_{A_n}$, the integral $\int_{P_{A_n}} e^{-\mathrm{Tr}(y\eta)}\Delta^{(M)}_{\underline{s}}(y)dy$ converges if and only if $s_i > -\frac{3}{2}$ for all $i \neq M$, and $s_M > -1$. In this case, we have*

$$\int_{P_{A_n}} e^{-\mathrm{Tr}(y\eta)}\Delta^{(M)}_{\underline{s}}(y)dy = \pi^{(n-1)/2}\Big\{\prod_{i\neq M}\Gamma(s_i + \frac{3}{2})\Big\}\Gamma(s_M+1)\delta^{(M)}_{-\underline{s}}(\eta)\varphi_{A_n}(\eta). \tag{4.27}$$

Proof. We will proceed by induction on the number n of vertices.

- For $n = 1$,

$$\int_0^\infty e^{-y\eta}\Delta_s^{(1)}(y)dy = \int_0^\infty e^{-y\eta}y^s dy = \Gamma(s+1)\eta^{-s}\varphi_{A_1}(\eta),$$

if and only if $s > -1$.

- Now assume that the assertion holds for some number of vertices $n-1$.

Case $M > 1$:

Let $\eta = \Psi_n(\alpha, \beta, x)$ and let us make the change of variable $y = \Phi_n(a, b, z)$; then the induction hypothesis gives

$$\int_{P_{A_{n-1}}} e^{-\operatorname{Tr}(zx)}\Delta^{(M)}_{(s_2,\ldots,s_n)}(z)dz \tag{4.28}$$

$$= \pi^{(n-2)/2}\Big\{\prod_{i\neq 1,M}\Gamma(s_i+\frac{3}{2})\Big\}\Gamma(s_M+1)\delta^{(M)}_{-(s_2,\ldots,s_n)}(x)\varphi_{A_{n-1}}(x).$$

By Lemma 3, the change of variable $y = \Phi_n(a, b, z)$ gives $dy = a^d a db dz$. Thus, we have

$$\int_{P_{A_n}} e^{-\operatorname{Tr}(y\eta)}\Delta^{(M)}_{\underline{s}}(y)dy$$

$$= \int_0^\infty\int_{-\infty}^\infty\int_{P_{A_{n-1}}} e^{-(a\alpha+ax_{22}(b+\beta)^2+\operatorname{Tr}(zx))}a^{s_1}\Delta^{(M)}_{(s_2,\ldots,s_n)}(z)\, a^d\, da\, db\, dz$$

$$= \int_0^\infty\int_{-\infty}^\infty\int_{P_{A_{n-1}}} e^{-(a\alpha+ax_{22}(b+\beta)^2+\operatorname{Tr}(zx))}a^{s_1+1}\Delta^{(M)}_{(s_2,\ldots,s_n)}(z)\, a^d\, da\, db\, dz,$$

where we used Lemma 3 and part 1 of Lemma 4.3.3, for the first equality.

Now, using the Gaussian integral

$$\int_{-\infty}^\infty e^{-ax_{22}(b+\beta)^2}db = \pi^{1/2}a^{-1/2}x_{22}^{-1/2}$$

and the gamma integral

$$\int_0^\infty e^{-a\alpha}a^{s_1+1/2}da = \alpha^{-(s_1+3/2)}\Gamma(s_1+\frac{3}{2}),$$

if and only if $s_1 > -\frac{3}{2}$, we get

$$\int_{P_{A_n}} e^{-\operatorname{Tr}(y\eta)} \Delta_{\underline{s}}^{(M)}(y)dy = \pi^{1/2}\Gamma(s_1+\frac{3}{2})x_{22}^{-1/2}\alpha^{-(s_1+3/2)} \int_{P_{A_{n-1}}} e^{-\operatorname{Tr}(zx)} \Delta_{(s_2,\ldots,s_n)}^{(M)}(z)dz.$$

Finally, using (4.28), (4.14) and (4.18) completes the proof.

Case $M = 1$:

Let $\eta = \tilde{\Psi}_n(\alpha, \beta, x)$ and let us make the change of variable $y = \tilde{\Phi}_n(a, b, z)$; then the induction hypothesis gives

$$\int_{P_{A_{n-1}}} e^{-\operatorname{Tr}(zx)} \Delta_{(s_1,\ldots,s_{n-1})}^{(1)}(z)dz \tag{4.29}$$

$$= \pi^{(n-2)/2}\Big\{\prod_{i\neq n,1} \Gamma(s_i + \frac{3}{2})\Big\}\Gamma(s_1 + 1)\delta_{-(s_1,\ldots,s_{n-1})}^{(1)}(x)\varphi_{A_{n-1}}(x).$$

By Lemma 3, the change of variable $y = \tilde{\Phi}_n(a, b, z)$ gives $dy = a^d a db dz$. Thus, we have

$$\int_{P_{A_n}} e^{-\operatorname{Tr}(y\eta)} \Delta_{\underline{s}}^{(1)}(y)dy$$

$$= \int_0^\infty \int_{-\infty}^\infty \int_{P_{A_{n-1}}} e^{-(a\alpha+ax_{n-1,n-1}(b+\beta)^2+\operatorname{Tr}(zx))} a^{s_n} \Delta_{(s_1,\ldots,s_{n-1})}^{(1)}(z)\, a^d\, da\, db\, dz$$

$$= \int_0^\infty \int_{-\infty}^\infty \int_{P_{A_{n-1}}} e^{-(a\alpha+ax_{n-1,n-1}(b+\beta)^2+\operatorname{Tr}(zx))} a^{s_n+1} \Delta_{(s_1,\ldots,s_{n-1})}^{(M)}(z)\, a^d\, da\, db\, dz,$$

where we used Lemma 3 and part 1 of Lemma 4.3.3, for the first equality.

Now, using the Gaussian integral

$$\int_{-\infty}^\infty e^{-ax_{n-1,n-1}(b+\beta)^2} db = \pi^{1/2} a^{-1/2} x_{n-1,n-1}^{-1/2}$$

and the gamma integral

$$\int_0^\infty e^{-a\alpha} a^{s_n+1/2} da = \alpha^{-(s_n+3/2)}\Gamma(s_n + \frac{3}{2}),$$

if and only if $s_n > -\frac{3}{2}$, we get

$$\int_{P_{A_n}} e^{-\operatorname{Tr}(y\eta)} \Delta_{\underline{s}}^{(1)}(y)dy$$

$$= \pi^{1/2}\Gamma(s_n + \frac{3}{2}) x_{n-1,n-1}^{-1/2} \alpha^{-(s_n+3/2)} \int_{P_{A_{n-1}}} e^{-\operatorname{Tr}(zx)} \Delta_{(s_1,\ldots,s_{n-1})}^{(1)}(z)dz.$$

Finally, using (4.28), (4.16) and (4.19) completes the proof.

For $n > 1$ and the order $\prec_M$, we first make $M-1$ times use of case $M > 1$ (with Φ and Ψ) and next we make $n - M$ times use of case $M = 1$ (with $\tilde{\Phi}$ and $\tilde{\Psi}$).

□

Corollary 4.4.3. $\varphi_{Q_{A_n}} = \text{const}\,.\,\varphi_{A_n}$.

Proof. The result, $\left(\frac{4}{\pi^2}\right)^{\frac{n-1}{2}} \int_{P_{A_n}} e^{-\operatorname{Tr}(y\eta)}dy = \varphi_{A_n}(\eta)$, is obtained by substituting $\underline{s} = (0,\ldots,0)$ into Theorem 4.4.2. □

Remark 4.4.4. *Formulas* (4.22) *and* (4.27) *may seem similar but in* (4.27) *the integrand does not contain the characteristic function of the cone* P_{A_n}. *This function is unknown except for* A_4 *when it is not a power function (Letac and Massam, 2007, Prop.3.2).*

4.5 Wishart exponential families on Q_{A_n}

Let us define the Riesz measure $R_{\underline{s}}^{(M)}$ on Q_G by

$$R_{\underline{s}}^{(M)}(dx) = \frac{1}{\Gamma_{Q_{A_n}}^{(M)}(\underline{s})} \delta_{\underline{s}}^{(M)}(x)\varphi_{A_n}(x)1_{Q_{A_n}}(x)dx, \tag{4.30}$$

where $\Gamma_{Q_{A_n}}^{(M)}(\underline{s}) = \pi^{(n-1)/2}\left(\prod_{i\neq M}\Gamma(s_i - \frac{1}{2})\right)\Gamma(s_M)$. From Theorem 4.4.1, the Laplace transform of the measure $R_{\underline{s}}^{(M)}$ is given for all $s_i > \frac{1}{2}$, $i \neq M$ and $s_M > 0$ by

$$\mathcal{L}(R_{\underline{s}}^{(M)})(y) = \int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)} dR_{\underline{s}}^{(M)}(d\eta) = \Delta_{-\underline{s}}^{(M)}(y), \quad y \in P_{A_n}. \tag{4.31}$$

The Wishart natural exponential family $\gamma^{(M)}_{\underline{s},y}$ on Q_{A_n} is, by definition, generated by the Riesz measure $dR^{(M)}_{\underline{s}}$. The density function of the Wishart distribution on Q_{A_n} is given by

$$\gamma^{(M)}_{\underline{s},y}(dx) = \frac{1}{\Gamma^{(M)}_{Q_{A_n}}(\underline{s})} e^{-\operatorname{Tr}(yx)} \Delta^{(M)}_{\underline{s}}(y) \delta^{(M)}_{\underline{s}}(x) \varphi_{A_n}(x) 1_{Q_{A_n}}(x) dx. \tag{4.32}$$

The Laplace transform of $\gamma^{(M)}_{\underline{s},y}(dx)$ is

$$\mathcal{L}(\gamma^{(M)}_{\underline{s},y})(z) = \frac{\mathcal{L}(R^{(M)}_{\underline{s}})(z+y)}{\mathcal{L}(R^{(M)}_{\underline{s}})(y)} = \frac{\Delta^{(M)}_{-\underline{s}}(z+y)}{\Delta^{(M)}_{-\underline{s}}(y)}.$$

The family $\gamma^{(M)}_{\underline{s},y}$ does not depend on the normalization of the Riesz measure.

4.5.1 Mean and covariance of the Wishart distributions on Q_{A_n}

In this subsection we derive a formula for the mean of the Wishart exponential family on the cones Q_{A_n}. It is known from the general theory of exponential families of distributions, that the mean of $\gamma^{(M)}_{\underline{s},y}$ is obtained by differentiation with respect to y of the Laplace transform of the Riesz measure:

$$m^{(M)}_{\underline{s}}(y) = -\nabla_y \ln \Delta^{(M)}_{-\underline{s}}(y) \in Q_{A_n}. \tag{4.33}$$

For all matrix A in Z_G and a subset $B \subset V$ of the set of vertices V of A_n we note $(A_B)^0$ the matrix in Z_{A_n} such that $(A_B)^0_{ij} = \begin{cases} A_{ij} & \text{if} \quad i,j \in B, \\ 0 & \text{otherwise.} \end{cases}$

Proposition 4.5.1. *The mean function of the Wishart family $\gamma^{(M)}_{\underline{s},y}$ on Q_{A_n} is equal to*

$$m^{(M)}_{\underline{s}}(y) = \pi \left(\sum_{i=1}^{M-1} (s_i - s_{i+1}) [(y_{\{1:i\}})^{-1}]^0 + s_M y^{-1} + \sum_{i=M+1}^{n} (s_i - s_{i-1}) [(y_{\{i:n\}})^{-1}]^0 \right). \tag{4.34}$$

Proof. Use formulas (4.3), (4.33) and $\nabla_y \ln |y_A| = ((y_A)^{-1})^0$.

□

Proposition 4.5.2. *For all $y \in P_{A_n}$, we have*

$$\langle m_{\underline{s}}^{(M)}(y), y\rangle = \kappa(\underline{s}),$$

where the constant $\kappa(\underline{s})$ is $\sum_{i=1}^{n} s_i - (n-M)s_M$.

Proof. Observe that by (4.3), for any $c > 0$, $\Delta_{-\underline{s}}^{(M)}(cy) = c^{-\kappa(\underline{s})}\Delta_{-\underline{s}}^{(M)}(y)$.

Let $F : P_{A_n} \to \mathbb{R}$, $y \mapsto \ln \Delta_{-\underline{s}}^{(M)}(y)$ and $h_y : \mathbb{R} \to P_{A_n}$, $t \mapsto e^t y$.

We have $\nabla_y F = -m_{\underline{s}}^{(M)}(y)$ and $\nabla_t h_y = e^t y$.

Set $\varphi_y = F \circ h_y$. We have $\varphi_y(t) = -t\kappa(\underline{s}) + F(y)$ and $\nabla_t \varphi_y = -\kappa(\underline{s})$.

On the other hand, the chain rule gives $\nabla_t \varphi_y = \langle \nabla_{h_y(t)} F, \nabla_t h_y\rangle = \langle -m_{\underline{s}}^{(M)}(e^t y),\, e^t y\rangle$.

Thus, $\langle -m_{\underline{s}}^{(M)}(e^t y),\, e^t y\rangle = -\kappa(\underline{s})$. Taking $t = 0$ gives the desired result.

□

Differentiating the mean function gives the covariance function. For $A \in S_n$, let $\mathbb{P}(A) : Z_{A_n} \to I_G$ be the quadratic operator defined by $\mathbb{P}(A)u = \pi(AuA), \quad u \in Z_G$.

Proposition 4.5.3. *The covariance function of the Wishart family $\gamma_{\underline{s},y}^{(M)}$ on Q_{A_n} is equal*

$$v(y) = -\nabla_y m_{\underline{s}}^{(M)}(y) = \sum_{i=1}^{M-1} (s_i - s_{i+1})\, \mathbb{P}\left[\left((y_{\{1:i\}})^{-1}\right)^0\right] + s_M\, \mathbb{P}(y^{-1}) \tag{4.35}$$

$$+ \sum_{i=M+1}^{n} (s_i - s_{i-1})\, \mathbb{P}\left[\left((y_{\{i:n\}})^{-1}\right)^0\right].$$

4.5.2 Inverse mean map

In the study of the exponential family $(\gamma_{\underline{s},y}^{(M)})_{y\in P_{A_n}}$, it is important to determine explicitly the inverse of the mean map $\psi_{\underline{s}}^{(M)} : \; m = m_{\underline{s}}^{(M)}(y) \mapsto y$, which we refer to as the inverse mean map in the sequel. The following theorem is known for Wishart exponential families on homogeneous cones (Ishi, 2014). Surprisingly, it is also true on Q_{A_n}.

Theorem 4.5.4. *The inverse mean map $\psi_{\underline{s}}^{(M)}$ is given by the formula*

$$\psi_{\underline{s}}^{(M)}(m) = \nabla_m \ln \delta_{\underline{s}}^{(M)}(m),\ m \in Q_{A_n}. \tag{4.36}$$

The proof consists in following steps:

1. One shows that there exists a constant $c_{\underline{s}}$ depending only on $\underline{s}$ such that for any $y \in P_{A_n}$

$$\delta_{\underline{s}}^{(M)}(m_{\underline{s}}^{(M)}(y)) = c_{\underline{s}}\Delta_{-\underline{s}}^{(M)}(y) = c_{\underline{s}}\delta_{\underline{s}}^{(M)}(\pi(y^{-1})).$$

This is done in Proposition 4.5.5 below.

2. One uses a differential calculus argument, based on the Legendre transform methods.

Proposition 4.5.5. *The following formula holds for any $y \in P_{A_n}$ and $\underline{s} \in \mathbb{R}^n$:*

$$\delta_{\underline{s}}^{(M)}(m_{\underline{s}}^{(M)}(y)) = \left(\prod_{i=1}^{n} s_i^{s_i}\right) \Delta_{-\underline{s}}^{(M)}(y) = \left(\prod_{i=1}^{n} s_i^{s_i}\right)\ \delta_{\underline{s}}^{(M)}(\pi(y^{-1})).$$

The proof of Proposition 4.5.5 will need a generalization of Lemma 4.2.11, where coefficients of inverse matrices of principal submatrices $y_{\{1:k\}}$ (or of $y_{\{k:n\}}$) are simultanously considered. Define for $y \in P_{A_n}$, $\eta^{(k)} = (y_{\{1:k\}})^{-1}$, $\eta^{[k]} = (y_{\{k:n\}})^{-1}$. The rows and the columns of the matrix $\eta^{(k)}$ are numbered by $i = 1, \ldots, k$ and the rows and the columns of the matrix $\eta^{[k]}$ are numbered by $i = k, \ldots, n$.

Lemma 4.5.6. *Let $y \in P_{A_n}$.*

1. *For all $i \in V$ and $k, j \geqslant i+1$ we have*

$$D_i^{k,j} := \begin{vmatrix} \eta_{ii}^{(k)} & \eta_{i,i+1}^{(j)} \\ \eta_{i,i+1}^{(k)} & \eta_{i+1,i+1}^{(j)} \end{vmatrix} = |y_{\{1:j\}}|^{-1}|y_{\{1:j\}\setminus\{i,i+1\}}|. \tag{4.37}$$

2. *For all $i \in V$ and $k, j \leqslant i < n$ we have*

$$D_i^{[k,j]} := \begin{vmatrix} \eta_{ii}^{[k]} & \eta_{i,i+1}^{[j]} \\ \eta_{i,i+1}^{[k]} & \eta_{i+1,i+1}^{[j]} \end{vmatrix} = |y_{\{k:n\}}|^{-1}|y_{\{k:n\}\setminus\{i,i+1\}}|. \tag{4.38}$$

Proof. Similar to the proof of Lemma 4.2.11; instead of y use $y_{\{1:k\}}$ or $y_{\{k:n\}}$. □

Proof. (of Proposition 4.5.5) We will deal with $\delta_{\underline{s}}^{(M)}(m_{\underline{s}}^{(M)}(y)) = \delta_{\underline{s}}^{\prec_M}(m_{\underline{s}}^{(M)}(y))$ where the order $\prec_M$ was defined in (4.6). By formula (4.34) and by the definition of $\delta_{\underline{s}}^{\prec_M}$ we obtain that $\delta_{\underline{s}}^{\prec_M}(m_{\underline{s}}(y))$ equals

$$\prod_{i=1}^{M-1}\left(\frac{1}{c_i}\begin{vmatrix} x_i+a_i & b_i \\ b_i & c_i\end{vmatrix}\right)^{s_i}(s_M\eta_{MM}^{(n)})^{s_M}\prod_{i=M+1}^{n}\left(\frac{1}{c_i'}\begin{vmatrix} x_i'+a_i' & b_i' \\ b_i' & c_i'\end{vmatrix}\right)^{s_i},$$

where $x_i = (s_i - s_{i+1})\eta_{ii}^{(i)}$, $a_i = \sum_{k=i+1}^{M-1}(s_k - s_{k+1})\eta_{ii}^{(k)} + s_M\eta_{ii}^{(n)}$,

$$b_i = \sum_{k=i+1}^{M-1}(s_k - s_{k+1})\eta_{i,i+1}^{(k)} + s_M\eta_{i,i+1}^{(n)},$$

$$c_i = \sum_{k=i+1}^{M-1}(s_k - s_{k+1})\eta_{i+1,i+1}^{(k)} + s_M\eta_{i+1,i+1}^{(n)},$$

$$a_i' = \sum_{k=M+1}^{i-1}(s_k - s_{k-1})\eta_{ii}^{[k]} + s_M\eta_{ii}^{[1]},$$

$$b_i' = \sum_{k=M+1}^{i-1}(s_k - s_{k-1})\eta_{i,i-1}^{[k]} + s_M\eta_{i,i-1}^{[1]},$$

$$c_i' = \sum_{k=M+1}^{i-1}(s_k - s_{k-1})\eta_{i-1,i-1}^{[k]} + s_M\eta_{i-1,i-1}^{[1]},$$

and $x_i' = (s_i - s_{i-1})\eta_{ii}^{[i]}$.

Let us first compute the factors $\begin{vmatrix} x_i+a_i & b_i \\ b_i & c_i\end{vmatrix}/c_i$ for $i \in \{1,\dots,M-1\}$. We will show that

$$\frac{1}{c_i}\begin{vmatrix} x_i+a_i & b_i \\ b_i & c_i\end{vmatrix} = s_i\eta_{ii}^{(i)}, \quad i \in \{1,\dots,M-1\}. \tag{4.39}$$

We have $\frac{1}{c_i}\begin{vmatrix} x_i+a_i & b_i \\ b_i & c_i\end{vmatrix} = x_i + \frac{1}{c_i}\begin{vmatrix} a_i & b_i \\ b_i & c_i\end{vmatrix}$, so in order to prove (4.39), it is sufficient to

prove that

$$\frac{1}{c_i}\begin{vmatrix} a_i & b_i \\ b_i & c_i \end{vmatrix} = s_{i+1}\eta_{ii}^{(i)}. \tag{4.40}$$

In order to prove (4.40), we first use the multilinearity of the determinant with respect to its columns and we write, using the notation $D_i^{k,j}$ from Lemma 4.5.6,

$$\begin{aligned}\begin{vmatrix} a_i & b_i \\ b_i & c_i \end{vmatrix} &= \sum_{k,j=i+1}^{M-1}(s_k - s_{k+1})(s_j - s_{j+1})D_i^{k,j} + s_M\sum_{k=i+1}^{M-1}(s_k - s_{k+1})D_i^{k,n} \\ &+ s_M\sum_{j=i+1}^{M-1}(s_j - s_{j+1})D_i^{n,j} + s_M^2 D_i^{n,n}.\end{aligned}$$

By Part 1 of Lemma 4.5.6 we have $D_i^{k,j} = |y_{\{1:j\}}|^{-1}|y_{\{1:j\}\setminus\{i,i+1\}}|$, which is independent of the left index k. The last fact allows to write

$$\begin{aligned}\begin{vmatrix} a_i & b_i \\ b_i & c_i \end{vmatrix} &= s_{i+1}\sum_{j=i+1}^{M-1}(s_j - s_{j+1})D_i^{n,j} + s_{i+1}s_M D_i^{n,n} \\ &= s_{i+1}\left(\sum_{j=i+1}^{M-1}(s_j - s_{j+1})\frac{|y_{\{1:j\}\setminus\{i,i+1\}}|}{|y_{\{1:j\}}|} + s_M\frac{|y_{\{1:n\}\setminus\{i,i+1\}}|}{|y|}\right).\end{aligned}$$

We factorize the determinants $|y_{\{1:j\}\setminus\{i,i+1\}}|$ and $|y_{\{1:n\}\setminus\{i,i+1\}}|$ in the last sum according to Lemma 4.2.10 and we write this sum as

$$\frac{|y_{\{1:i-1\}}|}{|y_{\{1:i\}}|}\left(\sum_{j=i+1}^{M-1}(s_j - s_{j+1})\frac{|y_{\{1:i\}}||y_{\{i+2:j\}}|}{|y_{\{1:j\}}|} + s_M\frac{|y_{\{1:i\}}||y_{\{i+2:n\}}|}{|y|}\right).$$

We have $|y_{\{1:j\}}|^{-1}|y_{\{1:i\}}||y_{\{i+2:j\}}| = \eta_{i+1,i+1}^{(j)}$. By definition of c_i we finally obtain

$$\begin{vmatrix} a_i & b_i \\ b_i & c_i \end{vmatrix} = s_{i+1}\frac{|y_{\{1:i-1\}}|}{|y_{\{1:i\}}|}c_i = s_{i+1}\eta_{ii}^{(i)}c_i$$

and formulas (4.40) and (4.39) are proved.

