Subspace Identification Via Convex Optimization

by

James Saunderson

BE (Hons) in Electrical Engineering and BSc (Hons) in Mathematics
The University of Melbourne, 2008

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2011

Author ....
Department of Electrical Engineering and Computer Science
May 20, 2011

Certified by ....
Pablo A. Parrilo
Professor of Electrical Engineering
Thesis Supervisor

Certified by ....
Alan S. Willsky
Edwin Sibley Webster Professor of Electrical Engineering
Thesis Supervisor

Accepted by ....
Leslie A. Kolodziejski
Chair, Department Committee on Graduate Students

Subspace Identification Via Convex Optimization

by

James Saunderson

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

In this thesis we consider convex optimization-based approaches to the classical problem of identifying a subspace from noisy measurements of a random process taking values in the subspace. We focus on the case where the measurement noise is component-wise independent, known as the factor analysis model in statistics.

We develop a new analysis of an existing convex optimization-based heuristic for this problem. Our analysis indicates that in high-dimensional settings, where both the ambient dimension and the dimension of the subspace to be identified are large, the convex heuristic, minimum trace factor analysis, is often very successful. We provide simple deterministic conditions on the underlying 'true' subspace under which the convex heuristic provably identifies the correct subspace. We also consider the performance of minimum trace factor analysis on 'typical' subspace identification problems, that is problems where the underlying subspace is chosen randomly from subspaces of a particular dimension. In this setting we establish conditions on the ambient dimension and the dimension of the underlying subspace under which the convex heuristic identifies the subspace correctly with high probability.

We then consider a refinement of the subspace identification problem where we aim to identify a class of structured subspaces arising from Gaussian latent tree models. More precisely, given the covariance at the finest scale of a Gaussian latent tree model, and the tree that indexes the model, we aim to learn the parameters of the model, including the state dimensions of each of the latent variables. We do so by extending the convex heuristic, and our analysis, from the factor analysis setting to the setting of Gaussian latent tree models. We again provide deterministic conditions on the underlying latent tree model that ensure our convex optimization-based heuristic successfully identifies the parameters and state dimensions of the model.

Thesis Supervisor: Pablo A. Parrilo
Title: Professor of Electrical Engineering

Thesis Supervisor: Alan S. Willsky
Title: Edwin Sibley Webster Professor of Electrical Engineering

Acknowledgments

I am very fortunate to have Alan Willsky and Pablo Parrilo as my research advisors. Alan provides a constant stream of research ideas (perhaps even programs) and encouragement, is a model of organization, and fosters the supportive and collaborative environment of the Stochastic Signals Group. Pablo has taught me, among many things, the value of developing intuition through computation and the value of thinking both cleanly and clearly.

It has been a pleasure to collaborate with Venkat Chandrasekaran over the past two years—indeed the main thread of this thesis came out of a discussion we had over lunch one day. I have also enjoyed sharing an office with Matt Johnson, for his friendship and willingness to act as a sounding board for my thoughts. I would like to thank the other current and former members of the Stochastic Signals Group with whom I have overlapped—Ying Liu, Vincent Tan, Jin Choi, and Kush Varshney—for their part in fostering the stimulating, fun, and positive environment in which I have had the fortune to work.

Finally I owe a great debt to Been Kim for her understanding and for reminding me of what is and is not important, and, above all, my parents for their quiet, constant, support in all things.

Contents

Chapter 1

Introduction

Many modern problems in engineering, science, and beyond involve analyzing, understanding, and working with high dimensional data. This poses challenges in terms of interpretability and computation—performing inference tasks with complex high-dimensional models is generally intractable. One common approach to dealing with high dimensional complex data is to build a parsimonious statistical model to explain the data. Very often such models assume the high dimensional data arise as perturbations (by 'noise') of points in a low-dimensional object, such as a manifold. Modeling, in this context, involves trying to identify the low-dimensional object, and perhaps some parameters of the noise model, from the observed data. Depending on the assumptions on the noise, even the simplest such models, where the low-dimensional object is a linear subspace, are challenging to work with, especially in high-dimensional settings. In this thesis we consider two related subspace identification problems, and convex optimization-based methods for solving them in high-dimensional settings.

1.1 Subspace Identification

We now introduce the two problems on which this thesis focuses and indicate how they can be thought of as subspace identification problems.

1.1.1 Factor Analysis

Suppose z_h is a zero-mean Gaussian[1] random variable taking values in a subspace $U^\star \subset \mathbb{R}^n$. Suppose we do not have access to z_h, instead having access to $x_o = z_h + w_o$ where w_o can be thought of as zero-mean Gaussian noise with unknown covariance. Given x_o, we would like to form an estimate

[1] Throughout this thesis we only consider first and second order statistics of random variables, so we can make this assumption without loss of generality.

$\hat{U}$ of the subspace in which z_h lies. Since we would like to have as simple a model as possible for z_h, we would like to find take $\hat{U}$ to be the lowest dimensional subspace in which z_h could possibly lie. With no further assumptions on the model, the problem is trivial: assume that $z_h = 0$ (and so $\hat{U} = \{0\}$) and that all of the measurement is just noise.

For the model to make sense we must make an assumption about the noise. The assumption we make is that w_o is independent of z_h and the components of w_o are independent, that is $\mathbb{E}[w_o w_o^T] = Q_o$ is diagonal. The resulting model is known as *factor analysis* [2] in the statistics literature [53, 54] and the *Frisch scheme* in the system identification literature [27, 32] and has received a great deal of attention since its introduction in the early twentieth century. We discuss a selection of previous work related to this model in Chapter 3.

If A is a matrix with column space $U^\star$, we can write $z_h = Ax_h$ and think of factor analysis as a linear Gaussian model:

$$x_h \sim \mathcal{N}(0, R), \quad w_o \sim \mathcal{N}(0, Q_o), \quad x_o = Ax_h + w_o, \tag{1.1}$