A 'mirror' proof based on Part 2 of Lemma 4.5.6 shows that

$$\frac{1}{c_i'}\begin{vmatrix} x_i' + a_i' & b_i' \\ b_i' & c_i' \end{vmatrix} = s_i\eta_{ii}^{[i]}, \quad i = M+1,\ldots,n \tag{4.41}$$

and that $\delta_{\underline{s}}^{(M)}(m_{\underline{s}}^{(M)}(y)) = \prod_{i=1}^{n} s_i^{s_i} \prod_{i=1}^{M-1} (\eta_{ii}^{(i)})^{s_i} (\eta_{MM}^{(n)})^{s_M} \prod_{i=M+1}^{n} (\eta_{ii}^{[i]})^{s_i}$.

Recall that

$$\eta_{ii}^{(i)} = \frac{|y_{\{1:i-1\}}|}{|y_{\{1:i\}}|}, \quad \eta_{ii}^{[i]} = \frac{|y_{\{i+1:n\}}|}{|y_{\{i:n\}}|}, \quad \eta_{MM}^{(n)} = \frac{|y_{\{1:M-1\}}||y_{\{M+1:n\}}|}{|y|},$$

so we deduce, using formula (4.3) that

$$\prod_{i=1}^{M-1} (\eta_{ii}^{(i)})^{s_i} (\eta_{MM}^{(n)})^{s_M} \prod_{i=M+1}^{n} (\eta_{ii}^{[i]})^{s_i} = \Delta_{-\underline{s}}^{(M)}(y).$$

Applying Theorem 4.2.9, we see that $\delta_{\underline{s}}^{(M)}(m_{\underline{s}}^{(M)}(y)) = \prod_{i=1}^{n} s_i^{s_i} \delta_{\underline{s}}^{(M)}(\pi(y^{-1}))$. □

Proof. (of Theorem 4.5.4).

Let $m_0 \in Q_{A_n}$ and $y_0 = \Psi_{\underline{s}}^{(M)}(y_0)$. Then, by formula (4.33), we have

$$m_0 = m_{\underline{s}}^{(M)}(y_0) = -\nabla_y \ln \Delta_{-\underline{s}}^{(M)}(y_0) = \nabla_y f(y_0),$$

where $f(y) = -\ln \Delta_{-\underline{s}}^{(M)}(y)$.

Let f^* be the Legendre-Fenchel conjugate of $f : Q_{A_n} \to \mathbb{R}$:

$f^*(m) = \sup_{y \in P_G} \{\langle m, y\rangle - f(y)\} = \sup_{y \in P_G} g_m(y)$ with $g_m(y) = \langle m, y\rangle - f(y)$.

We have $\nabla_y g_m(y) = m - \nabla_y f(y)$ and the Hessian of g_m is given by

$H(g_m)(y) = -H(f)(y)$. Since f is convex, g_{m_0} has a unique maximum y^* which satisfies $\nabla_y g_{m_0}(y^*) = 0$ and thus $m_0 = \nabla_y f(y^*)$. Hence, $y^* = y_0$, $f^*(m_0) = g_{m_0}(y_0) = \langle m_0, y_0\rangle - f(y_0)$ and $\nabla_{m_0} f^*(m_0) = y_0$. Now using Proposition 4.5.2, we get

$$y_0 = \nabla_{m_0} (\langle m_0, y_0\rangle - f(y_0)) = \nabla_{m_0} (\kappa_{\underline{s}} - f(y_0)) = -\nabla_{m_0} f(y_0) = \nabla_{m_0} \ln \Delta_{-\underline{s}}^{(M)}(y_0).$$

Finally, Proposition 4.5.5 gives

$$y_0 = \nabla_{m_0} \ln c_{\underline{s}}^{-1} \delta_{\underline{s}}^{(M)}(m_0) = \nabla_{m_0} \ln \delta_{\underline{s}}^{(M)}(m_0).$$

□

Corollary 4.5.7. *The inverse mean map $\psi_{\underline{s}}^{(M)} : Q_{A_n} \to P_{A_n}$ is given by*

$$\psi_{\underline{s}}^{(M)}(m) = \sum_{k=1}^{M-1} s_k \left((m_{\{k:k+1\}})^{-1}\right)^0 + \sum_{k=M+1}^{n} s_k \left((m_{\{k-1:k\}})^{-1}\right)^0$$
$$- \sum_{k=2}^{M-1} s_{k-1} \left((m_{\{kk\}})^{-1}\right)^0 - (s_{M-1} - s_M + s_{M+1}) \left((m_{\{MM\}})^{-1}\right)^0$$
$$- \sum_{k=M+1}^{n-1} s_{k+1} \left((m_{\{kk\}})^{-1}\right)^0. \tag{4.42}$$

Proof. The result is obtained by computing the gradient of $\ln \delta_{\underline{s}}^{(M)}(m)$, as indicated in (4.36). We use the formula (4.4). □

The Lauritzen formula (Lauritzen, 1996) is an explicit formula for a bijection between Q_G and P_G. It states that for all $x \in Q_{A_n}$, the unique $y \in P_{A_n}$ such that $\pi(y^{-1}) = x$ is given by

$$y = \sum_{i=1}^{n-1} (x_{\{i:i+1\}}^{-1})^0 - \sum_{i=2}^{n-1} (x_{ii}^{-1})^0. \tag{4.43}$$

Setting $s_1 = \ldots = s_n = 1$ in formula (4.34) for the mean function, we get

$$m_{(1,\ldots,1)}^{(M)}(y) = \pi(y^{-1}) = x. \tag{4.44}$$

Thus,

$$\psi_{(1,\ldots,1)}^{(M)}(x) = y \tag{4.45}$$

is the Lauritzen formula. Indeed, for $s_1 = \ldots = s_n = 1$, formula (4.42) gives

$$\psi_{(1,\ldots,1)}^{(M)}(m) = \sum_{i=1}^{n-1} (m_{\{i:i+1\}}^{-1})^0 - \sum_{i=2}^{n-1} (m_{ii}^{-1})^0. \tag{4.46}$$

Thus we found a new proof of the Lauritzen formula, based on the observation that the Lauritzen map is the inverse mean map for $\underline{s} = \mathbf{1} = (1, 1, \ldots, 1)$. At the same time we

find an infinite number of explicit isomorphisms from Q_{A_n} onto P_{A_n}, given by the inverse mean maps $\psi_{\underline{s}}^{(M)}$. It is an essential generalization of the Lauritzen formula. Each map $\psi_{\underline{s}}^{(M)}$ is a generalized Lauritzen map.

Remark 4.5.8. We note that if $\bar{X}$ is the mean of a sample from $\gamma_{\underline{s},y}^{(M)}$, then the maximum likelihood estimator of y is $\psi_{\underline{s}}^{(M)}(\bar{X})$.

4.5.3 Variance function

Properties of lower-upper M-triangular matrices

Here, we define and prove basic properties of lower-upper M-triangular matrices, that we will denote by LU(M). They are very important in proofs of this section.

Definition 4.5.9. A matrix T is said to be an LU(M) triangular matrix if for all $i < M$, $T_{ij} = 0$ if $j > i$ and for all $i > M$, $T_{ij} = 0$ if $i > j$.

In particular, T is an LU(n) triangular matrix if and only if it is lower triangular, and T is an LU(1) triangular matrix if and only if it is upper triangular. An LU(M) triangular matrix T is a succession of an $M \times M$ lower triangular matrix $L = T_{\{1:M\}}$ and an $(N - M) \times (N - M)$ upper triangular matrix $U = T_{\{M:n\}}$ with diagonal term T_{MM} in common. We write $T = s(L, U)$.

$$T = \begin{bmatrix} & 0 & \\ L & \ddots & T_{MM} \\ 0 & & U \end{bmatrix}$$

T can be decomposed in blocks as

$$T = \begin{pmatrix} T_{1:M,1:M} & T_{1:M,M+1:n} \\ 0 & T_{M+1:n,M+1:n} \end{pmatrix} \text{ or } T = \begin{pmatrix} T_{1:M-1,1:M-1} & 0 \\ T_{M:n,1:M-1} & T_{M:n,M:n} \end{pmatrix}, \tag{4.47}$$

where $T_{1:M,1:M}$ is a lower triangular matrix, $T_{M+1:n,M+1:n}$ is an upper triangular matrix, $T_{M:n,1:M-1}$ is a matrix with all rows but the last one having zero elements, $T_{1:M-1,1:M-1}$ is a lower triangular matrix, $T_{M:n,M:n}$ is an upper triangular matrix and $T_{M:n,1:M-1}$ is a matrix with all rows but the first one having zero elements.

Proposition 4.5.10. *1.* $s(L,U)s(L',U') = s(LL',UU')$.

2. *If* $s(L,U)$ *is invertible, then* $(s(L,U))^{-1}$ *is also an LU(M) triangular matrix and* $(s(L,U))^{-1} = s(L^{-1},U^{-1})$.

3. *The set of* $LU(M)$ *triangular matrices is a group.*

Proof. Part 1 is proved by block matrix multiplication. Part 2 is straightforward using Part 1. Part 3 follows from Parts 1 and 2.

□

Lemma 4.5.11. *Let* S *and* T *be LU(M) triangular* $n \times n$ *matrices.*

1. *(a) Let* $A = K^0$ *with* $K = A_{\{1:k\}}$. *If* $k \leqslant M-1$, *then* ${}^tSAT = \left({}^tS_{\{1:k\}}KT_{\{1:k\}}\right)^0$.

   (b) Let $B = K^0$ *with* $K = B_{\{k:n\}}$. *If* $k \geqslant M+1$, *then* ${}^tSBT = \left({}^tS_{\{k:n\}}KT_{\{k:n\}}\right)^0$.

2. *Let* A *be an* $n \times n$ *matrix. Then* $(TA\,{}^tS)_{\{1:i\}} = T_{\{1:i\}}A_{\{1:i\}}\,{}^tS_{\{1:i\}}$ *for* $i \leqslant M-1$, *and* $(TA\,{}^tS)_{\{i:n\}} = T_{\{i:n\}}A_{\{i:n\}}\,{}^tS_{\{i:n\}}$ *for* $i \geqslant M+1$.

3. *If* T *is invertible, then*

   (a) $(T_{\{1:k\}})^{-1} = (T^{-1})_{\{1:k\}}$ *for all* $k \leqslant M-1$;

   (b) $(T_{\{k:n\}})^{-1} = (T^{-1})_{\{k:n\}}$ *for all* $k \geqslant M+1$.

Proof. Part 1 is straightforward using block matrix multiplication and Part 1 of Lemma 4.8.1 in Appendix; for Part 2, just imagine which lines and columns intervene in the computation; Part 3 follows from Part 2 of Proposition 4.5.10 and Part 3 of Lemma 4.8.1. □

Proposition 4.5.12. *For all $y \in P_{A_n}$, for all $1 \leqslant M \leqslant n$, there exists an LU(M) triangular matrix T satisfying $T_{ij} = 0$ if $i \not\prec j$ and such that $y = T\ {}^{\mathrm{t}}T$.*

Proof. We will proceed by induction on n. The statement is obviously true for $n = 1$. Let us assume that the statement is true for $n-1$. Let $y \in P_{A_n}$ and $M \neq 1$. Let us write $y = \Phi_n(a, b, z)$ with $z \in P_{A_{n-1}}$. The induction assumption implies that there exists V an $(n-1) \times (n-1)$ LU(M) triangular matrix such that $V_{ij} = 0$ if $i \not\prec j$ and such that $z = V\ {}^{\mathrm{t}}V$. Let us write

$$T = \begin{pmatrix} 1 & & & \\ b & 1 & & \\ \vdots & & \ddots & \\ 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{a} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & V & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} \sqrt{a} & 0 & \dots & 0 \\ \sqrt{a}b & & & \\ \vdots & & V & \\ 0 & & & \end{pmatrix}.$$

T is LU(M) triangular satisfying $T_{ij} = 0$ if $i \not\prec j$ and $y = T\ {}^{\mathrm{t}}T$.
For $M = 1$, we use $y = \tilde{\Phi}_n(a, b, z)$ with $z \in P_{A_{n-1}}$. □

Two formulas for the variance function

Let $m \in Q_{A_n}$. Let $\hat{m} \in S_n^+$ be the unique symmetric positive definite matrix verifying $\pi(\hat{m}) = m$ and $\hat{m}^{-1} \in P_{A_n}$. We note the following interpretation of $\hat{m}$: if m is the mean of a sample from the Wishart model $\gamma^{(M)}_{(1,\dots,1),y}$, then $\hat{m}$ is the maximum likelihood estimator of y^{-1}.

Define $y = \psi^{(M)}_{\underline{s}}(m) \in P_{A_n}$. Decompose $y = T\ {}^{\mathrm{t}}T$, with T an LU(M) triangular matrix such that $T_{ij} = 0$ when $i \not\prec j$.

Lemma 4.5.13. *We have*

$$\hat{m} = {}^{\mathrm{t}}T^{-1} \begin{pmatrix} s_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & s_n \end{pmatrix} T^{-1}. \tag{4.48}$$

Proof. Note that $y = \psi_{\underline{s}}^{(M)}(m)$ is equivalent to $m = m_{\underline{s}}^{(M)}(y)$. The formula of the mean function (4.34) gives $m = \pi(Z)$, where

$$Z = \sum_{i=1}^{M-1}(s_i - s_{i+1})[(y_{\{1:i\}})^{-1}]^0 + s_M y^{-1} + \sum_{i=M+1}^{n}(s_i - s_{i-1})[(y_{\{i:n\}})^{-1}]^0. \tag{4.49}$$

Using Part 2 of Lemma 4.5.11, we have $y_{\{1:i\}} = T_{\{1:i\}} I_{\{1:i\}}\, {}^tT_{\{1:i\}}$ for $i \leqslant M-1$. By Part 3 of Lemma 4.5.11, we get $(y_{\{1:i\}})^{-1} = {}^t(T^{-1})_{\{1:i\}} I_{\{1:i\}} (T^{-1})_{\{1:i\}}$. Finally, using Part 1 of Lemma 4.5.11, we obtain

$$[(y_{\{1:i\}})^{-1}]^0 = {}^tT^{-1}(I_{\{1:i\}})^0 T^{-1}, \quad i \leqslant M-1. \tag{4.50}$$

Similarly, we have

$$[(y_{\{i:n\}})^{-1}]^0 = {}^tT^{-1}(I_{\{i:n\}})^0 T^{-1}, \quad i \geqslant M+1. \tag{4.51}$$

Thus,

$$\begin{aligned} Z &= {}^{\mathrm{t}}T^{-1}\left(\sum_{i=1}^{M-1}(s_i - s_{i+1})(I_{\{1:i\}})^0 + s_M I + \sum_{i=M+1}^{n}(s_i - s_{i-1})(I_{\{i:n\}})^0\right)T^{-1} \\ &= {}^{\mathrm{t}}T^{-1}\begin{pmatrix} s_1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & s_n \end{pmatrix}T^{-1}. \end{aligned}$$

Therefore, Z is positive definite and $Z^{-1} = T\begin{pmatrix} s_1^{-1} & \dots & 0 \\ & \ddots & \\ 0 & \dots & s_n^{-1} \end{pmatrix}{}^{\mathrm{t}}T \in P_{A_n}$. Indeed, for all $i < i+1 < j$, we have $(Z^{-1})_{ij} = \sum_{k=1}^{n} T_{ik}T_{jk}s_k^{-1}$. Since $T_{ik} = 0$ for $|k-i| > 1$, $(Z^{-1})_{ij} = T_{i,i-1}T_{j,i-1}s_{i-1}^{-1} + T_{ii}T_{ji}s_i^{-1} + T_{i,i+1}T_{j,i+1}s_{i+1}^{-1}$. But since $|j-i| > 1$, we have $T_{j,i-1} = 0 = T_{ji}$ and $(Z^{-1})_{ij} = T_{i,i+1}T_{j,i+1}s_{i+1}^{-1}$. Now since T is LU(M), we have $T_{i,i+1}T_{j,i+1} = 0$. In fact, $T_{i,i+1} = 0$ for $i \leqslant M-1$ and $T_{j,i+1} = 0$ for $i \geqslant M$. In conclusion, we have shown that $m = \pi(Z)$ with $Z^{-1} \in P_{A_n}$, which implies $Z = \hat{m}$. □

The following Proposition derives the formula for the variance function $V(m)$ which, for each fixed $m \in Q_{A_n}$ is a continuous operator $V(m) : Z_{A_n} \to I_{A_n}$ (Casalis and Letac, 1996). Recall that $\mathbb{P}(A) : Z_{A_n} \to I_{A_n}$ is the quadratic operator defined by $\mathbb{P}(A)u = \pi(AuA)$. For $A, B \in S_n$, let $\mathbb{P}(A,B)u = \frac{1}{2}\pi(AuB + BuA)$. For all $m \in Q_{A_n}$ and $I \subset V$, we note

$$M_I = [((\hat{m}^{-1})_I)^{-1}]^0. \tag{4.52}$$

Proposition 4.5.14. *The variance function $V^{(M)}(m)$ of a Wishart exponential family on Q_{A_n} is equal to*

$$\begin{aligned} V^{(M)}(m) = & \sum_{i=1}^{M-1} (s_i - s_{i+1})\, \mathbb{P}\left(\sum_{j=1}^{i-1} \left(\frac{1}{s_j} - \frac{1}{s_{j+1}} \right) M_{\{1:j\}} + \frac{1}{s_i} M_{\{1:i\}} \right) \\ & + s_M\, \mathbb{P}\left(\frac{\hat{m}}{s_M} + \sum_{j=1}^{M-1} \left(\frac{1}{s_j} - \frac{1}{s_{j+1}} \right) M_{\{1:j\}} + \sum_{k=M+1}^{n} \left(\frac{1}{s_k} - \frac{1}{s_{k-1}} \right) M_{\{k:n\}} \right) \\ & + \sum_{i=M+1}^{n} (s_i - s_{i-1})\, \mathbb{P}\left(\frac{1}{s_i} M_{\{i:n\}} + \sum_{j=i+1}^{n} \left(\frac{1}{s_j} - \frac{1}{s_{j-1}} \right) M_{\{j:n\}} \right). \end{aligned} \tag{4.53}$$

Proof. The variance function is given for all $m \in Q_{A_n}$ by $V^{(M)}(m) = v(\psi_{\underline{s}}^{(M)}(m))$, where $v(y)$ is given by (4.35). Let $y = \psi_{\underline{s}}^{(M)}(m) = T\,{}^tT$, where T is LU(M). From Lemma 4.5.13, we have

$$\hat{m}^{-1} = T \begin{pmatrix} s_1^{-1} & \dots & 0 \\ & \ddots & \\ 0 & \dots & s_n^{-1} \end{pmatrix} {}^tT.$$

Using Lemma 4.5.11, we get

$$M_{\{1:i\}} = {}^tT^{-1} \left(diag(s_1, \dots, s_i) \right)^0 T^{-1}, \quad i \leqslant M-1 \tag{4.54}$$

and

$$M_{\{i:n\}} = {}^tT^{-1} \left(diag(s_i, \dots, s_n) \right)^0 T^{-1}, \quad i \geqslant M+1. \tag{4.55}$$

Thus, for all $2 \leqslant i \leqslant M-1$, we have

$$\frac{1}{s_1} M_1 = {}^tT^{-1} e_1 T^{-1}, \quad \frac{1}{s_i}(M_{\{1:i\}} - M_{1:i-1)} = {}^tT^{-1} e_i T^{-1}, \tag{4.56}$$

and for all $n-1 \geqslant i \geqslant M+1$, we have

$$\frac{1}{s_n} M_n = {}^tT^{-1} e_n T^{-1}, \quad \frac{1}{s_i}(M_{\{i:n\}} - M_{i+1:n}) = {}^tT^{-1} e_i T^{-1}, \tag{4.57}$$

where e_i is the matrix with $e_{ii} = 1$ and $e_{ij} = 0$ for all $i \neq j$. Observing that $(I_{\{1:i\}})^0 = \sum_{k=1}^{i} e_i$ and $(I_{\{i:n\}})^0 = \sum_{k=i}^{n} e_i$, and using (4.50) and (4.56), we obtain for $i \leqslant M-1$

$$\begin{aligned}
[(y_{\{1:i\}})^{-1}]^0 &= {}^tT^{-1}(I_{\{1:i\}})^0 T^{-1} = {}^tT^{-1}\left(\sum_{k=1}^{i} e_i\right) T^{-1} = \sum_{k=1}^{i} \left({}^tT^{-1} e_i T^{-1}\right) \\
&= \frac{1}{s_1} M_{\{1\}} + \frac{1}{s_2}(M_{\{1:2\}} - M_{\{1\}}) + \ldots + \frac{1}{s_i}(M_{\{1:i\}} - M_{\{1:i-1\}}) \\
&= \left(\frac{1}{s_1} - \frac{1}{s_2}\right) M_{\{1\}} + \ldots + \left(\frac{1}{s_{i-1}} - \frac{1}{s_i}\right) M_{\{1:i-1\}} + \frac{1}{s_i} M_{\{1:i\}}.
\end{aligned}$$

Similarly, using (4.51) and (4.57), we obtain for $i \geqslant M+1$,

$$[(y_{\{i:n\}})^{-1}]^0 = \frac{1}{s_i} M_{\{i:n\}} + \left(\frac{1}{s_{i+1}} - \frac{1}{s_i}\right) M_{\{i+1:n\}} + \ldots + \left(\frac{1}{s_n} - \frac{1}{s_{n-1}}\right) M_{\{n\}}.$$

We also observe that

$${}^tT^{-1} e_M T^{-1} = \frac{1}{s_M}\left(\hat{m} - M_{\{1:M-1\}} - M_{\{M+1:n\}}\right). \tag{4.58}$$

Thus, by (4.56), (4.57) and (4.58), we get

$$\begin{aligned}
y^{-1} &= \sum_{i=1}^{n} {}^tT^{-1} e_i T^{-1} = \sum_{i=1}^{M-1} {}^tT^{-1} e_i T^{-1} + {}^tT^{-1} e_M T^{-1} + \sum_{i=M+1}^{n} {}^tT^{-1} e_i T^{-1} \\
&= \frac{\hat{m}}{s_M} + \sum_{j=1}^{M-1} \left(\frac{1}{s_j} - \frac{1}{s_{j+1}}\right) M_{\{1:j\}} + \sum_{j=M+1}^{n} \left(\frac{1}{s_j} - \frac{1}{s_{j-1}}\right) M_{\{j:n\}}.
\end{aligned} \tag{4.59}$$

Substituting these expressions of $[(y_{\{1:i\}})^{-1}]^0$, y^{-1} and $[(y_{\{i:n\}})^{-1}]^0$ into $v(y)$ given by (4.35), we obtain the stated result. □

We prove now a much simpler formula for the variance function on Q_G, surprisingly similar to the variance function on a homogeneous cone, in particular on the symmetric cone S_n^+ (cf. Graczyk et al. (2016a)).

Theorem 4.5.15. *The variance function of the Wishart exponential family* $\gamma_{\underline{s},y}^{(M)}$ *is*

$$V^{(M)}(m) = \left(\frac{1}{s_1} + \frac{1}{s_n} - \frac{1}{s_M}\right)\mathbb{P}(\hat{m}) \tag{4.60}$$
$$+ \sum_{i=1}^{M-1}\left(\frac{1}{s_{i+1}} - \frac{1}{s_i}\right)\mathbb{P}(\hat{m} - M_{\{1:i\}}) + \sum_{i=M+1}^{n}\left(\frac{1}{s_{i-1}} - \frac{1}{s_i}\right)\mathbb{P}(\hat{m} - M_{\{i:n\}}),$$

where $M_{\{1:i\}}$ *and* $M_{\{i:n\}}$ *are defined in* (4.52).

Proof. Using $\mathbb{P}(a-b) = \mathbb{P}(a) + \mathbb{P}(b) - 2\,\mathbb{P}(a,b)$, we see that (4.60) is equivalent to

$$V^{(M)}(m) = \frac{1}{s_M}\mathbb{P}(\hat{m}) + \sum_{i=1}^{M-1}\left(\frac{1}{s_{i+1}} - \frac{1}{s_i}\right)\mathbb{P}(M_{\{1:i\}}) + \sum_{i=M+1}^{n}\left(\frac{1}{s_{i-1}} - \frac{1}{s_i}\right)\mathbb{P}(M_{\{i:n\}})$$
$$- 2\left(\sum_{i=1}^{M-1}\left(\frac{1}{s_{i+1}} - \frac{1}{s_i}\right)\mathbb{P}(\hat{m}, M_{\{1:i\}}) + \sum_{i=M+1}^{n}\left(\frac{1}{s_{i-1}} - \frac{1}{s_i}\right)\mathbb{P}(\hat{m}, M_{\{i:n\}})\right). \tag{4.61}$$

We show that the right hand sides of (4.53) and (4.61) are the same. Below, we expand (4.53) using $\mathbb{P}(a+b) = \mathbb{P}(a) + \mathbb{P}(b) + 2\,\mathbb{P}(a,b)$ and compute the coefficients in the expanded formula. Note that for all $Z \in Z_{A_n}$, $\mathbb{P}(M_{\{1:i\}}, M_{\{k:n\}})Z = 0$ for all $i \leqslant M-1$ and $k \geqslant M+1$, since $Z_{\{1:i\},\{k:n\}} = 0$.

For a fixed $r \leqslant M-1$, the coefficient of $\mathbb{P}(M_{\{1:r\}})$ is

$$\frac{s_r - s_{r+1}}{s_r^2} + \sum_{i=r+1}^{M-1}(s_i - s_{i+1})\left(\frac{1}{s_r} - \frac{1}{s_{r+1}}\right)^2 + s_M\left(\frac{1}{s_r} - \frac{1}{s_{r+1}}\right)^2 = \frac{1}{s_{r+1}} - \frac{1}{s_r}.$$

By a mirror argument, for a fixed $r \geqslant M+1$, the coefficient of $\mathbb{P}(M_{\{r:n\}})$ is $\frac{1}{s_{r-1}} - \frac{1}{s_r}$. On the other hand, the coefficient of $\mathbb{P}(\hat{m})$ is $\frac{1}{s_M}$.

For a fixed r, the coefficient of $\mathbb{P}(\hat{m}, M_{\{1:r\}})$ is $\frac{1}{s_r} - \frac{1}{s_{r+1}}$ if $r \leqslant M-1$, and the coefficient

of $\mathbb{P}(\hat{m}, M_{\{r:n\}})$ is $\frac{1}{s_r} - \frac{1}{s_{r-1}}$ if $r \geqslant M+1$. Moreover, if $k < r \leqslant M-1$, the coefficient of $\mathbb{P}(M_{\{1:r\}},\ M_{\{1:k\}})$ is

$$
\begin{aligned}
&(s_r - s_{r+1})\frac{1}{s_r}\left(\frac{1}{s_k} - \frac{1}{s_{k+1}}\right) + \sum_{i=r+1}^{M-1}(s_i - s_{i+1})\left(\frac{1}{s_r} - \frac{1}{s_{r+1}}\right)\left(\frac{1}{s_k} - \frac{1}{s_{k+1}}\right) \\
&+\ s_M\left(\frac{1}{s_r} - \frac{1}{s_{r+1}}\right)\left(\frac{1}{s_k} - \frac{1}{s_{k+1}}\right) \\
&= \left(\frac{1}{s_k} - \frac{1}{s_{k+1}}\right)\left(1 - \frac{s_{r+1}}{s_r} + s_{r+1}\left(\frac{1}{s_r} - \frac{1}{s_{r+1}}\right)\right) = 0.
\end{aligned}
$$

By a mirror argument, for a fixed $M+1 \leqslant k < r$, the coefficient of $\mathbb{P}(M_{\{k:n\}}, M_{\{r:n\}})$ is 0. □

Remark 4.5.16. *$\hat{m}$ can be computed using the Lauritzen formula:*

$$
\hat{m}^{-1} = \sum_{i=1}^{n-1}\left((m_{\{i:i+1\}})^{-1}\right)^0 - \sum_{i=2}^{n-1}\left((m_{ii})^{-1}\right)^0.
$$

Alternatively, to compute the missing entries $\hat{m}_{ij}$ for non-adjacent i and j, one can use the formula (Letac and Massam, 2007, p.1279):

$$
\hat{m}_{ij} = \hat{m}_{i,V\setminus\{i,j\}}(\hat{m}^{-1}_{V\setminus\{i,j\},V\setminus\{i,j\}})\hat{m}_{V\setminus\{i,j\},j} \tag{4.62}
$$

For $n = 3$, only $\hat{m}_{13}$ needs to be computed and (4.62) gives: $\hat{m}_{13} = m_{12}m_{22}^{-1}m_{23}$. But for $n > 3$, (4.62) does not give the missing elements directly. For example for $n = 4$, we need $\hat{m}_{13}$, $\hat{m}_{14}$ and $\hat{m}_{24}$ to complete m. Formula (4.62) gives

$$
\hat{m}_{13} = (m_{12}, \hat{m}_{14})\begin{pmatrix} m_{22} & \hat{m}_{24} \\ \hat{m}_{24} & m_{44}\end{pmatrix}^{-1}\begin{pmatrix} m_{23} \\ m_{43}\end{pmatrix} \tag{4.63}
$$

$$
\hat{m}_{14} = (m_{12}, \hat{m}_{13})\begin{pmatrix} m_{22} & m_{23} \\ m_{23} & m_{33}\end{pmatrix}^{-1}\begin{pmatrix} m_{23} \\ m_{34}\end{pmatrix} \tag{4.64}
$$

$$\hat{m}_{24} = (m_{21}, m_{23}) \begin{pmatrix} m_{11} & \hat{m}_{13} \\ \hat{m}_{13} & m_{33} \end{pmatrix}^{-1} \begin{pmatrix} \hat{m}_{14} \\ m_{34} \end{pmatrix} \tag{4.65}$$

The matrix m is completed by by solving the system of three equations with three unknowns above.

An extensive literature exists on methods of computation $\hat{m}$ and iterative algorithms are available (Grone et al., 1984; Paulsen et al., 1989; Johnson, 1990; Laurent, 1998; Glunt et al., 1999). The problem is sometimes referred to as positive definite matrix completion with maximum determinant constraint.

In the next Corollary, we consider $\underline{s} = p\mathbf{1}, p > 1/2$. We note that $\delta^{(M)}_{p\mathbf{1}}$ and $\gamma^{(M)}_{p\mathbf{1},y} := \gamma_{p,y}$ do not depend on M.

Corollary 4.5.17. *The variance function of the Wishart exponential family $\gamma_{p,y}$ is*

$$V(m) = \frac{1}{p}\,\mathbb{P}(\hat{m}).$$

A relation between the inverse mean map and $m_{\frac{1}{\underline{s}}}$

Recall that for the classical Wishart exponential families $W_{s\mathbf{1},y}$ on the symmetric cone S_n^+ the bijection between the cone Q_G and P_G is given by $L(m) = m^{-1}$. The mean map is $m_s(y) = sy^{-1}$ and the inverse mean map $\psi_s(m) = sm^{-1}$. It follows that

$$\psi_s = L \circ m_{\frac{1}{\underline{s}}} \circ L,$$

that is, the maps ψ_s and $m_{\frac{1}{\underline{s}}}$ are intertwined by the bijection L.

An analogous property holds on the cone Q_{A_n}, with the intertwiner given by the Lauritzen map. The bijection $L : Q_{A_n} \to P_{A_n}$ is the Lauritzen map $L(m) = (\hat{m})^{-1}$. Its inverse $L^{-1} : P_{A_n} \to Q_{A_n}$ is $L^{-1}(y) = \pi(y^{-1})$.

Proposition 4.5.18. *The inverse mean map* $\psi_{\underline{s}}^{(M)} : Q_G \to P_G$ *satisfies*

$$\psi_{\underline{s}}^{(M)} = L \circ m_{\frac{1}{\underline{s}}}^{(M)} \circ L.$$

Equivalently, for any $m \in Q_G$, $\pi(\psi_{\underline{s}}^{(M)}(m)^{-1}) = m_{\frac{1}{\underline{s}}}^{(M)}(\hat{m}^{-1})$.

$$\begin{array}{ccc} Q_{A_n} & \xrightarrow{\psi_{\underline{s}}^{(M)}} & P_{A_n} \\ {\scriptstyle L}\downarrow & & \uparrow{\scriptstyle L} \\ P_{A_n} & \xrightarrow[m_{\frac{1}{\underline{s}}}^{(M)}]{} & Q_{A_n} \end{array}$$

Proof. Using formula (4.34) of the mean function and definition (4.52) of $M_{\{1:i\}}$ and $M_{\{i:n\}}$, we see that $m_{\frac{1}{\underline{s}}}^{(M)}(\hat{m}^{-1})$ equals

$$\pi\left(\sum_{j=1}^{M-1}\left(\frac{1}{s_j}-\frac{1}{s_{j+1}}\right)M_{\{1:j\}}+\frac{\hat{m}}{s_M}+\sum_{j=M+1}^{n}\left(\frac{1}{s_j}-\frac{1}{s_{j-1}}\right)M_{\{j:n\}}\right).$$

Confronting this result with (4.59), we obtain $m_{\frac{1}{\underline{s}}}^{(M)}(\hat{m}^{-1}) = \pi\left(\psi_{\underline{s}}^{(M)}(m)^{-1}\right)$.

□

4.6 Wishart exponential families on the cone P_{A_n}.

The Diaconis-Ylvisaker conjugate family (Diaconis and Ylvisaker, 1979; Gutiérrez-Peña and Smith, 1997) for the exponential family of Wishart distributions $\gamma_{\underline{s},y}^{(M)}(x)$ is

$$\pi_{\underline{s},\eta,\beta}(y) = e^{-\langle\eta,y\rangle-\beta\ln\Delta_{-\underline{s}}^{(M)}(y)-\ln\int_{P_{A_n}}e^{-\langle\eta,y\rangle}\Delta_{\beta\underline{s}}^{(M)}(y)dy}dy.$$

Using Theorem 4.4.2, we obtain, for all $\underline{s}' \in \mathbb{R}^n$ such that $s'_M > -1$ and $s'_i > -3/2$, $i \neq M$,

$$\begin{aligned}\pi_{\underline{s},\eta,\beta}(y) &= c_{\beta\underline{s}}e^{-\langle\eta,y\rangle}\Delta_{\beta\underline{s}}^{(M)}(y)\delta_{-\beta\underline{s}}^{(M)}(\eta)\varphi_{Q_{A_n}}(\eta)dy = c_{\underline{s}'}e^{-\langle\eta,y\rangle}\Delta_{\underline{s}'}^{(M)}(y)\delta_{-\underline{s}'}^{(M)}(\eta)\varphi_{Q_{A_n}}(\eta)dy \\ &:= \tilde{\gamma}_{\underline{s}',\eta}^{(M)}(y),\end{aligned}$$

where

$$c_{\underline{s}}^{-1} = \pi^{(n-1)/2}\Big\{\prod_{i\neq M}\Gamma(s_i+\frac{3}{2})\Big\}\Gamma(s_M+1).$$

We call the probability distribution corresponding to the density function

$$\tilde{\gamma}_{\underline{s},x}^{(M)}(dy) = c_{\underline{s}}e^{-\langle x,y\rangle}\Delta_{\underline{s}}^{(M)}(y)\delta_{-\underline{s}}^{(M)}(x)\varphi_{Q_{A_n}}(x)1_{P_{A_n}}(y)dy \tag{4.66}$$

the Wishart distribution on P_{A_n}.

For a fixed $\underline{s}$, $\tilde{\gamma}_{\underline{s},x}^{(M)}$ is an exponential family generated by the measure $\tilde{R}_{\underline{s}}^{(M)}(dy) = C_{\underline{s}}\Delta_{\underline{s}}^{(M)}(y)dy$ with Laplace transform

$$\mathcal{L}_{\tilde{R}_{\underline{s}}^{(M)}}(x) = \int_{P_{A_n}} e^{-\langle x,y\rangle}\tilde{R}_{\underline{s}}^{(M)}(dy) = \delta_{-\underline{s}}^{(M)}(x)\varphi_{Q_{A_n}}(x). \tag{4.67}$$

Its Laplace transform is

$$\mathcal{L}_{\tilde{\gamma}_{\underline{s},x}^{(M)}}(\theta) = \int_{P_{A_n}} e^{-\langle\theta,y\rangle}\tilde{\gamma}_{\underline{s},x}^{(M)}(dy) = \frac{\mathcal{L}_{\tilde{R}_{\underline{s}}^{(M)}}(\theta+x)}{\mathcal{L}_{\tilde{R}_{\underline{s}}^{(M)}}(x)} = \frac{\delta_{-\underline{s}}^{(M)}(\theta+x)\varphi_{Q_{A_n}}(\theta+x)}{\delta_{-\underline{s}}^{(M)}(x)\varphi_{Q_{A_n}}(x)}. \tag{4.68}$$

4.6.1 Mean and covariance

Theorem 4.6.1. *The mean function of the Wishart exponential family on P_{A_n} is for all $s_i > -\frac{3}{2}$ and $x \in Q_{A_n}$,*

$$\begin{aligned}\tilde{m}_{\underline{s}}^{(M)}(x) = &\sum_{i=1}^{M-1}(s_i+\tfrac{3}{2})(x_{\{i:i+1\}}^{-1})^0 + \sum_{i=M+1}^{n}(s_i+\tfrac{3}{2})(x_{\{i-1:i\}}^{-1})^0 \\ &- \sum_{i=2}^{M-1}(s_{i-1}+1)(x_{ii}^{-1})^0 - (s_{M-1}-s_M+s_{M+1}+1)(x_{MM}^{-1})^0 \\ &- \sum_{i=M+1}^{n-1}(s_{i+1}+1)(x_{ii}^{-1})^0.\end{aligned} \tag{4.69}$$

The covariance function $\tilde{v}(x) : I_{A_n} \to Z_{A_n}$ of the Wishart exponential family on P_{A_n} equals

$$\tilde{v}(x) = \sum_{i=1}^{M-1}(s_i+\tfrac{3}{2})\mathbb{P}\left[(x_{\{i:i+1\}}^{-1})^0\right] + \sum_{i=M+1}^{n}(s_i+\tfrac{3}{2})\mathbb{P}\left[(x_{\{i-1:i\}}^{-1})^0\right]$$

$$- \sum_{i=2}^{M-1} (s_{i-1}+1)\mathbb{P}\left[(x_{ii}^{-1})^0\right] - (s_{M-1}-s_M+s_{M+1}+1)\mathbb{P}\left[(x_{MM}^{-1})^0\right]$$

$$- \sum_{i=M+1}^{n-1} (s_{i+1}+1)\mathbb{P}\left[(x_{ii}^{-1})^0\right],$$

where we identify I_{A_n} with Z_{A_n} by the trace inner product.

Proof. We have $\tilde{m}_{\underline{s}}^{(M)}(x) = -\nabla_x \ln L_{\mu_{\underline{s}}^{(M)}}(x) = -\nabla_x \ln \delta_{-\underline{s}}^{(M)}(x)\varphi_{Q_{A_n}}(x)$. The covariance operator is obtained by differentiation of (4.69). □

4.7 Relations with the type I and type II Wishart distributions of Letac and Massam (2007)

In this section, we will explain the relation between our work and type 1 and type 2 Wishart distributions constructed by Letac and Massam (2007).

Letac and Massam (2007) introduced, studied and used the function $H(\alpha, \beta, x)$ on Q_G as a generalized power function for constructing type I and type II Wishart distributions. The reader is referred to the cited paper for the general definition of the function $H(\alpha, \beta, x)$ as well as for graphical theoretic notions such as cliques, separators and perfect order of cliques (see also Lauritzen (1996)). For our purpose, it is sufficient to recall that for $\alpha \in \mathbb{R}^{n-1}$ and $\beta \in \mathbb{R}^{n-2}$

$$H(\alpha, \beta; x) = \frac{\prod_{i=1}^{n-1} |x_{\{i,i+1\}}|^{\alpha_i}}{\prod_{i=2}^{n-1} x_{ii}^{\beta_i}}, \quad x \in Q_{A_n}, \tag{4.70}$$

that the cliques (i.e. the sets of vertices of maximal complete subgraphs) are $\{1,2\}, \ldots,$ $\{n-1, n\}$ and the separators $\{2\}, \ldots, \{n-1\}$. The definition of the function $H(\alpha, \beta; x)$ does not include any restrictions on the values of the parameter (α, β) of dimension $2n-3$.

However, the existence of type I Wishart distributions on Q_G is only showed for (α, β) belonging to some set A_P dependent on a perfect order of cliques P, i.e. for $(\alpha, \beta) \in \mathcal{A}_0 =$

$\cup_P A_P$, where the union is on all perfect order of cliques. Proposition 4.7.3 describes this set for A_n graphs. It also makes clear a phenomenon observed by Letac and Massam (2007) for the graph A_4, where there are only two different sets A_P although there are 4 perfect orders of cliques. To prove Proposition 4.7.3 we use the following explicit relation between two concepts: perfect orders of cliques used by Letac and Massam (2007) and perfect elimination orders of vertices used in this work.

Definition 4.7.1. Let P be a perfect order $P : C'_1 < C'_2 < \ldots < C'_{n-1}$ on $\mathcal{C}$ and let $S'_2, \ldots, S'_{n-1}$ be the sequence of separators associated to this order. Let $\alpha : \mathcal{C} \mapsto \mathbb{R}$ and $\beta : \mathcal{S} \mapsto \mathbb{R}$ be two real functions of cliques and separators. The pair (α, β) belongs to A_P if

(1) $\alpha(C'_k) = \beta(S'_k)$ for $k = 3, \ldots, n-1$

(2) $\alpha(C) > \frac{1}{2}$ for all $C \in \mathcal{C}$

(3) $\alpha(C'_1) + \alpha(C'_2) - \beta(S'_2) > 0$.

Proposition 4.7.2. *Let $G = A_n : 1 - 2 - 3 - \ldots - n$. A clique ordering $C'_1 < \ldots < C'_{n-1}$ is perfect if and only if $C'_{n-1} \prec \ldots \prec C'_1$ is a perfect elimination order on the A_{n-1} graph $G' : C_1 - C_2 \ldots - C_{n-1}$. There are 2^{n-2} perfect orders of cliques on A_n.*

Proof. The proof is in two parts, for the two inclusions of the claimed equality. Both parts are straightforward and based on the definitions of a perfect order of cliques and of a perfect elimination order on a graph. We omit the details. □

Proposition 4.7.3. *Let $P' : C'_1 < C'_2 < \ldots < C'_{n-1}$ and $P'' : C''_1 < C''_2 < \ldots < C''_{n-1}$ be two perfect orders of cliques on $G = A_n$. Let S'_2 and S''_2 be the first separators of P' and P'', respectively. If $S'_2 = S''_2$ then $A_{P'} = A_{P''}$, i.e. the parameter set A_P depends only on the first separator S_2 with respect to the clique order P. If $S_2 = \{M\}$ then the set A_P is*

described by the conditions:

$$\begin{cases} \alpha_j = \beta_{j+1} \text{ if } 1 \leqslant j \leqslant M-2, \\ \alpha_j = \beta_j \text{ if } M+1 \leqslant j \leqslant n-1, \end{cases} \tag{4.71}$$

and

$$\alpha_j > \frac{1}{2} \text{ for all } 1 \leqslant j \leqslant n-1;\ \alpha_{M-1} + \alpha_M - \beta_M > 0. \tag{4.72}$$

Thus $\mathcal{A}_0 = \cup_P A_P$ is the set of (α, β) such that there exists $2 \leqslant M \leqslant n-1$ for which (4.71) and (4.72) are satisfied.

Proof. We use Propositions 4.2.5 and 4.7.2. □

The reference measure μ_G used by Letac and Massam (2007) is, on the cone Q_{A_n},

$$\mu_{A_n}(x)(dx) = H_{A_n}(-\frac{3}{2}\mathbb{1}, -\mathbb{1}; x)1_{Q_{A_n}}(x)dx. \tag{4.73}$$

By (4.17), we observe that $\mu_{A_n}(x)(dx) = \varphi_{Q_{A_n}}(x)1_{Q_{A_n}}(x)dx$. Namely, the reference measure μ_{A_n} is the characteristic measure of the cone Q_{A_n}.

Theorem 4.7.4. *(Letac and Massam (2007) Theorem 3.3) If $(\alpha, \beta) \in \mathcal{A}_0$, then, for a constant $\Gamma_{1(\alpha,\beta)}$, and for all $y \in P_{A_n}$*

$$\int_{Q_{A_n}} e^{-\operatorname{Tr}(xy)} H(\alpha, \beta; x)\mu_{A_n}(x)(dx) = \Gamma_{1(\alpha,\beta)} H(\alpha, \beta; \pi(y^{-1})).$$

The methods developed in this thesis give a new simple proof of Theorem 4.7.4, see the proof of Corollary 4.7.6 below.

Let us compare now the functions $H(\alpha, \beta; x)$ and $H(\alpha, \beta; \pi(y^{-1}))$ with the generalized power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$.

Proposition 4.7.5. *1. Let $\alpha \in \mathbb{R}^{n-1}$ and $\beta \in \mathbb{R}^{n-2}$. There exists $\underline{s} \in \mathbb{R}^n$ such that $H(\alpha, \beta; x) = \delta_{\underline{s}}^{(M)}(x)$ if and only if (4.71) holds for some $2 \leqslant M \leqslant n-1$.*

Then $s_j = \alpha_j$ if $1 \leqslant j \leqslant M-1$, $s_M = \alpha_{M-1} + \alpha_M - \beta_M$ and $s_j = \alpha_{j-1}$ if $M+1 \leqslant j \leqslant n$.

2. Moreover, under the hypothesis of Part 1, we have $H(\alpha, \beta; \pi(y^{-1})) = \Delta_{-\underline{s}}^{(M)}(y)$.

Proof. The equality of $H(\alpha, \beta; x)$ and $\delta_{\underline{s}}^{(M)}(x)$ is easily verified by confronting their definitions (4.70) and (4.4). Part 2 follows from Theorem 4.2.9. □

Corollary 4.7.6. *The type I Wishart distributions indexed by the set $\mathcal{A}_0$ are equal to the subset $\bigcup_{M=2}^{n-1}(\gamma_{\underline{s},y}^{(M)})_{y \in P_G}$ of Wishart NEF families defined in Section 4.5. Thus they are strictly contained in the set of all Wishart NEF families on Q_{A_n}, equal to $\bigcup_{M=1}^{n}(\gamma_{\underline{s},y}^{(M)})_{y \in P_{A_n}}$.*

Proof. It is a direct application of Proposition 4.7.5 and Theorem 4.4.1. Note that Theorem 4.4.1 implies Theorem 4.7.4 of Letac and Massam (2007).

The family of functions $H(\alpha, \beta, x)$ does not contain the power functions $\delta_{\underline{s}}^{(1)}$ or $\delta_{\underline{s}}^{(n)}$. In fact, the last functions contain powers of $n-1$ diagonal elements x_{ii}, whereas the function $H(\alpha, \beta, x)$ contains powers of $n-2$ such elements. □

Similar comparisons can be done on the cones P_{A_n}. In this case, Letac and Massam (2007) define type II Wishart distributions on P_{A_n} indexed by a set $\mathcal{B}_0$, analogous to the set $\mathcal{A}_0$ for Q_{A_n}. Similar arguments as on the cone Q_{A_n} lead to

Corollary 4.7.7. *The type II Wishart distributions on P_{A_n} indexed by the set $\mathcal{B}_0$ are equal to the subset $\bigcup_{M=2}^{n-1}(\tilde{\gamma}_{\underline{s},x}^{(M)})_{x \in Q_{A_n}}$ of Wishart NEF families defined in Section 4.6. Thus they are strictly contained in the set of all Wishart NEF families on P_{A_n}, equal to $\bigcup_{M=1}^{n}(\tilde{\gamma}_{\underline{s},x}^{(M)})_{x \in Q_{A_n}}$.*

4.8 Appendix

We list here some properties of triangular matrices, used in proofs.

Lemma 4.8.1. *1. Let $A = K^0$, where $K = A_{\{1:k\}}$ and let L be lower triangular and U upper triangular $n \times n$ matrices. Then $UAL = \left(U_{\{1:k\}} K L_{\{1:k\}}\right)^0$.*

2. Let B, L, U be matrices $n \times n$, with L lower triangular and U upper triangular. Then, for all $i = 1, \ldots, n$, $(LBU)_{\{1:i\}} = L_{\{1:i\}} B_{\{1:i\}} U_{\{1:i\}}$ and $(UBL)_{\{i:n\}} = U_{\{i:n\}} B_{\{i:n\}} L_{\{i:n\}}$.

3. If T is an invertible triangular matrix then $(T_{\{1:k\}})^{-1} = (T^{-1})_{\{1:k\}}$ for all $k = 1, \ldots, n$.

All these properties are elementary and easy to prove, by block multiplication of matrices (1,2) or by inverse matrix formula with cofactors (3).

Chapter 5

ON THE LETAC-MASSAM CONJECTURE

5.1 Introduction

In this chapter, we solve on an important class of cones, the conjecture stated by Letac and Massam in (Letac and Massam, 2007, p.1314), and called "Letac-Massam conjecture" in (Ben-David and Rajaratnam, 2014). This conjecture on the set parameters of type I and type II Wishart distributions is of fundamental importance in harmonic analysis of Riesz and Wishart measures on convex cones connected to graphs and in its applications to modern multivariate statistics. More generally, the Letac-Massam conjecture is closely related to an important problem in a wide range of analysis on cones:

$(\mathcal{P})$ *Is the Laplace transform of a product of powers of given polynomials equal to a product of powers of some polynomials?*

According to (Letac and Massam, 2007, Corollary 3.1), the Letac-Massam conjecture is true on the cones Q_{A_4} and P_{A_4}, but these results are "obtained by a nontrivial and long computation" and the proofs are omitted. Letac and Massam (2007) states that for $n = 5$ "calculations are terrifying." Our method of proof is simple and based on tools introduced in Chapter 4: triangular changes of variables on Q_{A_n} and using natural "future" and "past" power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$ on Q_{A_n} and on P_{A_n}. We show that the Letac-Massam conjecture is true on the cones Q_{A_n} and P_{A_4}.

5.2 Letac-Massam conjecture on Q_{A_n}

The Letac-Massam Conjecture is a conjecture on the Laplace transform of functions $\eta \mapsto H(\alpha, \beta, \eta)$, $\eta \in Q_{A_n}$, $\alpha = (\alpha_1, \ldots, \alpha_{n-1})$, $\beta = (\beta_2, \ldots, \beta_{n-1})$, introduced in Letac and Massam (2007). When needed, we will use a more precise notation H_n for the function H on Q_{A_n}. Let $\mu_{A_n}(d\eta)$ be the reference measure on the cone Q_{A_n}, defined in Letac and Massam (2007) by

$$\mu_{A_n}(d\eta) = \varphi_{Q_{A_n}}(\eta)d\eta = \prod_{i=1}^{n-1} |\eta_{\{i,i+1\}}|^{-3/2} \prod_{i\neq 1,n} \eta_{ii} d\eta, \tag{5.1}$$

where $d\eta$ is the Lebesgue measure on Q_{A_n}.

The Letac-Massam conjecture on the cone Q_{A_n} says that there exists $C_{\alpha,\beta} > 0$ such that

$$\int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)} H(\alpha, \beta, \eta) d\mu_{A_n}(\eta) = C_{\alpha,\beta} H(\alpha, \beta, \pi(y^{-1})) \qquad (y \in P_{A_n}) \tag{5.2}$$

if and only if $(\alpha, \beta) \in \mathcal{A}$, where $\mathcal{A} = \bigcup_{M=2}^{n-1} A_M$ and the sets A_M are defined by the following conditions (C) and (I):

(C) $\alpha_j = \beta_{j+1}$ *if* $1 \leqslant j \leqslant M-2$, and $\alpha_j = \beta_j$ *if* $M+1 \leqslant j \leqslant n-1$,

(I) $\alpha_j > \frac{1}{2}$ for all $j = 1, \ldots, n-1$, and $\alpha_{M-1} + \alpha_M - \beta_M > 0$.

The sufficiency of the condition $(\alpha, \beta) \in \mathcal{A}$ was shown in Letac and Massam (2007) and the necessity conjectured and proved true for $n = 4$. For $n = 2$ and $n = 3$ the equivalence of (5.2) with $(\alpha, \beta) \in \mathcal{A}$ is well known. The necessity of (I) is evident (consider diagonal $y \in P_{A_n}$, cf. Lemma 5.3.1 below), so the necessity of (C) is to be proved for $n \geqslant 4$.

In the sequel, the equality (5.2) will be referred to as the Letac-Massam formula on Q_{A_n} and the conditions (C) as Letac-Massam conditions. The main result of this chapter is the following:

Theorem 5.2.1. *Let $n \geqslant 4$. The formula (5.2) implies conditions* (C).

5.2.1 Letac-Massam conjecture in terms of power functions $\delta_{\underline{s}}^{(M)}$ and $\Delta_{\underline{s}}^{(M)}$

Now we recall the power functions $\delta_{\underline{s}}^{(M)}$ on Q_{A_n} and $\Delta_{\underline{s}}^{(M)}$ on P_{A_n}. For all $2 \leqslant M \leqslant n-1$,

$$\begin{aligned}
\delta_{\underline{s}}^{(M)}(\eta) &= \frac{\prod_{i=1}^{M-1} |\eta_{\{i:i+1\}}|^{s_i} \prod_{i=M+1}^{n} |\eta_{\{i-1:i\}}|^{s_i}}{\prod_{i=2}^{M-1} \eta_{ii}^{s_{i-1}} \eta_{MM}^{s_{M-1}-s_M+s_{M+1}} \prod_{i=M+1}^{n-1} \eta_{ii}^{s_{i+1}}} \\
\Delta_{\underline{s}}^{(M)}(y) &= \prod_{i<M} |y_{\{1:i\}}|^{s_i-s_{i+1}} |y|^{s_M} \prod_{i>M} |y_{\{i:n\}}|^{s_i-s_{i-1}},
\end{aligned}$$

where, for $I \subset \{1, \ldots, n\}$, the matrix A_I is the submatrix of A indexed by I, and the symbol $\{a : b\}$ with $1 \leqslant a \leqslant b \leqslant r$ denotes the set of i for which $a \leqslant i \leqslant b$.

Define $r_i = \alpha_{i+1} - \beta_i$ and $p_i = \alpha_i - \beta_i$ for all $2 \leqslant i \leqslant n-1$. We have, as defined in Letac and Massam (2007),

$$H(\alpha, \beta, \eta) = \frac{\prod_{i=1}^{n-1} |\eta_{\{i:i+1\}}|^{\alpha_i}}{\prod_{i=2}^{n-1} \eta_{ii}^{\beta_i}} \tag{5.3}$$

so that $H(\alpha, \beta, \eta) = \delta_{\underline{s}}^{(M)}(\eta) \prod_{i=2}^{M-1} \eta_{ii}^{r_i} \prod_{i=M+1}^{n-1} \eta_{ii}^{p_i}$, where $s_i = \alpha_i$, for all $1 \leqslant i \leqslant M-1$; $s_i = \alpha_{i-1}$, for all $M+1 \leqslant i \leqslant n$ and $s_M = \alpha_{M-1} + \alpha_M - \beta_M$. This implies $r_M = s_M - s_{M+1}$ and $p_M = s_M - s_{M-1}$. We notice that $\underline{s} = (s_i)$ depends on M, whereas neither $\underline{r} = (r_i)$ nor $\underline{p} = (p_i)$ does.

Let $\varphi(y) = \pi(y^{-1})$.

The Letac-Massam formula (5.2) is equivalent, for each $2 \leqslant M \leqslant n-1$, to

$$\begin{aligned}
&\int_{Q_{A_n}} e^{-\operatorname{Tr}(y\eta)} \delta_{\underline{s}}^{(M)}(\eta) \prod_{i=2}^{M-1} \eta_{ii}^{r_i} \prod_{i=M+1}^{n-1} \eta_{ii}^{p_i} d\mu_{A_n}(\eta) \\
&\qquad = C_{\alpha,\beta} \Delta_{-\underline{s}}^{(M)}(y) \prod_{i=2}^{M-1} \varphi(y)_{ii}^{r_i} \prod_{i=M+1}^{n-1} \varphi(y)_{ii}^{p_i}.
\end{aligned} \tag{5.4}$$

The Letac-Massam conditions (C) are equivalent to the following $n-2$ alternative conditions:

$$r_2 = \cdots = r_{M-1} = p_{M+1} = \cdots = p_{n-1} = 0 \tag{5.5}$$

for an $M \in \{2, \dots, n-1\}$, or, in other words, to the equality $H(\alpha, \beta, \cdot) = \delta_{\underline{s}}^{(M)}$ for an $M \in \{2, \dots, n-1\}$.

A positive answer to the Letac-Massam conjecture implies that the natural generalized power functions on Q_{A_n} are the families $\delta_{\underline{s}}^{(M)}(\eta)$, motivated by analysis on symmetric and homogeneous cones, with n-dimensional parameter $\underline{s}$. Power functions $H(\alpha, \beta, \eta)$, $\eta \in Q_{A_n}$ are motivated by advanced graph theory, more exactly by cliques and separators formalism. The parameters α, β have dimension $2n-3$. Even if we start with a larger family $H(\alpha, \beta, \eta)$, in order to have the property $(\mathcal{P})$ satisfied, we boil down to the families $\delta_{\underline{s}}^{(M)}(\eta)$, with $M = 2, \dots, n-1$. Moreover, the families $\delta_{\underline{s}}^{(1)}(\eta)$ and $\delta_{\underline{s}}^{(n)}(\eta)$ are "forgotten" in the graph theory approach of Letac and Massam (2007).

5.3 Proof

We are going to prove the Letac-Massam conjecture by induction on n. The proof of the initiation part ($n = 4$) and the heredity part ($n \geqslant 5$) are the same, so they are given together.

First, in the following lemma, we express, for each M, the constant $C_{\alpha,\beta}$ as a function of $M, \underline{s} = (s_i), \underline{r} = (r_i)$ and $\underline{p} = (p_i)$. This is convenient and needed in further study of the formula (5.4).

Lemma 5.3.1. *If the formula (5.4) holds for all $y \in P_{A_n}$ then we have*

$$C_{\alpha,\beta} = \pi^{(n-1)/2}\Gamma(s_M)\Big\{\prod_{i\neq M}\Gamma(s_i - \frac{1}{2})\Big\}\prod_{i=2}^{M-1}\frac{\Gamma(s_i + r_i)}{\Gamma(s_i)}\prod_{i=M+1}^{n-1}\frac{\Gamma(s_i + p_i)}{\Gamma(s_i)}. \tag{5.6}$$

If y is diagonal, then (5.4) holds if and only if $s_i > \frac{1}{2}$ for $i \neq M$, $s_M > 0$, $s_i + r_i > 0$ for $2 \leqslant i < M$, and $s_i + p_i > 0$ for $M < i \leqslant n-1$.

Proof. We take y diagonal. The proof is a by-product of the Step 1 of the main proof. □

Step 1 (descent in Letac-Massam formula, from Q_{A_n} to $Q_{A_{n-1}}$). *Let $n \geqslant 4$, $\alpha = (\alpha_1, \dots, \alpha_{n-1})$ and $\beta = (\beta_2, \dots, \beta_{n-1})$.*

Suppose that the Letac-Massam formula (5.2) *holds for* $H_n(\alpha,\beta,\cdot)$ *on* Q_{A_n}. *Then the Letac-Massam formula holds on* $Q_{A_{n-1}}$ *for:*

(i) $H_{n-1}((\alpha_1,\dots,\alpha_{n-2}),(\beta_2,\dots,\beta_{n-2}),\cdot)$ *and the graph* $1-\cdots-(n-1)$

(ii) $H_{n-1}((\alpha_2,\dots,\alpha_{n-1}),(\beta_3,\dots,\beta_{n-1}),\cdot)$ *and the graph* $2-\cdots-n$.

Proof. Let us prove (i). We choose $2\leqslant M\leqslant n-2$. For all $y\in P_{A_n}$, let, successively, $y=\tilde{\Phi}_n(a',b',z)$ and $z=\Phi_{n-1}(a'',b'',Z)$. We easily check that for $2\leqslant i\leqslant n-1$, $\varphi(y)_{ii}=\varphi(z)_{ii}=\varphi(Z)_{ii}$, see Lemma 4.3.4 (by our convention, z is indexed by $1,\dots,n-1$ and Z is indexed by $2,\dots,n-1$). Integration on Q_{A_n} with two successive changes of variables $\eta=\tilde{\Psi}_n(\mu',\nu',\xi)$ and then $\xi=\Psi_{n-1}(\mu'',\nu'',\Xi)$ gives

$$\int_{Q_{A_{n-2}}} e^{-\operatorname{Tr}(Z\Xi)}\delta^{(M)}_{(s_2,\dots,s_{n-1})}(\Xi)\prod_{i=2}^{M-1}\Xi_{ii}^{r_i}\prod_{i=M+1}^{n-1}\Xi_{ii}^{p_i}d\mu_{A_{n-2}}(\Xi) \tag{5.7}$$

$$=C^{(n-2)}_{\alpha,\beta}\Delta^{(M)}_{-(s_2,\dots,s_{n-1})}(Z)\prod_{i=2}^{M-1}\varphi(Z)_{ii}^{r_i}\prod_{i=M+1}^{n-1}\varphi(Z)_{ii}^{p_i},$$

where $C^{(n-2)}_{\alpha,\beta}=\frac{C_{\alpha,\beta}}{\pi\Gamma(s_1-\frac{1}{2})\Gamma(s_n-\frac{1}{2})}$ and the rows and columns of Ξ and Z are numbered $2,\dots,n-1$. Now, we apply one more change of variable $\Xi=\tilde{\Psi}_{n-2}(\mu,\nu,\Theta)$ in formula (5.7) and we set $Z=\tilde{\Phi}_{n-2}(a,0,T)$. The lines and columns of Θ and T are numbered $2,\dots,n-2$.

Let $F(\mu,\nu,\Theta)$ be the integrand. We first compute $J=\int_{-\infty}^{\infty}\int_0^{\infty}Fd\mu d\nu=2\int_0^{\infty}\int_0^{\infty}Fd\mu d\nu$. Using the change of variables $u=a\mu, t=a\Theta_{n-2,n-2}\nu^2$, we get

$$J=2a^{-p_{n-1}}\int_0^{\infty}\int_0^{\infty}e^{-\left(a\mu+a\Theta_{n-2,n-2}\nu^2\right)}\mu^{s_{n-1}-3/2}(a\mu+a\Theta_{n-2,n-2}\nu^2)^{p_{n-1}}d\mu d\nu$$

$$=a^{-(s_{n-1}+p_{n-1})}\Theta^{-1/2}_{n-2,n-2}\int_0^{\infty}\int_0^{\infty}e^{-(u+t)}u^{s_{n-1}-\frac{3}{2}}t^{-\frac{1}{2}}(u+t)^{p_{n-1}}dudt.$$

Using the change of variables $(u,v)=(u,u+t)$, we get

$$J=a^{-(s_{n-1}+p_{n-1})}\Theta^{-1/2}_{n-2,n-2}\int_0^{\infty}\left(\int_0^v u^{s_{n-1}-\frac{3}{2}}(v-u)^{-\frac{1}{2}}du\right)e^{-v}v^{p_{n-1}}dv \tag{5.8}$$

$$= a^{-(s_{n-1}+p_{n-1})}\Theta_{n-2,n-2}^{-1/2}B(s_{n-1}-\frac{1}{2},\frac{1}{2})\Gamma(s_{n-1}+p_{n-1}),$$

where, in the integral with respect to du we made a change of variable $x = u/v$. We get

$$\int_{Q_{A_{n-3}}} e^{-\operatorname{Tr}(T\Theta)}\delta^{(M)}_{(s_2,\dots,s_{n-2})}(\Theta)\prod_{i=2}^{M-1}\Theta_{ii}^{r_i}\prod_{i=M+1}^{n-2}\Theta_{ii}^{p_i}d\mu_{A_{n-3}}(\Theta) \tag{5.9}$$

$$= C_{\alpha,\beta}^{(n-3)}\,\Delta^{(M)}_{-(s_2,\dots,s_{n-2})}(T)\prod_{i=2}^{M-1}\varphi(T)_{ii}^{r_i}\prod_{i=M+1}^{n-2}\varphi(T)_{ii}^{p_i}, \tag{5.10}$$

where

$$C_{\alpha,\beta}^{(n-3)} = \frac{C_{\alpha,\beta}}{\pi^{\frac{3}{2}}\Gamma(s_1-\frac{1}{2})\Gamma(s_n-\frac{1}{2})\Gamma(s_{n-1}-\frac{1}{2})}\frac{\Gamma(s_{n-1})}{\Gamma(p_{n-1}+s_{n-1})}. \tag{5.11}$$

Recall that throughout the thesis $C_{\alpha,\beta}$ denotes the constant from formulas (5.2) and (5.4). By the same argument as to obtain formula (5.7), we observe that the Letac-Massam formula for the function $H_{n-1}((\alpha_1,\dots,\alpha_{n-2}),(\beta_2,\dots,\beta_{n-2}),\cdot)$ on $Q_{A_{n-1}}$ and the graph $1-2-\cdots-(n-1)$ is equivalent to formula (5.9). This finishes the proof of (i).

By a similar 'mirror-like' argument with the change of variables $\Xi = \Psi_{n-2}(\mu,\nu,\Theta)$ in (5.7), we get the Letac-Massam formula for $H_{n-1}((\alpha_2,\dots,\alpha_{n-1}),(\beta_3,\dots,\beta_{n-1}),\cdot)$ and the graph $2-\cdots-n$, and we prove part (ii) of Step 1. □

Proof of Lemma 5.3.1. For y diagonal, formula (5.11) leads by induction to formula (5.6), observing that the last equation we get is $a^{-s_M}\int_0^\infty e^{-ax}x^{s_M}\frac{dx}{x} = C_{\alpha,\beta}^{(1)}a^{-s_M}$, so that $C_{\alpha,\beta}^{(1)} = \Gamma(s_M)$.

Step 2 (induction step). *The Letac-Massam conjecture on $Q_{A_{n-1}}$ implies the Letac-Massam conjecture on Q_{A_n}.*

Proof. Let $n \geqslant 4$. Suppose that the Letac-Massam formula (5.2) holds for some α and β and suppose that the Letac-Massam conjecture is true on $Q_{A_{n-1}}$.

For $n \geqslant 5$, we use Step 1 and the induction hypothesis. Thus one of the following $n-3$ conditions has to be satisfied:

$$r_2 = \cdots = r_{M-1} = p_{M+1} = \cdots = p_{n-2} = 0,$$

for an $M \in \{2, \cdots, n-2\}$, and, simultaneously, one of the following $n-3$ "shifted" conditions has to be satisfied:

$$r_3 = \cdots = r_M = p_{M+2} = \cdots = p_{n-1} = 0,$$

for an $M \in \{2, \ldots, n-2\}$. This implies that either conditions (5.5) are satisfied or

$$p_3 = \cdots = p_{n-2} = 0;\ r_3 = \cdots = r_{n-2} = 0. \tag{5.12}$$

Let us assume this single remaining case and show that it also implies conditions (5.5).

The equality $r_M = 0$ implies $s_M = s_{M+1}$ and $p_M = 0$ implies $s_M = s_{M-1}$. Also, from $p_j = r_j$ for all $3 \leqslant j \leqslant n-2$, we get $s_2 = \cdots = s_{M-1}$ and $s_{M+1} = \cdots = s_{n-1}$. Thus, $s_2 = \cdots = s_{n-1} = s$. In the case (5.12), using the cofactor formula for Z^{-1}, equation (5.7) reduces to

$$\begin{aligned} &\int_{Q_{A_{n-2}}} e^{-\operatorname{Tr}(Z\Xi)} \delta^{(M)}_{(s,\ldots,s)}(\Xi) \Xi_{22}^{r_2} \Xi_{n-1,\,n-1}^{p_{n-1}} d\mu_{A_{n-2}}(\Xi) \\ &= C^{(n-2)}_{\alpha,\beta} |Z|^{-s} \left(\frac{|Z_{\{3:n-1\}}|}{|Z|} \right)^{r_2} \left(\frac{|Z_{\{2:n-2\}}|}{|Z|} \right)^{p_{n-1}}. \end{aligned} \tag{5.13}$$

We apply the second derivative with respect to $Z_{n-2,n-1}$ on both sides of (5.13) and we take $Z_{n-2,n-1} = 0$. Theorem 2.7.1 in Lehmann and Romano (2005) ensures that the derivatives of all orders of the integral (5.13) can be computed under the integral sign. We obtain

$$\begin{aligned} &\int_{Q_{A_{n-2}}} e^{-\operatorname{Tr}(Z\Xi)} \delta^{(M)}_{(s,\ldots,s)}(\Xi) \Xi_{22}^{r_2} \Xi_{n-1,\,n-1}^{p_{n-1}} \Xi_{n-2,n-1}^{2} d\mu_{A_{n-2}}(\Xi) \big|_{Z_{n-2,n-1}=0} \\ &= \frac{C^{(n-2)}_{\alpha,\beta}}{4} \frac{\partial^2}{\partial Z^2_{n-2,n-1}} \big|_{Z_{n-2,n-1}=0} g(Z), \end{aligned} \tag{5.14}$$

where $g(Z) = |Z|^{-s} \left(\frac{|Z_{\{3:n-1\}}|}{|Z|}\right)^{r_2} \left(\frac{|Z_{\{2:n-2\}}|}{|Z|}\right)^{p_{n-1}}$.

Let us change the variables $\Xi = \tilde{\Psi}_{n-2}(\tilde{\mu}, \tilde{\nu}, \Theta)$ and set $Z = \tilde{\Phi}_{n-2}(a, 0, T)$, i.e. $Z_{n-2,n-1} = 0$. Similarly as in the proof of (5.9) in Step 1, we find that the left hand side of (5.14) is

$$a^{-(s+p_{n-1}+1)}\Gamma(s+p_{n-1}+1)B\left(s-\tfrac{1}{2},\,\tfrac{3}{2}\right)\int_{Q_{A_{n-3}}} e^{-\operatorname{Tr}(T\Theta)}\delta^{(M)}_{(s,\dots,s)}(\Theta)\Theta_{22}^{r_2}\Theta_{n-2,n-2}d\mu_{An-3}(\Theta). \tag{5.15}$$

We write $\frac{1}{4}C^{(n-2)}_{\alpha,\beta}D$ the right hand side of (5.14) and we compute D. Denoting $S = -(s + r_2 + p_{n-1})$ and $h(Z) = |Z|^S|Z_{\{3:n-1\}}|^{r_2}$ we have

$$D = |Z_{\{2:n-2\}}|^{p_{n-1}} \left.\frac{\partial^2}{\partial Z^2_{n-2,n-1}}\right|_{Z_{n-2,n-1}=0} h(Z).$$

We apply formulas

$$|Z| = Z_{n-1,n-1}|Z_{\{2:n-2\}}| - Z^2_{n-2,n-1}|Z_{\{2:n-3\}}|,$$

$$|Z|^S = (Z_{n-1,n-1}|Z_{\{2:n-2\}}|)^S(1 - S\frac{Z^2_{n-2,n-1}|Z_{\{2:n-3\}}|}{Z_{n-1,n-1}|Z_{\{2:n-2\}}|} + o(Z^2_{n-2,n-1})).$$

Thus, for $Z_{n-2,n-1} = 0$, we get $\frac{\partial |Z|^S}{\partial Z_{n-2,n-1}} = 0$ and

$\frac{\partial^2 |Z|^S}{\partial Z^2_{n-2,n-1}} = -2S(Z_{n-1,n-1}|Z_{\{2:n-2\}}|)^{S-1}|Z_{\{2:n-3\}}|.$

Similarly,

$$|Z_{\{3:n-1\}}| = Z_{n-1,n-1}|Z_{\{3:n-2\}}| - Z^2_{n-2,n-1}|Z_{\{3:n-3\}}|$$

(for $n = 5$ we set $|Z_{\{3:n-3\}}| = 1$) and

$$\left.\frac{\partial^2 |Z_{\{3:n-1\}}|^{r_2}}{\partial Z^2_{n-2,n-1}}\right|_{Z_{n-2,n-1}=0} = -2r_2(Z_{n-1,n-1}|Z_{\{3:n-2\}}|)^{r_2-1}|Z_{\{3:n-3\}}|.$$

Using $Z = \tilde{\Phi}_{n-2}(a, 0, T)$, where the matrix T is indexed by $2, \dots, n-2$, we obtain $Z_{n-1,n-1} = a$, $Z_{\{2:n-2\}} = T$, $Z_{\{3:n-2\}} = T_{\{3:n-2\}}$, $|Z_{\{3:n-1\}}| = a|T_{\{3:n-2\}}|$ and $|Z| = a|T|$. By Leibniz formula,

$$D = -2a^{r_2+S-1}|T|^{p_{n-1}+S-1}|T_{\{3:n-2\}}|^{r_2-1}\left(S|T_{\{3:n-2\}}||T_{\{2:n-3\}}| + r_2|T_{\{3:n-3\}}||T|\right),$$

where for $n = 5$ we set $|T_{\{3:n-3\}}| = 1$. Hence, for $Z_{n-2,n-1} = 0$, the right hand side of (5.14) is

$$\frac{C_{\alpha,\beta}^{(n-2)}}{2} a^{-(s+p_{n-1}+1)} |T|^{-(s+r_2+1)} |T_{\{3:n-2\}}|^{r_2-1} f(T), \tag{5.16}$$

where

$$f(T) = (s + r_2 + p_{n-1})|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - r_2|T_{\{3:n-3\}}||T| \,.$$

Equating (5.16) and (5.15), we obtain, using (5.11),

$$\int_{Q_{A_{n-3}}} e^{-\operatorname{Tr}(T\Theta)} \delta^{(M)}_{(s,\dots,s)} \Theta_{22}^{r_2} \Theta_{n-2,n-2} d\mu_{A_{n-3}}(\Theta) \frac{s d(s, r_2, T)}{s + p_{n-1}} f(T), \tag{5.17}$$

where

$$d(s, r_2, T) = C_{\alpha,\beta}^{(n-3)} |T|^{-(s+r_2+1)} |T_{\{3:n-2\}}|^{r_2-1}.$$

Formula (5.17) is supposed to be true for our $p_{n-1} = \alpha_{n-1} - \beta_{n-1}$. It is surely true for $p_{n-1} = 0$, because the Letac-Massam conditions (5.5) are then satisfied. Equating (5.17) for these two values of p_{n-1}, and noting that by (5.6) the constant $C_{\alpha,\beta}^{(n-3)}$ does not depend on p_{n-1}, we get

$$\begin{aligned} &\frac{(s + r_2 + p_{n-1})|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - r_2|T_{\{3:n-3\}}||T|}{s + p_{n-1}} \\ &= \frac{(s + r_2)|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - r_2|T_{\{3:n-3\}}||T|}{s}, \end{aligned}$$

which is equivalent to

$$r_2 p_{n-1} \left(|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - |T_{\{3:n-3\}}||T| \right) = 0, \tag{5.18}$$

where for $n = 5$ we set $|T_{\{3:n-3\}}| = 1$. We observe that $|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - |T_{\{3:n-3\}}||T| \neq 0$, for example for T such that $T_{ii} = 2$ for all $2 \leqslant i \leqslant n-2$, $T_{i,i+1} = T_{i+1,i} = 1$ for $2 \leqslant i \leqslant n-3$ and $T_{ij} = 0$ for all other i, j (in this case, this expression equals 1). Thus,

for $n \geqslant 5$, in the remaining case (5.12), we also have $r_2 = 0$ or $p_{n-1} = 0$. In both cases we fall in the Letac-Massam conditions (5.5) and the proof of the induction step is finished.

For $n = 4$, we get formula (5.7) for $M = 2$, the computations are simpler (no use of Leibniz formula is needed), and no condition $s_2 = s_3 = s$ appears. The analogue of formula (5.17) is

$$\Gamma(s_3 + p_3 + 1)B(s_3 - \frac{1}{2}, \frac{3}{2}) \int_0^\infty e^{-tu} u^{s_2} u \frac{1}{u} du \frac{C^{(2)}_{\alpha,\beta}}{2}(s_2 + p_3) t^{-(s_2+1)}, \quad t > 0. \tag{5.19}$$

After substitution of the constant

$$C^{(2)}_{\alpha,\beta} = \pi^{\frac{1}{2}} \Gamma(s_2)\Gamma(s_3 - \frac{1}{2}) \frac{\Gamma(s_3 + p_3)}{\Gamma(s_3)},$$

one gets $(s_3 + p_3)s_2 = s_3(s_2 + p_3)$ equivalent to $r_2 p_3 = 0$, so $r_2 = 0$ or $p_3 = 0$. We get the Letac-Massam conditions for Q_{A_4}. □

Remark 5.3.2. *The expression on the RHS of* (5.18)*, i.e.* $|T_{\{3:n-2\}}||T_{\{2:n-3\}}| - |T_{\{3:n-3\}}||T|$*, where* $T = T_{\{2:n-2\}}$ *is known in matrix theory. It is treated in Desnanot-Jacobi identity* (*(Bressoud, 1999, Thm 3.12)*)*, called also Lewis Caroll* (*or Dodgson's*) *identity* (*Chenevier and Renard (2008)*) *and is equal to* $(\prod_{i=2}^{n-3} T_{i,i+1})^2$*, the square of the monomial of the off-diagonal entries.*

Remark 5.3.3. *The same method applies in order to prove the Letac-Massam Conjecture on* P_{A_4}*. We take* $M = 2$ *and apply two changes of variables* Φ_4 *and* $\tilde{\Phi}_3$ *on* P_{A_4} *and* P_{A_3}*, (see Lemma 4.3.3). We obtain an integral on* $P_{A_2} = S_2^+$*, which is the same as the integral on* $Q_{A_2} = S_2^+$ *in the proof above. The work on the Letac-Massam Conjecture on* P_{A_n} *for* $n \geqslant 5$ *is in progress. The analysis on these cones is more difficult.*

Remark 5.3.4. *Our method of differentiating the Letac-Massam formula with respect to* Z_{12} *gives a simple proof of the "Mellin transform" Lemma 3.1 in (Letac and Massam, 2007, p. 1302), announced without proof. However, instead of the second derivative in* Z_{12}*, the complete Taylor expansion in* Z_{12} *is needed.*

5.4 Generalized Letac-Massam conjecture

In the first part of the proof of Theorem 5.2.1, we showed that the Letac-Massam formula (5.2) on Q_{A_n}, with $M \in \{2, \ldots, n-1\}$, is equivalent to a Laplace transform formula (5.4) on $Q_{A_{n-2}}$, for a function $\delta^{(M)}_{(s_2,\ldots,s_{n-1})}$. Next we proved that (5.4) implies that the formula is rewritten for an $M' \in \{2, \ldots, n-1\}$ with $r_i = 0 = p_j, i = 2, \ldots, M'-1, j = M'+1, \ldots, n-1$. Thus, in fact we showed a stronger result that we call Generalized Letac-Massam Conjecture (GLMC):

Theorem 5.4.1. *Let $M \in \{1, \ldots, n\}$. There exists a multi-index $\underline{s} \in \mathbb{R}^n$ and a constant $C > 0$ such that for all $y \in P_{A_n}$*

$$\int_{Q_{A_n}} e^{-\mathrm{Tr}(y\eta)} \delta^{(M)}_{\underline{s}}(\eta) \prod_{i=1}^{M-1} \eta_{ii}^{r_i} \prod_{i=M+1}^{n} \eta_{ii}^{p_i} d\mu_{A_n}(\eta) = C\, \Delta^{(M)}_{-\underline{s}}(y) \prod_{i=1}^{M-1} \varphi(y)_{ii}^{r_i} \prod_{i=M+1}^{n} \varphi(y)_{ii}^{p_i}$$

if and only if the formula is rewritten with $M' \in \{1, \ldots, n\}$ such that $r_i = 0 = p_j, i = 1, \ldots, M'-1, j = M'+1, \ldots, n$ and $s_i > \frac{1}{2}, i \neq M', s_{M'} > 0$.

The GLMC gives a partial answer to the question which products of powers of well-defined minors on Q_{A_n} have the property $(\mathcal{P})$.

5.5 Discussions on (Ben-David and Rajaratnam, 2014)

Recent work by Ben-David and Rajaratnam (2014) suggests that the Letac-Mass conjecture may not hold for some non-homogeneous graphs. The rationale of their work is the following: According to Ben-David and Rajaratnam (2014) the Letac-Massam conjecture implies that the number r_D of separators of G that are ancestral in the DAG (Directed Acyclic Graph) version D of G is such that $r_D \leqslant 1$; any graphs such that $r_D > 1$ would therefore constitute a counterexample to the Letac-Massam conjecture.

In this section, we prove that for all A_n graphs, we have $r_D = 1$ and therefore, A_n graphs are not members of those potential counterexamples suggested by Ben-David and

Rajaratnam (2014). To this end, we need to first quickly recall some concepts associated to directed acyclic graphs that we will use in this section. The reader is referred to Lauritzen (1996) for details on any other concepts not clearly defined here.

A perfect DAG is a DAG in which all parents of the same vertex are adjacent. A DAG version D of an undirected graph G is any perfect directed acyclic graph that yields the graph G when all directed edges are replaced by undirected ones.

Consider a decomposable graph $G = (V, \mathcal{E})$ with set of cliques $(C_1, \dots C_r)$ and let P be a perfect of order of the cliques of G. Set $H_1 = C_1$ and define for all $2 \leqslant i \leqslant r$, $H_i = H_{i-1} \cup C_i$ and $S_i = H_{i-1} \cap C_i$.

1. A subset A of V is said to be ancestral in a directed graph with set of vertices V if it contains all parents of all vertices in it.

2. A DAG version D of a decomposable graph G is said to be induced by a perfect order P of the cliques of G if $H_1, \dots, H_{r-1}$ are all ancestral in D.

From the above definitions, we have the following:
A DAG D is a version of G induced by P such that S_2 is ancestral if D can be obtained by replacing the undirected edges of G by directed edges such that

- for all $1 \leqslant i \leqslant r-1$, H_i contains all parents of all vertices in it;

- S_2 contains all parents of all vertices in it.

(Ben-David and Rajaratnam, 2014, Lemma 5.1) states that for all perfect order P of a decomposable graph G there exists a DAG version of G induced by P. So this should be true for A_n graphs in particular and we have the following result.

Lemma 5.5.1. *Consider an A_n graph and P a perfect order of the cliques of A_n. Let D be a DAG version of A_n induced by P. At most one separator $\{i\}$ of A_n is ancestral in D and in this case D is unique and is given by*

$$1 \leftarrow \cdots \leftarrow i-1 \leftarrow i \rightarrow i+1 \rightarrow \cdots \rightarrow n.$$

Proof. Let $\{i\}$, $2 \leqslant i \leqslant n$, be a separator of A_n which is ancestral in D. Then, for all $j \in V$, we cannot have $j \to i$. This means that D has the pattern $i-1 \leftarrow i \to i+1$.
Also, we cannot have $i-2 \longrightarrow i-1$ because then $i-2 \longrightarrow i-1 \longleftarrow i$ would constitute an immorality (the parents of $i-1$ are not adjacent) and we know from Lemma 5.1 in [Ben-David and Rajaratnam] that D is a perfect DAG (it has no immoralities). Repeating the same argument inductively shows that we cannot have $k-1 \longrightarrow k$ for all $k < i$. Therefore, we have $k-1 \longleftarrow k$, for all $1 < k < i$.
The same reasoning shows that we have $k \longrightarrow k+1$, for all $i < k < n$.
D is therefore the directed graph $1 \leftarrow \cdots \leftarrow i-1 \leftarrow i \to i+1 \to \cdots \to n$ and no other separator of A_n is ancestral in D. □

Theorem 5.5.2. *For A_n graphs ($n > 2$), the DAG version of A_n induced by a perfect order P, such that $S_2 = \{i\}$ is ancestral, is given by*

$$1 \leftarrow \cdots \leftarrow i-1 \leftarrow i \to i+1 \to \cdots \to n.$$

Moreover, we always have $r = n-1$ and $r_D = 1$.

Proof. The proof follows from Lemma 5.5.1. □

We have thus proved that A_n graphs are not members of those potential counterexamples suggested by (Ben-David and Rajaratnam, 2014).

Chapter 6

FISHER INFORMATION AND EXPONENTIAL FAMILIES PARAMETRIZED BY A SEGMENT OF MEANS

6.1 Introduction

Fisher information is a key concept in mathematical statistics. Its importance stems from the Cramér-Rao inequality which says that the covariance of any unbiased estimator $T(X_1, \ldots X_n)$ of an unknown parameter θ, is bounded by the inverse of the Fisher information: $\mathrm{Var}_\theta(T) - (I(\theta))^{-1}$ is semi-positive definite. Fisher information is therefore a measure of the maximum precision attainable in parameter estimation. The efficiency of an estimator is based on whether this precision is achieved. This justifies the use of Fisher information in experimental design for predicting the maximum precision an experiment can provide on model parameters. This also justifies the important role Fisher information plays in estimation theory where it provides bounds for confidence regions, and also in Bayesian analysis where it provides a basis for noninformative priors. Fisher information can be used to investigate the trade-off between parsimony of parameters and precision of the estimation of the parameters (Andersson and Handel, 2006). Besides its importance in statistical theory, Fisher information has different interpretations that lead

to some practical applications. For example, the interpretation of Fisher information as a measure of the state of disorder of a dynamic system leads to the use of Fisher information in stochastic optimal control as a tuning tool to stabilise the performance of a dynamic system (Ramirez et al., 2010). Viewing Fisher information as a measure of information, leads to the statement of a "minimum information principle" akin to the well-known maximum entropy principle for determining the "maximally unpresumptive distribution" satisfying some predefined constraints (Bercher and Vignat, 2009). Gupta and Kundu (2006) describe the use of Fisher information in model selection as a tool to discriminate between two models with otherwise very similar fit to some data. The use of Fisher information however goes far beyond statistics; Frieden (2004) explains that Fisher information is in fact a key concept in the unification of science in general, as it allows a systematic approach to deriving Lagrangians.

The objective of this chapter is the study of exponential families $(Q_m)_{m\in\mathcal{M}}$ parametrized by a segment of means $[m_1, m_2]$ with a particular emphasis on Fisher information. These models were first considered by Letac (2012). Exponential families of distributions are extensively used in statistics and intensively studied, cf. Lehmann and Casella (1998); Lehmann and Romano (2005); Letac (1992); Letac and Casalis (2000). They are the only models for which the Cramér-Rao bound is always attained. A parametrization of the family by a segment instead of the whole means domain allows to obtain a parsimonious model when the mean domain is high-dimensional. The parametrization of the mean parameter by a segment is particularly useful in practical situations when hesitating between two equally convenient mean values m_1 and m_2. Such parametrization will also serve in sequential data collection, when an updated estimate of a parameter largely differs from the previous estimate.

In this chapter, we prove explicit formulas for the Fisher information of $(Q_{\theta m_1+(1-\theta)m_2})_{\theta\in[0,1]}$ if the full model is either the multivariate Gaussian family of known mean and unknown covariance matrix or a family of Wishart distributions with unknown

scaling parameter.

The chapter is organised as follows. In Section 6.2, basic definitions and results on Fisher information and exponential families are recalled. Section 6.3 contains new results on the Fisher information of exponential Gaussian and Wishart sub-families parametrized by a segment of means $[m_1, m_2]$. When m_1 and m_2 are colinear, we construct efficient estimators for the segment parameter θ.

6.2 Preliminaries

Definition 6.2.1. Consider a σ-finite measurable space $(\Omega, \mathcal{A}, \nu)$ with a family of strictly positive probability density functions f_s, $s \in S \subset \mathbb{R}^d$ with respect to ν. Let $l_s = \ln f_s$. Assume that the function $s \mapsto l_s(\omega)$ is differentiable for every $\omega \in \Omega$. Consider the gradient l'_s of the map $s \mapsto l_s$ as a random vector on the statistical model $(\Omega, \mathcal{A}, f_s d\nu)$. Suppose that it satisfies $\mathbb{E}_s(\|l'_s\|^2) < \infty$, where $\|.\|$ is the Euclidean norm.

The Fisher information matrix is defined by $I(s) = \mathbb{E}_s(l'_s {l'_s}^T)$.

In the sequel we restrict our attention to exponential statistical models. We first recall some important concepts and results on exponential families of distribution.

Definition 6.2.2. Let $T : \Omega \to \mathbb{R}^d$. Set

$$S = \{s : K(s) = \ln \int \exp\{\langle s, T\rangle\} d\nu < \infty\} \subset \mathbb{R}^d.$$

We suppose that the set S has non-empty interior S^0.

The general exponential family generated by the measure ν and the map T is the family

$$\{P_s(T, \nu) = \exp\{\langle s\,,\, T\rangle - K(s)\} d\nu = f_s d\nu : \quad s \in S\}. \tag{6.1}$$

Let μ be the image of the measure ν by T. We assume that μ is not concentrated on a strict affine subspace of $\mathbb{R}^d$. The natural exponential family associated with the above general exponential family is the family of probability distributions defined by

$$\{P_s(\mu) = \exp\{\langle s\,,\, .\rangle - K(s)\} d\mu : \quad s \in S\}. \tag{6.2}$$

Natural exponential families may be viewed as a special case of general exponential families with $\Omega \subset \mathbb{R}^d$, $T(\omega) = \omega$ and $\nu = \mu$.

As usual, $\mathbb{E}_s$ will denote the integral on the exponential model $(\Omega, \mathcal{A}, (P_s)_{s\in S})$ where $P_s = P_s(T, \nu)$. We have $l_s = \langle s\,,\, T\rangle - K(s)$. Theorem 2.7.1 in (Lehmann and Romano, 2005) ensures that the cumulant function K and the function $s \mapsto l_s(\omega)$ are analytic on S^0. Moreover the differentiation with respect to s can be carried out under the integral sign in

$$1 = \int \exp\{\langle s\,,\, T\rangle - K(s)\} d\nu$$

as long as $s \in S^0$. This gives, by taking the derivatives and by integration by parts

$$\begin{aligned} \mathbb{E}_s l'_s &= 0 \\ -\mathbb{E}_s l''_s &= I(s). \end{aligned} \tag{6.3}$$

Similarly, we obtain the mean and the covariance

$$m(s) = \mathbb{E}_s(T) = K'(s) \tag{6.4}$$

$$v(s) = \mathrm{Cov}_s(T) = K''(s). \tag{6.5}$$

From (6.3) and (6.5) it follows that the Fisher information of a general exponential family $P_s(T, \nu)$ equals for $s \in S^0$

$$I(s) = K''(s) = v(s). \tag{6.6}$$

The following important result is proved in (Letac and Casalis, 2000).

Proposition 6.2.3. The map $s \mapsto m(s) = \mathbb{E}_s(T) = K'(s)$ is an analytic diffeomorphism from S^0 to the open set $M = m(S^0) \subset \mathbb{R}^d$ called the domain of the means of the family.

Let $\psi : M \to S^0$, $m \mapsto \psi(m) = (K')^{-1}(m)$ denote the inverse of the "mean" diffeomorphism K'. The general exponential family, parametrized by the domain of the means M is given by the family of distributions

$$Q(m, T, \nu)(d\omega) = e^{\langle \psi(m)\,,\, T(\omega)\rangle - K(\psi(m))} \nu(d\omega), \quad m \in M. \tag{6.7}$$

The mean of the family (6.7) is equal to m. Let $V(m)$ denote the covariance of the family (6.7). Then, by (6.5) we have

$$V(m) = v(\psi(m)) = K''(\psi(m)). \tag{6.8}$$

The function $V : m \in M \to V(m)$ is called the *variance function* of the exponential family.

In order to avoid confusion, when the parameter of an exponential family is the mean m, the Fisher information will be noted $J(m)$.

Theorem 6.2.4. The Fisher information of the exponential family (6.7) equals

$$J(m) = V(m)^{-1} = \psi'(m), \tag{6.9}$$

where $V(m)$ is the variance function of the exponential family, given by (6.8).

Proof. By Definition 6.2.1 and by the chain rule,
$J(m) = \psi'(m)^T I(\psi(m))\psi'(m)$ on M. Since $\psi(m) = (K')^{-1}(m)$, we have
$\psi'(m) = [K''(\psi(m))]^{-1}$. Thus, using formula (6.6), we get
$J(m) = [K''(\psi(m))]^{-1} = V(m)^{-1}$. □

Remark 6.2.5. Note a striking contrast in the formulas (6.6) and (6.9) for the Fisher information of an exponential family parametrized either by the canonical parameter $s \in S^0$ or by the mean $m \in M$; in the first case we have
$I(\psi(m)) = V(m)$, in the second $J(m) = V(m)^{-1}$.

Finally, consider the general exponential family $Q(m, T, \nu)$ parametrized by the means domain. Let $A \neq 0, B \in \mathbb{R}^d$. Define $\Theta = \{\theta \in \mathbb{R} : \theta A + B \in M\}$. The set $\Theta \subset \mathbb{R}$ is open because M is open. Suppose that $\Theta \neq \varnothing$. The parametrization by a segment of means $I \subset \Theta$ consists in considering the submodel

$$\{Q(\theta A + B, T, \nu) : \theta \in I\}. \tag{6.10}$$

Such models contain the case $\{Q(\theta m_1 + (1-\theta)m_2, T, \nu) : \theta \in [0,1]\}$ which is suitable when one hesitates between two different estimations $m_1, m_2 \in M$ of the true mean m of an exponential family (6.7).

The following corollary gives the Fisher information of a general exponential family parametrized by a segment of means. By analogy to the notation $J(m)$, this information is noted $J(\theta)$.

Corollary 6.2.6. The Fisher information of the model $\{Q(\theta A + B, T, \nu) : \theta \in I\}$ equals

$$J(\theta) = A^T V(\theta A + B)^{-1} A. \tag{6.11}$$

Proof. We use Definition 6.2.1 and the chain rule similarly as in the proof of Theorem 6.2.4, for the reparametrization $f : I \to M$, $f(\theta) = \theta A + B$, with $f'(\theta) = A$. We conclude by Theorem 6.2.4. □

6.3 Fisher information of Gaussian and Wishart families parametrized by a segment of means

In this section, we study the Fisher information for multivariate Gaussian and Wishart exponential families. These families are parametrized by symmetric positive definite matrices. Therefore we first adapt the presentation to suit this case. We denote by $\mathbb{R}^{k\times m}$ the space of real matrices with k rows and m columns and by $A \otimes B$ the Kronecker product of two matrices. We use the usual notation $\langle A, B\rangle = \mathrm{Tr}({}^{\mathtt{t}}AB)$ for the scalar product of two matrices. The operator Vec converts a $k \times m$ matrix A into a vector $\mathrm{Vec}(A) \in \mathbb{R}^{km}$ by stacking the columns one underneath the other. The Vec operator is commonly used in applications of the matrix differential calculus in statistics, cf. (Magnus and Neudecker, 2007; Muirhead, 2005).

The following properties of the Kronecker product are used in this work (Magnus and Neudecker, 2007, p.32,35). For non-singular squared matrices A, B we have

$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$. For all matrices A, B and C such that the product ABC is well defined

$$\mathrm{Vec}(A\,B\,C) = (C^T \otimes A)\,\mathrm{Vec}(B). \tag{6.12}$$

We use the following convention of matrix differential calculus: if a function $f : \mathbb{R}^{k\times p} \to \mathbb{R}^{n\times m}$ is differentiable then its derivative is a matrix $f'(x) \in \mathbb{R}^{nm\times kp}$ such that

$$\mathrm{Vec}(df(x)(u)) = f'(x)\,\mathrm{Vec}(u), \quad u \in \mathbb{R}^{k\times p}. \tag{6.13}$$

The only exception we will make is the derivative of a function $K : \mathbb{R}^{k\times m} \to \mathbb{R}$, for which the following convention is used: the derivative of K is not a row vector but the matrix $K'(x) \in \mathbb{R}^{k\times m}$, related to the differential of K by
$dK(x)(u) = \langle K'(x), u\rangle = \mathrm{Tr}(K'(x)^T u)$, for all $u \in \mathbb{R}^{k\times m}$. This convention is needed to give sense to formula (6.4) for the mean of an exponential family.

The following Lemma is useful for the derivation of an alternative formula for the Fisher information of an exponential family parametrized by a segment of means and verifying an additional condition (6.14). We will see that this condition holds for Gaussian and Wishart models.

Lemma 6.3.1. Assume that for all $m \in \mathcal{M}$,

$$\langle m\,,\,\psi(m)\rangle = c, \tag{6.14}$$

for some constant $c \in \mathbb{R}$. Then, for all $u \in \mathcal{M}$,

$$\langle m\,,\,d\psi(m)(u)\rangle = -\langle u\,,\,\psi(m)\rangle. \tag{6.15}$$

Proof. By (6.14) the differential of the function $g : \mathcal{M} \to \mathbb{R}$, $m \mapsto \langle m\,,\,\psi(m)\rangle$ is zero. Therefore, $dg(m)(u) = \langle m\,,\,d\psi(m)(u)\rangle + \langle u\,,\,\psi(m)\rangle = 0$ for all $m, u \in \mathcal{M}$ and (6.15) follows. □

Corollary 6.3.2. Let $\{Q(\theta A+B, T, \nu)(d\omega) : \theta \in I\}$ be an exponential model parametrized by a segment of means. If the condition (6.14) holds then the Fisher information of the model equals

$$J(\theta) = -\frac{d^2}{d\theta^2}\left[K(\psi(\theta A + B))\right]. \tag{6.16}$$

Proof. Let $h(\theta) = K(\psi(\theta A + B))$ and $f(\theta) = \theta A + B$. We want to compute $h''(\theta)$. If $\theta, u \in \mathbb{R}$,

$$\begin{aligned} dh(\theta)(u) &= dK(\psi(f(\theta)))\big(d\psi(f(\theta))(df(\theta)(u))\big) \\ &= \langle K'(\psi(f(\theta)))\,,\, d\psi(f(\theta))(df(\theta)(u))\rangle \\ &= \langle f(\theta)\,,\, d\psi(f(\theta))(df(\theta)(u))\rangle \\ &= -\langle df(\theta)(u)\,,\, \psi(f(\theta))\rangle \\ &= -u\langle A\,,\, \psi(f(\theta))\rangle, \end{aligned}$$

where we used successively: the convention on K' introduced after (6.13), the equality $K' \circ \psi(m) = m$, Lemma 6.3.1 and the formula $df(\theta)(u) = uA$. Thus we have $h'(\theta) = -\langle A\,,\, \psi(f(\theta))\rangle$. Now, starting as in the computation of $h'(\theta)$ and using (6.13), we get

$$\begin{aligned} h''(\theta) &= -\langle A\,,\, d\psi(f(\theta))(A)\rangle = -\operatorname{Vec}(A)^T \operatorname{Vec}(d\psi(f(\theta))(A)) \\ &= -\operatorname{Vec}(A)^T \psi'(\theta A + B) \operatorname{Vec}(A). \end{aligned}$$

We conclude using (6.9) and Corollary 6.2.6. □

6.3.1 Exponential families of Gaussian distributions

We denote by S_d the vector space of $d \times d$ symmetric matrices and by S_d^+ the open cone of positive definite matrices.

Let us recall the construction of the multivariate Gaussian model $\{N(u, \Sigma);\ \Sigma \in S_d^+\}$ as a general exponential family. Here u is a fixed vector of $\mathbb{R}^d$. We consider $\Omega = \mathbb{R}^d$

equiped with a rescaled Lebesgue measure $\nu(d\omega) = d\omega/(2\pi)^{d/2}$, the vector space S_d and the map

$$T : \mathbb{R}^d \to S_d, \;\; T(\omega) = -\frac{1}{2}(\omega - u)(\omega - u)^T.$$

The image of T is contained in the opposite of the cone of semi-positive definite matrices of rank one. For $s \in \mathcal{S}_d^+$, we have

$$\int_\Omega e^{\langle s, T(\omega)\rangle} \nu(d\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\frac{1}{2}\operatorname{Tr}\left(s(\omega-u)(\omega-u)^T\right)} d\omega = (\det s)^{-1/2}$$

and the integral is infinite otherwise. Thus $S = \mathcal{S}_d^+$ and the cumulant function is

$$K(s) = -\frac{1}{2}\ln\det(s), \quad s \in S = S_d^+.$$

The general exponential family is therefore

$$\begin{aligned} P_s(T,\nu)(d\omega) &= \frac{1}{(2\pi)^{d/2}} e^{\langle s\,, -\frac{1}{2}(\omega-u)(\omega-u)^T\rangle + \frac{1}{2}\ln\det(s)} d\omega \\ &= \frac{(\det s)^{1/2}}{(2\pi)^{d/2}} e^{-\frac{1}{2}(\omega-u)^T s(\omega-u)} d\omega, \end{aligned} \tag{6.17}$$

which is the family of Gaussian distributions $N(u, s^{-1})$ on $\mathbb{R}^d$ with a fixed mean $u \in \mathbb{R}^d$, parametrized by $s = \Sigma^{-1}$. The derivative of the function $X \in \mathbb{R}^{d\times d} \to \det X$ is the cofactor matrix $X^\sharp$ which equals $(\det X)(X^{-1})^T$ when X is invertible. It follows that

$$m(s) = K'(s) = -\frac{1}{2}s^{-1}, \quad s \in S_d^+.$$

The means domain is $\mathcal{M} = -S_d^+$ and the inverse mean map is $\psi(m) = -\frac{1}{2}m^{-1}$. The Gaussian general exponential family parametrized by $m \in \mathcal{M} = -S_d^+$ is therefore the family

$$Q(m, T, \nu) = N(u, -2m). \tag{6.18}$$

Up to a change of parameter $\Sigma = -2m$, this parametrization by the covariance parameter is more natural than the parametrization of the family $(N(u, s^{-1}))_{s\in S_d^+}$ by the canonical parameter s.

In order to compute the variance function, recall that $XX^{-1} = I_d$ implies that $dX^{-1} = -X^{-1}(dX)\,X^{-1}$ and $(X^{-1})' = -X^{-1} \otimes X^{-1}$.

Thus $K''(s) = \frac{1}{2}s^{-1} \otimes s^{-1}$ and formula (6.8) implies that

$$V(m) = 2m \otimes m. \tag{6.19}$$

The Fisher information of the family $(N(u, s^{-1}))_{s\in S_d^+}$ is $I(s) = \frac{1}{2}s^{-1} \otimes s^{-1}$. By Theorem 6.2.4 and formula (6.19), the Fisher information of the model $(N(u, -2m))_{m\in -S_d^+}$ equals $J(m) = \frac{1}{2}m^{-1} \otimes m^{-1}$.

Corollary 6.3.3. The Fisher information matrix of the Gaussian model $(N(u, \Sigma))_{\Sigma\in S_d^+}$ is

$$J(\Sigma) = \frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1}.$$

Proof. Using chain rule and a reparametrization $\Sigma = -2m$ we see that the information for the new parameter Σ is $\tilde{J}(\Sigma) = \frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1} = J(\Sigma)$. □

Let us now consider Gaussian models parametrized by a segment of covariances.

Corollary 6.3.4. Let C and D be two symmetric matrices and let $I \subset \mathbb{R}$ be a non-empty segment such that $I \subset \Theta = \{\theta \in \mathbb{R} : \theta C + D \in S_d^+\}$. The Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ is

$$J(\theta) \;=\; \frac{1}{2}\operatorname{Tr}\left(C(\theta C + D)^{-1}C(\theta C + D)^{-1}\right).$$

Proof. We use Corollary 6.3.3 and the chain rule with $f(\theta) = \theta C + D$. It follows that

$$J(\theta) \;=\; \operatorname{Vec}(C)^T J(\theta C + D)\operatorname{Vec}(C)$$

$$= \frac{1}{2}\operatorname{Vec}(C)^T \left((\theta C + D)^{-1} \otimes (\theta C + D)^{-1}\right) \operatorname{Vec}(C).$$

Applying (6.12) we get

$$\begin{aligned} J(\theta) &= \frac{1}{2}\operatorname{Vec}(C)^T \operatorname{Vec}\left((\theta C + D)^{-1} C (\theta C + D)^{-1}\right) \\ &= \frac{1}{2}\operatorname{Tr}\left(C(\theta C + D)^{-1} C (\theta C + D)^{-1}\right). \end{aligned}$$

□

On the other hand, we have the following alternative formula for the information $J(\theta)$.

Corollary 6.3.5. The Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ is

$$J(\theta) = -\frac{1}{2}\frac{d^2}{d\theta^2}(\ln \det(\theta C + D)). \tag{6.20}$$

Proof. Observe that the condition (6.14) holds for the Gaussian exponential families $Q(m, t, \nu)$:

$$\langle m\,,\, \psi(m)\rangle = -\frac{1}{2}\operatorname{Tr}(mm^{-1}) = -\frac{d}{2}.$$

The model $N(u, \theta C + D) = N(u, -2m) = Q(m, T, \nu)$, with $m = \theta A + B \in \mathcal{M} = -S_d^+$ where $A = -\frac{C}{2}$ and $B = -\frac{D}{2}$. We apply Corollary 6.3.2 and the fact that

$$K(\psi(\theta A + B)) = -\frac{1}{2}\ln \det(\theta C + D).$$

Formula (6.20) follows. □

Now we characterize the information $J(\theta)$ in terms of the eigenvalues of the matrix $D^{-1/2} C D^{-1/2}$.

Theorem 6.3.6. Let C and D be two symmetric matrices and let $I \subset \mathbb{R}$ be a segment such that $IC + D \subset S_d^+$. Let $a_1, \ldots, a_d$ be the eigenvalues of the matrix $D^{-1/2}CD^{-1/2}$.

The Fisher information of the Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ equals

$$J(\theta) = \frac{1}{2} \sum_{j=1}^{d} \left(\frac{a_j}{1 + a_j \theta} \right)^2. \tag{6.21}$$

Proof. The idea of the proof is to use formula (6.20). Let $P(\lambda)$ be the characteristic polynomial of the matrix $D^{-1/2}CD^{-1/2}$. We have

$$\begin{aligned} P(\lambda) &= \det(D^{-1/2}CD^{-1/2} - \lambda I_d) \\ &= \det(D^{-1}C - \lambda I_d) = (\det D)^{-1} \det(C - \lambda D). \end{aligned}$$

On the other hand, $P(\lambda) = \prod_{j=1}^{d}(a_j - \lambda)$. It follows that

$$|\theta C + D| = |\theta D(D^{-1/2}CD^{-1/2} - \frac{1}{\theta} I_d)| = |D| \left(\theta^d P(-1/\theta) \right) = |D| \prod_{j=1}^{d} (\theta a_j + 1).$$

The last formula allows to compute easily $\frac{d^2}{d\theta^2}(\ln \det(\theta C + D))$. First we see that

$$\frac{d}{d\theta}(\ln \det(\theta C + D)) = \frac{\frac{d}{d\theta} \det(\theta C + D)}{\det(\theta C + D)} = \sum_{j=1}^{d} \frac{a_j}{\theta a_j + 1}.$$

One more derivation and formula (6.20) lead to (6.21). □

We finish by computing the Fisher information of two Gaussian models in $\mathbb{R}^d$, parametrized by an explicitly given segment of covariances. First, let A be a circulant matrix with the first row $e_2 + e_d = (0, 1, 0, \ldots, 0, 1)$. Then for a segment $I \subset \mathbb{R}$ containing 0 and $\theta \in I$

$$\theta A + I_d = \begin{pmatrix} 1 & \theta & 0 & \ldots & 0 & \theta \\ \theta & 1 & \theta & 0 & \ldots & 0 \\ 0 & \theta & 1 & \theta & 0 & \ldots \\ & & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & \theta & 1 & \theta \\ \theta & 0 & \ldots & 0 & \theta & 1 \end{pmatrix} \in S_d^+. \tag{6.22}$$

Corollary 6.3.7. The Fisher information of the model $(N(0, \theta A + I_d))_{\theta \in I}$ is given by

$$J(\theta) \quad = \quad \frac{1}{2} \sum_{j=0}^{d-1} \left(\frac{2\cos(\frac{2\pi j}{d})}{1 + 2\theta \cos(\frac{2\pi j}{d})} \right)^2 . \tag{6.23}$$

Proof. Let $\mathcal{A}$ be a circulant matrix with the first row $(r_0, r_1, \ldots, r_{d-1})$. It is well known (see e.g. (Gray, 2006)) and easy to check that if ϵ is a d-th root of unity, $\epsilon^d = 1$, then $a = \sum_{l=0}^{d-1} r_l \epsilon^l$ is an eigenvalue of $\mathcal{A}$ with an eigenvector $(1, \epsilon, \epsilon^2, \ldots, \epsilon^{d-1})$.

Therefore if $\epsilon_j = e^{\frac{2\pi j i}{d}}$, $j = 0, \ldots, d-1$ are the d distinct d-th roots of unity, then the matrix $\mathcal{A}$ has d distinct eigenvalues $a_j = \sum_{l=0}^{d-1} r_l \epsilon_j^l$. In our particular case,

$$a_j = e^{\frac{2\pi j i}{d}} + e^{\frac{2(d-1)\pi j i}{d}} = 2\cos\left(\frac{2\pi j}{d}\right).$$

Formula (6.23) follows from Theorem 6.3.6. □

Now, let us consider a tridiagonal matrix C such that

$$\theta C + I_d = \begin{pmatrix} 1 & \theta & 0 & 0 & 0 & \ldots \\ \theta & 1 & \theta & 0 & 0 & \ldots \\ 0 & \theta & 1 & \theta & 0 & \ldots \\ & \ddots & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & \theta & 1 & \theta \\ 0 & \ldots & 0 & 0 & \theta & 1 \end{pmatrix}. \tag{6.24}$$

As in the preceding case, there exists a segment $I \subset \mathbb{R}$ such that $\theta C + I_d \in S_d^+$ for $\theta \in I$.

Corollary 6.3.8. The Fisher information of the model $(N(0, \theta C + I_d))_{\theta \in I}$ is given by

$$J(\theta) = \frac{1}{2} \sum_{j=1}^{d} \left(\frac{2\cos\left(\frac{j}{d+1}\pi\right)}{1 + 2\theta \cos\left(\frac{j}{d+1}\pi\right)} \right)^2 . \tag{6.25}$$

Proof. We will apply Theorem 6.3.6 with C and $D = I_d$.

Expanding $\psi_d(\lambda) = \det(C - \lambda I_d)$ along the first row, we get

$\psi_d(\lambda) = -\lambda\psi_{d-1}(\lambda) - M^{1,2}$. Expanding the minor $M^{1,2}$ along its first column gives $M^{1,2} = \psi_{d-2}(\lambda)$ and

$$\psi_d(\lambda) = -\lambda\psi_{d-1}(\lambda) - \psi_{d-2}(\lambda),\ d \geqslant 3.$$

We set $\varphi_d(\lambda) = (-1)^d\psi_d(2\lambda)$ and we obtain

$$\varphi_d(\lambda) = 2\lambda\varphi_{d-1}(\lambda) - \varphi_{d-2}(\lambda),\ d \geqslant 3$$

with initial conditions $\varphi_1(\lambda) = 2\lambda$, $\varphi_2(\lambda) = 4\lambda^2 - 1$. Therefore φ_d is a Tchebyshev polynomial of the second kind (Mason and Handscomb, 2003) and it satisfies

$$\varphi_d(\cos x) = \frac{\sin(d+1)x}{\sin x},\ d \geqslant 1.$$

We have, for all $\lambda \in [-2, 2]$,

$$\psi_d(\lambda) = 0 \iff \varphi_d\left(\frac{\lambda}{2}\right) = 0 \implies \frac{\sin(d+1)x}{\sin x} = 0, \quad x = \arccos\frac{\lambda}{2}.$$

Therefore $\lambda_j = 2\cos\left(\frac{j}{d+1}\pi\right)$, $1 \leqslant j \leqslant d$, are d distinct eigenvalues of the matrix C. □

6.3.2 Exponential families of Wishart distributions

Wishart distributions on the cone $\overline{S_d^+}$ are defined as elements of natural exponential families generated by Riesz measures, (Faraut and Korányi, 1994). Recall that the Riesz measures μ_p on the cone $\overline{S_d^+}$ are unbounded positive measures such that their Laplace transform equals for $t \in S_d^+$

$$\mathcal{L}\,\mu_p(t) = \int_{\overline{S_d^+}} e^{-\langle t,x\rangle} d\mu_p(x) = (\det t)^{-p}.$$

By the celebrated Gindikin theorem, such measures exist if and only if p belongs to the Gindikin set $\Lambda_d = \{\frac{1}{2}, \ldots, \frac{d-1}{2}\} \cup \left(\frac{d-1}{2}, \infty\right)$. Their support is equal to the cone $\overline{S_d^+}$ if and

only if $p > \frac{d-1}{2}$ and they are absolutely continuous in that case. Otherwise, when $p \in \{\frac{1}{2}, \ldots, \frac{d-1}{2}\}$, the measures μ_p are singular and concentrated on semipositive symmetric matrices of rank $2p$.

The family of Wishart distributions $W(p;s)$ on $\overline{S_d^+}$ is defined as the natural exponential family generated by the Riesz measure μ_p. It means that $p \in \Lambda_d$, $s \in S = -S_d^+$ and

$$\begin{aligned} W(p;s)(dx) &= \frac{e^{\langle s\,,\,x\rangle}}{\mathcal{L}\,\mu_p(-s)}\mu_p(dx) \\ &= e^{\langle s\,,\,x\rangle}(\det(-s))^p\mu_p(dx) = e^{\langle s\,,\,x\rangle - K_p(s)}\mu_p(dx), \end{aligned}$$

with $K_p(s) = -p\ln\det(-s)$. It follows that $\mathcal{L}\,W(p;s)(t) = \det(I_d + (-s)^{-1}t)^{-p}$ and that $\mu_p(dx) = e^{\operatorname{Tr} x}W(p;-I_d)$.

Wishart distributions are multivariate analogs of the gamma distributions with density $\lambda^p\Gamma(p)^{-1}e^{-\lambda x}x^{p-1}dx$ on $\mathbb{R}^+(\,p > 0, \lambda > 0)$, considered with a canonical parameter $s = -\lambda < 0$. Similarly as in dimension 1, the Wishart distributions are often parametrized by a scale parameter $\sigma = (-s)^{-1} \in S_d^+$ and then the notation $\gamma(p;\sigma) = W(p;(-\sigma)^{-1})$ is used, cf. (Letac and Massam, 2008). The study of Wishart distributions is motivated by their importance as estimators of the covariance matrix of a Gaussian model in $\mathbb{R}^d$.

Let us apply our results on the Fisher information to a natural exponential family of Wishart distributions $\{W(p;s) :\ s \in -S_d^+\}$. The mean equals
$m(s) = K_p'(s) = p(-s)^{-1} \in \mathcal{M} = S_d^+$ and the inverse mean map $\psi : S_d^+ \to -S_d^+$ is $\psi(m) = -pm^{-1}$.

Thus the Wishart family $Q(m,\mu_p)$ parametrized by the domain of means is, up to a trivial reparametrization $m \to \frac{1}{p}m$, the family parametrized by its scale parameter:

$$Q(m,\mu_p) = W(p;-pm^{-1}) = \gamma(p;\frac{1}{p}m), \quad m \in S_d^+. \tag{6.26}$$

As $v(s) = K_p''(s) = p(s^{-1} \otimes s^{-1})$, it follows that the variance function is

$$V(m) = \frac{1}{p}(m \otimes m). \tag{6.27}$$

The Fisher information of the model $\{W(p;s) : \ s \in -S_d^+\}$ is $I(s) = ps^{-1} \otimes s^{-1}$. By Theorem 6.2.4 the Fisher information of the model $\{Q(m, \mu_p), \ m \in \mathcal{M}\}$ is $J(m) = pm^{-1} \otimes m^{-1}$.

Consequently, using the reparametrization $m \to \frac{1}{p}m = \sigma$ and the chain rule, we see that the Fisher information matrix of the Wishart model $\{\gamma(p;\sigma) : \ \sigma \in S_d^+\}$ parametrized by the scale parameter σ equals
$J(\sigma) = p\sigma^{-1} \otimes \sigma^{-1}$.

Theorem 6.3.9. Let $I = (a,b) \subset \mathbb{R}$ and $C, D \in S_d$ be such that $IC + D \subset S_d^+$. The Fisher information $J(\theta)$ of the Wishart model
$\{\gamma(p;\theta C + D) : \ \theta \in I\}$ verifies the formulas

$$J(\theta) = p \operatorname{Tr}\left(C(\theta C + D)^{-1}\right)^2 \tag{6.28}$$

$$J(\theta) = -p\frac{d^2}{d\theta^2}(\ln \det(\theta C + D))$$

$$J(\theta) = p\sum_{j=1}^{d}\left(\frac{a_j}{1 + a_j\theta}\right)^2 \tag{6.29}$$

where $a_1, \ldots, a_d$ are the eigenvalues of the matrix $D^{-1/2}CD^{-1/2}$.

Proof. The proofs are similar to the proofs of the analogous results for exponential Gaussian families in the previous subsection. The condition (6.14) holds true: $\langle m \, , \, \psi(m) \rangle = -pd$, the model $\{\gamma(p;\theta C + D) : \ \theta \in I\}$ is equal to the model $\{Q(\theta pC + pD, \mu_p) : \ \theta \in I\}$ parametrized by the means and we have $K_p(\psi(\theta pC + pD)) = p \ln \det(\theta C + D)$. □

Corollary 6.3.10. Let $\sigma_1, \sigma_2 \in \mathcal{S}_d^+$ and let I be the open interval containing θ such that

$\sigma_\theta = \theta\sigma_1 + (1-\theta)\sigma_2 \in \mathcal{S}_d^+$. The Fisher information of the model $\{\gamma(p;\sigma_\theta) : \theta \in I\}$ is equal to $J(\theta) = p\,\mathrm{Tr}\left(\left((\sigma_1-\sigma_2)\,\sigma_\theta^{-1}\right)^2\right)$.

Proof. We write $\theta\sigma_1 + (1-\theta)\sigma_2 = \theta(\sigma_1-\sigma_2)+\sigma_2$ and we apply formula (6.28). □

Using (6.29) we obtain the following corollary, analogous to Corollaries 6.3.7 and 6.3.8.

Corollary 6.3.11. 1. Consider the model $\{\gamma(p;\theta A + I_d) : \theta \in I\}$ with $\theta A + I_d$ as in (6.22). Then its Fisher information equals

$$J(\theta) = p\sum_{j=0}^{d-1}\left(\frac{2\cos(\frac{2\pi j}{d})}{1+2\theta\cos(\frac{2\pi j}{d})}\right)^2.$$

2. Consider the model $\{\gamma(p;\theta C + I_d) : \theta \in I\}$ with $\theta C + I_d$ as in (6.24). Then its Fisher information equals $J(\theta) = p\sum_{j=1}^{d}\left(\frac{2\cos\left(\frac{j}{d+1}\pi\right)}{1+2\theta\cos\left(\frac{j}{d+1}\pi\right)}\right)^2$.

Remark 6.3.12. Let $P_s(\mu)$ be the natural exponential family corresponding to the Gaussian general exponential family (6.17). If W has the law $N(u,s^{-1})$ given by (6.17), then $T(W)$ has the law $P_s(\mu)$. On the other hand, it is well known that $-T(W) = \frac{1}{2}(W-u)(W-u)^T$ has the Wishart law $\gamma(\frac{1}{2};2s^{-1})$. This explains why the formulas for the Fisher information are the same for the Gaussian family and for the Wishart family with $p=\frac{1}{2}$.

Exponential families of noncentral Wishart distributions

Let us finish the section on the Wishart models by considering the non-central case. The main reference is (Letac and Massam, 2008). Let $p \in \Lambda_d = \{\frac{1}{2},\ldots,\frac{d-2}{2}\}$, $a \in \overline{S_d^+}$ and $\sigma \in S_d^+$. The noncentral Wishart distribution $\gamma(p,a;\sigma)$ is defined by its Laplace transform

$$\mathcal{L}_\gamma(p,a;\sigma)(t) = \int_{\overline{S_d^+}} e^{-\mathrm{Tr}(tx)}\gamma(p,a;\sigma)(dx) = \det(I_d+\sigma t)^{-p}e^{-\mathrm{Tr}\left(t(I_d+\sigma\,t)^{-1}\sigma a\sigma\right)},$$

for all $t \in S_d^+$.

When $p \geqslant \frac{d-1}{2}$, then non-central Wishart laws exist for all $a \in \overline{S_d^+}$; when $p \in \Lambda_d$ then a must be of rank at most $2p$ (Letac and Massam, 2011). When $p = \frac{n}{2}$, $n \in \mathbb{N}$, the non-central Wishart distributions are constructed in the following way from n independent d-dimensional Gaussian vectors $Y_1, \ldots, Y_n$. Let $Y_j \sim N(u_j\,, \Sigma)$ and let u be the $d \times n$ matrix $[u_1, \ldots, u_n]$. Then, the $d \times d$ matrix $W = Y_1 Y_1^T + \ldots + Y_n Y_n^T$ has the noncentral Wishart distribution $\gamma(p, a; \sigma)$ with $p = \frac{n}{2}$, $\sigma = 2\Sigma$ and $\sigma a \sigma = uu^T$. Such Wishart distributions are studied in (Muirhead, 2005).

The non-central Wishart distributions may be constructed as a natural exponential family $\{W(p, a; s) : \ s \in -S_d^+\}$ generated by the positive measure
$\mu = \mu_{a,p}(dx) = e^{\mathrm{Tr}(a+x)}\gamma(p, a; I_d)(dx)$. Its moment generating function is given for $s \in -S_d^+$ by

$$\int_{\overline{S_d^+}} e^{\mathrm{Tr}(sx)}\mu_{a,p}(dx) = \det(-s)^{-p} e^{\mathrm{Tr}\left(a(-s)^{-1}\right)}.$$

We have $W(p, a; s) = \gamma(p, a; (-s)^{-1})$. Like for central Wishart families,
$S = -S_d^+$. The cumulant function is

$$K(s) = -p \log\det(-s) + \mathrm{Tr}(a(-s)^{-1}).$$

As before, we denote $\sigma = (-s)^{-1}$. We see that the mean equals

$$m(s) = K'(s) = p(-s)^{-1} + (-s)^{-1}a(-s)^{-1} = p\sigma + \sigma a \sigma \tag{6.30}$$

and the covariance

$$\begin{aligned} v(s) &= K''(s) = p\sigma \otimes \sigma + (\sigma a \sigma) \otimes \sigma + \sigma \otimes (\sigma a \sigma) \\ &= -p\sigma \otimes \sigma + m \otimes \sigma + \sigma \otimes m. \end{aligned} \tag{6.31}$$

When the matrix a is non-singular, the inverse mean map $\psi(m) = s$ is such that

$$(-s)^{-1} = \sigma = -\frac{p}{2}a^{-1} + a^{-1/2}\left(a^{1/2} m a^{1/2} + \frac{p^2}{4} I_d\right)^{1/2} a^{-1/2}. \tag{6.32}$$

For other cases see (Letac and Massam, 2008, Prop.4.5). In order to write the variance function $V(m) = v(\psi(m))$ we compose the last expression from (6.31) and the formula (6.32).

For a model $\{W(p, a; \psi(\theta A + B)) : \theta \in I\}$ parametrized by a segment of means, the Fisher information $J(\theta)$ is obtained from the expression of $V(m)$ and Theorem 6.2.6.

Example 6.3.13. *Suppose that $a = I_d$, $A = \alpha I_d$ and $B = \beta I_d$, $\alpha, \beta > 0$. The Fisher information on θ is*

$$J(\theta) = \alpha^2 d \left((p^2 + 2\theta\alpha + 2\beta)(\theta\alpha + \beta + \frac{p^2}{4})^{1/2} - 2p(\theta\alpha + \beta) - \frac{p^3}{2} \right)^{-1}.$$

6.3.3 Applications to estimation of the mean in exponential families parametrized by a segment of means

Consider a sample $X_1, \ldots, X_n$ of a random variable X from a natural exponential family $Q(m, \mu)$ parametrized by the domain of means $\mathcal{M}$, where the parameter $m = \mathbb{E}(X)$ is unknown and $\mathcal{M}$ is open.

Proposition 6.3.14. The sample mean $\bar{X}_n$ is an unbiased, consistent and efficient estimator of the parameter m. It is a maximum likelihood estimator of m.

Proof. By Theorem 6.2.4 we have $\mathrm{Cov}(X) = V(m) = J(m)^{-1}$, so the Cramér-Rao bound is attained by X. Consequently, the sample mean $\bar{X}_n$ is an efficient estimator of m. It follows by equating zero to the derivative with respect to m of the logarithm of expression (6.7) that the sample mean $\bar{X}_n$ is a maximum likelihood estimator of m. One can also first show by (6.2) that the maximum likelihood estimator of s is $\hat{s} = (K')^{-1}(X) = \psi(X)$ and next use the functional invariance of the maximum likelihood estimator (Casella and Berger, 2002, Theorem 7.2.10). □

Remark 6.3.15. For general exponential families $Q(m, T, \nu)$ parametrized by an open domain of means M, all these properties remain valid for $\hat{m} = \overline{T(X)}_n$ as an estimator of $m = \mathbb{E}\left(T(X)\right)$.

Consider an exponential family $Q(\theta A + B, \mu)$ parametrized by $\theta \in I$, a segment in $\mathbb{R}$ with fixed $A \neq 0$ and $B \in \mathcal{M}$. We will now discuss estimators of the real parameter θ when we know that the mean $\mathbb{E}(X) = m \in IA + B$. Determining a maximum likelihood estimator for θ seems impossible explicitly. This is the "price to pay" for the parsimony of the segment model parametrized by $m \in IA + B$. On the other hand, the efficiency of estimators of θ may be studied thanks to Corollary 6.2.6.

Knowing that

$$m = \theta A + B \tag{6.33}$$

for a value $\theta \in I$, we have many possibilities of writing down a solution θ of equation (6.33). If $A \neq 0$ then the solution θ is unique ($A\theta + B = A\theta' + B$ implies $\theta = \theta'$ when $A \neq 0$). We assume that m and therefore A and B are $d \times d$ matrices. For any C such that $\langle A\,,\, C\rangle \neq 0$ we have

$$\theta \;=\; \frac{\langle m - B\,,\, C\rangle}{\langle A\,,\, C\rangle}.$$

We define an estimator $\hat{\theta}_C$ of the parameter θ by

$$\hat{\theta}_C \;=\; \frac{\langle \bar{X}_n - B\,,\, C\rangle}{\langle A\,,\, C\rangle}$$

All the estimators $\hat{\theta}_C$ are unbiased and consistent. The natural question is whether they are efficient. The variance of $\hat{\theta}_C$ may be computed using the variance function $V(m)$ of the exponential family:

$$\operatorname{Var}\hat{\theta}_C = \frac{1}{\langle A, C\rangle^2}\operatorname{Var}\langle \bar{X}_n, C\rangle = \frac{1}{\langle A, C\rangle^2}\operatorname{Var}\left(\operatorname{Vec}(C)^T\ \operatorname{Vec}\bar{X}_n\right)$$

$$= \frac{\operatorname{Vec}(C)^T V(\theta A + B) \operatorname{Vec}(C)}{n\langle A, C\rangle^2}. \tag{6.34}$$

On the other hand, the Cramér-Rao bound is equal by Corollary 6.2.6 to

$$\frac{1}{nJ(\theta)} = \frac{1}{n \operatorname{Vec}(A)^T V(\theta A + B)^{-1} \operatorname{Vec}(A)}. \tag{6.35}$$

When the matrix A is invertible, we can take $C = A^{-1}$ and consider the estimator

$$\hat{\theta}_{A^{-1}} = \frac{\langle \bar{X}_n - B, A^{-1}\rangle}{d}.$$

The following theorem shows that for Gaussian and central Wishart exponential families and for linearly dependent A and B, the estimator $\hat{\theta}_{A^{-1}}$ is efficient as an estimator of the mean m (with X_i replaced by $T(X_i) = -\frac{1}{2}(X_i - u)(X_i - u)^T$ in the Gaussian case). In conclusion, we obtain efficient estimators for Gaussian models parametrized by a covariance segment parameter and for Wishart models parametrized by a scale segment parameter.

Theorem 6.3.16. 1. Let $I \subset \mathbb{R}^+$ be a non-empty segment. Let $c \geqslant 0$, $A \in S_d^+$ and $B = cA$.

(a) Consider an n-sample $(X_1, \ldots, X_n)$ from a Gaussian family $Q(m, T, \nu)$ defined by (6.18), where $m = \theta A + B$, $\theta \in I$. Then

$$\hat{\theta}_{A^{-1}} = \frac{\langle \overline{T(X)}_n - B, A^{-1}\rangle}{d}$$

is a uniformly minimum-variance unbiased estimator of the parameter θ.

(b) Consider an n-sample $(X_1, \ldots, X_n)$ from a Wishart model $Q(m, \mu_p)$ defined by (6.26), where $m = \theta A + B$, $\theta \in I$. Then

$$\hat{\theta}_{A^{-1}} = \frac{\langle \bar{X}_n - B, A^{-1}\rangle}{d}$$

is a uniformly minimum-variance unbiased estimator of the parameter θ.

2. Let $c \geqslant 0$, $C \in S_d^+$ and $D = cC$.

   (a) Consider an n-sample $(X_1, \ldots, X_n)$ from a Gaussian model $\{N(u, \theta C + D), \theta \in I\}$ parametrized by a segment of covariances. A uniformly minimum-variance unbiased estimator of θ is given by

$$\hat{\theta} = \frac{1}{d}\langle \frac{1}{n}\sum_{i=1}^{n}(X_i - u)(X_i - u)^T - D, C^{-1}\rangle.$$

   (b) Consider a sample $(X_1, \ldots, X_n)$ from a Wishart model $\{\gamma(p, \theta C + D), \theta \in I\}$ parametrized by a segment of scale parameters. A uniformly minimum-variance unbiased estimator of θ is given by

$$\hat{\theta} = \frac{\langle \frac{1}{p}\bar{X}_n - D, C^{-1}\rangle}{d}.$$

Proof. For the first part of the Theorem, we give the proof in the Wishart case. The proof in the Gaussian case is identical, with $p = \frac{1}{2}$, cf. Remark 6.3.12. By formulas (6.34) and (6.27)

$$\operatorname{Var}\hat{\theta}_{A^{-1}} = \frac{1}{pd^2n}\operatorname{Tr}((A\theta + B)A^{-1}(A\theta + B)A^{-1}) = \frac{(\theta + c)^2}{pdn}$$

On the other hand, by (6.35) and (6.27)

$$\frac{1}{nJ(\theta)} = \frac{1}{np\operatorname{Tr}(A(A\theta + B)^{-1}A(A\theta + B)^{-1})} = \frac{1}{np(\theta + c)^{-2}d}.$$

Thus $\operatorname{Var}\hat{\theta} = \frac{1}{nJ(\theta)}$ and the estimator $\hat{\theta}_{A^{-1}}$ is efficient.

The second part of the Theorem follows by necessary reparametrizations. For (2a), using (6.18), we write $\theta C + D = -2m$ with $m = \theta A + B$, where $A = -\frac{C}{2}$ and $B = -\frac{D}{2}$. The part (2b) follows similarly from (6.26). □

Remark 6.3.17. It is an open question whether $\hat{\theta}_{A^{-1}}$ may be efficient for independent A and B. Let $n = 1$. The equality $\operatorname{Var} \hat{\theta} = \frac{1}{J(\theta)}$ holds if and only if, writing $D_\theta = (A\theta + B)A^{-1}(A\theta + B)A^{-1}$, the equality $\frac{1}{d^2} \operatorname{Tr}(D_\theta) = \frac{1}{\operatorname{Tr}(D_\theta^{-1})}$ holds for all $\theta \in I$.

Conclusion

In this thesis, we propose some parameter parsimonious models which are of great importance in high dimensional data analysis. We first provide a motivation for the work and critically discuss the available literature on the subject. Then, we present a background of classical Wishart distributions and multiparameter Wishart distributions in terms of the canonical measures of cones. Next, we provide a harmonious construction of Wishart exponential families in nearest neighbours interactions graphical models. Our simple method is based on analysis on convex cones compared to existent work which relies more on graph theory. We define the Wishart distributions and explicitly determine their classical objects, such as the Riesz generating measures, the Wishart densities, the Laplace transforms and the mean functions. Wishart distributions on Q_{A_n} are constructed as the exponential family generated from the gamma function Q_{A_n}, defined by $\Gamma^{(M)}_{Q_{A_n}}(\underline{s}) = \int_{Q_{A_n}} e^{-\operatorname{Tr}(x)} \delta^{(M)}_{\underline{s}}(x)\varphi_{A_n}(x)dx$. Wishart distributions on P_{A_n} are then constructed as the Diaconis-Ylvisaker conjugate family for the exponential family of Wishart distributions on Q_G. For Wishart distributions on Q_{A_n}, explicit formulas for the inverse mean map and the variance function are derived. Later, the methods of construction of Wishart laws introduced in this thesis are used to solve the Letac-Massam Conjecture on the set of parameters of type I Wishart distributions on Q_{A_n} $n \geqslant 1$. Finally, we introduce and study exponential families of distributions parametrized by a segment of means with an emphasis on their Fisher information. This class of models will be useful in high-dimensional data analysis, particularly when one is hesitating between two parameter values. We derive the mean function, the variance

function and the Fisher information of the model. We also propose some estimators and explore their properties. The particular cases of Gaussian and Wishart exponential families parametrized by a segment of means are examined.

The work presented in Chapter 4 has resulted in the article (Graczyk et al., 2016b) written in collaboration with Piotr Graczyk and Hideyuki Ishi. It has been presented at the "Séminaire triangulaire Probabilités et Statistique" in Le Mans (France) in June 2015, and at the summer school "Mathematical Methods of Statistics" in Angers in June 2016. Based on the work presented in Chapter 5, an article (Graczyk et al., 2016c) was written in collaboration with Piotr Graczyk, Hideyuki Ishi and Hiroyuki Ochiai. The work presented in Chapter 6 has resulted in the article (Graczyk and Mamane, 2015) published in collaboration with Piotr Graczyk. It has been presented at the "Séminaire tournant Probabilités et Statistique", in Poitiers (France) in June 2015. The methods and tools developed in this thesis can be used for a future generalization of the construction of graphical Wishart exponential families to decomposable graphs. This generalization will set ground to the solution of the Letac-Massam conjecture in general. Future research can also explore concrete applications of the models proposed in this thesis.

References

Abdelwahab, A., Amor, O. and Abdelwahed, T. (2008), The analysis of the interdependence structure in international financial markets by graphical models, *International research journal of finance and economics*, vol. 15, pp. 291–306.

Andersson, S. and Wojnar, G. (2004), Wishart distributions on homogeneous cones, *Journal of Theoretical Probability*, vol. 17, no. 4, pp. 781–818.

Andersson, S.A. and Klein, T. (2010), On Riesz and Wishart distributions associated with decomposable undirected graphs, *Journal of Multivariate Analysis*, vol. 101, no. 4, pp. 789–810.

Andersson, T. and Handel, P. (2006), Cramer-Rao bound and the parsimony principle, *Instrumentation and Measurement, IEEE Transactions*, vol. 55, no. 1, pp. 44–53.

Avventi, E., Lindquist, A.G. and Wahlberg, B. (2013), ARMA identification of graphical models, *Automatic Control, IEEE Transactions on*, vol. 58, no. 5, pp. 1167–1178.

Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008), Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data, *The Journal of Machine Learning Research*, vol. 9, pp. 485–516.

Barndorff-Nielsen, O.E. (1978), *Information and Exponential Families in Statistical Theory*, John Wiley and Sons.

Ben-David, E. and Rajaratnam, B. (2014), The Letac-Massam conjecture and existence of high dimensional Bayes estimators for graphical models, *arXiv preprint arXiv:1408.0788*.

Bercher, J.F. and Vignat, C. (2009), On minimum Fisher information distributions with restricted support and fixed variance, *Information Sciences*, vol. 179, pp. 3832–3842.

Berrington, A., Hu, Y., Smith, P.W. and Sturgis, P. (2008), A graphical chain model for reciprocal relationships between women's gender role attitudes and labour force participation, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 1, pp. 89–108.

Bishop, C.M. (2006), *Pattern Recognition and Machine Learning*, Springer Science and Business Media.

Blalock, H.M. (1971), *Causal models in the social sciences*, Aldine-Atheston, Chicago.

Boutouria, I. and Hassairi, A. (2009), Riesz exponential families on homogeneous cones, *arXiv preprint arXiv:0906.1892*.

Bressoud, D.M. (1999), *Proofs and Confirmations: The Story of the Alternating-Sign Matrix Conjecture*, Cambridge University Press.

Brown, L.D. (1986), *Fundamentals of statistical exponential families with applications in statistical decision theory*, Lecture Notes-monograph, Institute of Mathematical Statistics, Hayward.

Bru, M.F. (1989), Diffusions of perturbed principal component analysis, *Journal of Multivariate Analysis*, vol. 29, pp. 127–136.

Bru, M.F. (1991), Wishart processes, *Journal of Theoretical Probability*, vol. 4, no. 4, pp. 725–751.

Callies, U., Kurowicka, D. and Cooke, R. (2003), Graphical models for the evaluation of multisite temperature forecasts: comparison of vines and independence graphs, *Proc. of ESREL 2003, Safety and Reliability*, vol. 1, pp. 363–371.

Caputo, A., Foraita, R., Klasen, S. and Pigeot, I. (2003), Undernutrition in Benin: an analysis based on graphical models, *Social science & medicine*, vol. 56, no. 8, pp. 1677–1691.

Carvalho, C. and West, M. (2007), Dynamic matrix-variate graphical models, *Bayesian Analysis*, vol. 2, no. 1, pp. 69–98.

Carvalho, C.M., Massam, H. and West, M. (2007), Simulation of hyper-inverse Wishart distributions in graphical models, *Biometrika*, vol. 94, no. 3, pp. 647–659.

Casalis, M. and Letac, G. (1996), The Lukacs-Olkin-Rubin characterization of Wishart distributions on symmetric cones, *The Annals of Statistics*, vol. 24, no. 2, pp. 763–786.

Casella, G. and Berger, R. (2002), *Statistical Inference*, Duxbury Advances Series.

Chandran, L.S., Ibarra, L., Ruskey, F. and Sawada, J. (2003), Generating and characterizing the perfect elimination orderings of a chordal graph, *Theoretical Computer Science*, vol. 307, no. 2, pp. 303–317.

Chenevier, G. and Renard, D. (2008), Characters of speh representations and lewis caroll identity, *Representation Theory of the American Mathematical Society*, vol. 12, no. 18, pp. 447–452.

Cox, D.R. and Wermuth, N. (1996), *Multivariate Dependencies Models*, Chapman and Hall.

Dahlhaus, R. (2000), Graphical interaction models for multivariate time series, *Metrika*, vol. 51, no. 2, pp. 157–172.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980), Markov fields and log-linear interaction models for contingency tables, *The Annals of Statistics*, pp. 522–539.

Dawid, A.P. and Lauritzen, S.L. (1993), Hyper Markov laws in the statistical analysis of decomposable graphical models, *The Annals of Statistics*, pp. 1272–1317.

Dempster, A.P. (1972), Covariance selection, *Biometrics*, pp. 157–175.

Diaconis, P. and Ylvisaker, D. (1979), Conjugate priors for exponential families, *The Annals of statistics*, vol. 7, no. 2, pp. 269–281.

Dobra, A., Eicher, T.S. and Lenkoski, A. (2010), Modeling uncertainty in macroeconomic growth determinants using Gaussian graphical models, *Statistical Methodology*, vol. 7, no. 3, pp. 292–306.

Eaton, M.L. (2007), *Multivariate Statistics*, Institute of Mathematical Statistics.

Edwards, D. (2000), *Introduction to graphical modelling*, Springer-Verlag New York.

Eichler, M. (2012), Graphical modelling of multivariate time series, *Probability Theory and Related Fields*, vol. 153, no. 1-2, pp. 233–268.

Faraut, J. and Korányi, A. (1994), *Analysis on symmetric cones*, Oxford Mathematical Monographs, Oxford University Press, New York.

Frieden, B.R. (2004), *Science from Fisher information: a unification*, Cambridge University Press.

Friedman, J., Hastie, T. and Tibshirani, R. (2008), Sparse inverse covariance estimation with the graphical lasso, *Biostatistics*, vol. 9, no. 3, pp. 432–441.

Gather, U., Imhoff, M. and Fried, R. (2002), Graphical models for multivariate time series from intensive care monitoring, *Statistics in medicine*, vol. 21, no. 18, pp. 2685–2701.

Gibbs, J.W. (1902), *Elementary principles in statistical mechanics*, Yale University Press.

Glunt, W., Hayden, T., Johnson, C.R. and Tarazaga, P. (1999), Positive definite completions and determinant maximization, *Linear algebra and its applications*, vol. 288, pp. 1–10.

Goodman, L.A. (1970), The multivariate analysis of qualitative data: Interactions among multiple classifications, *Journal of the American Statistical Association*, vol. 65, no. 329, pp. 226–256.

Graczyk, P. and Ishi, H. (2014), Riesz measures and Wishart laws associated to quadratic maps, *Journal of the Mathematical Society of Japan*, vol. 66, no. 1, pp. 317–348.

Graczyk, P., Ishi, H. and Kołodziejek, B. (2016a), Wishart exponential families and variance function on homogeneous cones, *Submitted*.

Graczyk, P., Ishi, H. and Mamane, S. (2016b), Riesz and Wishart distributions on the cones related to A_n graphs, *Submitted*.

Graczyk, P., Ishi, H., Mamane, S. and Ochiai, H. (2016c), On the Letac-Massam conjecture on cones Q_{A_n}, *Submitted*.

Graczyk, P. and Mamane, S. (2015), Fisher information and exponential families parametrized by a segment of means, *Probability and Mathematical Statistics*, vol. 35, no. 1, pp. 73–90.

Gray, R.M. (2006), *Toeplitz and Circulant Matrices: A review*, Foundations and Trends in Communications and Information Theory, Vol. 2 Issue 3, NOW Publishers.

Grone, R., Johnson, C.R., Sá, E.M. and Wolkowicz, H. (1984), Positive definite completions of partial hermitian matrices, *Linear algebra and its applications*, vol. 58, pp. 109–124.

Gupta, R.D. and Kundu, D. (2006), On the comparison of Fisher information of the Weibull and GE distributions, *Journal of Statistical Planning and Inference*, vol. 136, pp. 3130–3144.

Gutiérrez-Peña, E. and Smith (1997), Exponential and Bayesian conjugate families: review and extensions, *Test*, vol. 6, no. 1, pp. 1–90.

Hand, D., McConway, K. and Stanghellini, E. (1997), Graphical models of applicants for credit, *IMA Journal of Management Mathematics*, vol. 8, no. 2, pp. 143–155.

Hassairi, A. and Lajmi, S. (2001), Riesz exponential families on symmetric cones, *Journal of Theoretical Probability*, vol. 14, no. 4, pp. 927–948.

Hastie, T., Tibshirani, R. and Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer-Verlag New York.

Højsgaard, S., Edwards, D. and Lauritzen, S. (2012), *Graphical models with R*, Springer Science & Business Media.

Irvine, K.M. and Gitelman, A.I. (2011), Graphical spatial models: a new view on interpreting spatial pattern, *Environmental and ecological statistics*, vol. 18, no. 3, pp. 447–469.

Ishi, H. (2014), Homogeneous cones and their applications to statistics, in *Modern methods of multivariate statistics*, vol. 82, Hermann, Paris, pp. 135–154.

Johnson, C.R. (1990), Matrix completion problems: a survey, in *Matrix theory and applications*, vol. 40, Providence, RI, pp. 171–198.

Jordan, M.I., Ghahramani, Z., Jaakkola, T.S. and Saul, L.K. (1999), An introduction to variational methods for graphical models, *Machine learning*, vol. 37, no. 2, pp. 183–233.

Khare, K. and Rajaratnam, B. (2011), Wishart distributions for decomposable covariance graph models, *The Annals of Statistics*, pp. 514–555.

Koller, D. and Friedman, N. (2009), *Probabilistic Graphical Models*, MIT Press.

Konno, Y. (2007), Estimation of normal covariance matrices parametrized by irreducible symmetric cones under stein's loss, *Journal of Multivariate Analysis*, vol. 98, no. 2, pp. 295–316.

Konno, Y. (2009), Shrinkage estimators for large covariance matrices in multivariate real and complex normal distributions under an invariant quadratic loss, *Journal of Multivariate Analysis*, vol. 100, no. 10, pp. 2237–2253.

Kuriki, S. and Numata, Y. (2010), Graph presentations for moments of noncentral wishart distributions and their applications, *Annals of the Institute of Statistical Mathematics*, vol. 62, no. 4, pp. 645–672.

Laurent, M. (1998), A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems, *Topics in Semidefinite and Interior-Point Methods*, vol. 18, pp. 51–76.

Lauritzen, S.L. (1995), The EM algorithm for graphical association models with missing data, *Computational Statistics & Data Analysis*, vol. 19, no. 2, pp. 191–201.

Lauritzen, S.L. (1996), *Graphical models*, Oxford University Press.

Lauritzen, S.L. and Sheehan, N.A. (2003), Graphical models for genetic analyses, *Statistical Science*, vol. 18, no. 4, pp. 489–514.

Lauritzen, S.L. and Spiegelhalter, D.J. (1988), Local computations with probabilities on graphical structures and their application to expert systems, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 157–224.

Lauritzen, S.L. and Wermuth, N. (1989), Graphical models for associations between variables, some of which are qualitative and some quantitative, *The Annals of Statistics*, pp. 31–57.

Lehmann, E. and Casella, G. (1998), *Theory of point estimation*, Springer Texts in Statistics. Springer-Verlag, New York.

Lehmann, E. and Romano, J. (2005), *Testing Statistical Hypotheses*, Springer Texts in Statistics. Springer-Verlag, New York.

Letac, G. (1989), A characterization of the Wishart exponential families by an invariance property, *Journal of Theoretical Probability*, vol. 2, no. 1, pp. 71–86.

Letac, G. (1992), *Lectures on natural exponential families and their variance functions*, CNDCT, Instituto de Matemática Pura e Aplicada, Rio de Janeiro.

Letac, G. (2012), Private communication.

Letac, G. (2014), Decomposable graphs, in *Modern methods of multivariate statistics*, vol. 82, Hermann, Paris, pp. 155–196.

Letac, G. and Casalis, M. (2000), Natural exponential families, Wiley Series in Probability and Statistics, pp. 659–696.

Letac, G. and Massam, H. (2007), Wishart distributions for decomposable graphs, *The Annals of Statistics*, vol. 35, no. 3, pp. 1278–1323.

Letac, G. and Massam, H. (2008), The noncentral Wishart as an exponential family, and its moments, *Journal of Multivariate Analysis*, vol. 99, no. 7.

Letac, G. and Massam, H. (2011), Existence and non-existence of the noncentral Wishart distributions, *arXiv preprint arXiv:1108.2849*.

Magnus, J.R. and Neudecker, H. (2007), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

Mason, J.C. and Handscomb, D.C. (2003), *Chebyshev polynomials*, Chapman & Hall/CRC.

Mathai, A.M., Provost, S.B. and Hayakawa, T. (2012), *Bilinear forms and zonal polynomials*, vol. 102, Springer Science & Business Media.

Mohamed, W.N., Diamond, I. and Smith, P.W. (1998), The determinants of infant mortality in malaysia: a graphical chain modelling approach, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 161, no. 3, pp. 349–366.

Muirhead, R.J. (2005), *Aspects of multivariate statistical theory*, John Wiley & Sons, New Jersey.

Murphy, K., Torralba, A. and Freeman, W. (2003), Using the forest to see the trees: a graphical model relating features, objects and scenes, *Advances in neural information processing systems*, vol. 16, pp. 1499–1506.

Murphy, K.P. (2012), *Machine learning: a probabilistic perspective*, MIT press.

Oliver, N. and Pentland, A.P. (2000), Graphical models for driver behavior recognition in a smartcar, in *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, IEEE, pp. 7–12.

Paulsen, V.I., Power, S.C. and Smith, R.R. (1989), Schur products and matrix completions, *Journal of Functional Analysis*, vol. 85, no. 1, pp. 151–178.

Philipov, A. and Glickman, M. (2006), Multivariate stochastic volatility via Wishart processes, *Journal of Business and Economic Statistics*, vol. 24, no. 3, pp. 313–328.

Ramirez, V.R., Hernandez, P.Q., Castro, S.H. and Diwekar, U.M. (2010), Fisher information as a novel tool for process control applications, *20th European Symposium on Computer Aided Process Engineering - ESCAPE20 S. Pierucci and G. Buzzi Ferraris (Editors), 2010, Elsevier B.V.*

Roverato, A. (2000), Cholesky decomposition of a hyper inverse Wishart matrix, *Biometrika*, vol. 87, no. 1, pp. 99–112.

Sewart, P. and Whittaker, J. (1998), Graphical models in credit scoring, *IMA Journal of Management Mathematics*, vol. 9, no. 3, pp. 241–266.

Smith, P.W., Berrington, A. and Sturgis, P. (2009), A comparison of graphical models and structural equation models for the analysis of longitudinal survey data, *Methodology of Longitudinal Surveys*, pp. 381–392.

Songsiri, J., Dahl, J. and Vandenberghe, L. (2010), Graphical models of autoregressive processes, *Convex Optimization in Signal Processing and Communications*, pp. 89–116.

Speed, T. and Kiiveri, H. (1986), Gaussian Markov distributions over finite graphs, *The Annals of Statistics*, pp. 138–150.

Sugiura, N. and Konno, Y. (1988), Entropy loss and risk of improved estimators for the generalized variance and precision, *Annals of the Institute of Statistical Mathematics*, vol. 40, no. 2, pp. 329–341.

Tsukuma, H. and Konno, Y. (2006), On improved estimation of normal precision matrix and discriminant coefficients, *Journal of Multivariate Analysis*, vol. 97, no. 7, pp. 1477–1500.

Wainwright, M.J. and Jordan, M.I. (2008), *Graphical models, exponential families, and variational inference*, vol. 1, Now Publishers Inc.

Wermuth, N. (1976), Analogies between multiplicative models in contingency tables and covariance selection, *Biometrics*, pp. 95–108.

Whittaker, J. (1990), *Graphical models in applied multivariate analysis*, John Wiley & Sons.

Wishart, J. (1928), The generalised product moment distribution in samples from a normal multivariate population, *Biometrika*, pp. 32–52.

Wold, H. (1954), Causality and econometrics, *Econometrica: Journal of the Econometric Society*, pp. 162–177.

Wright, S. (1921), Correlation and causation, *Journal of agricultural research*, vol. 20, no. 7, pp. 557–585.

Wright, S. (1934), The method of path coefficients, *The Annals of Mathematical Statistics*, vol. 5, no. 3, pp. 161–215.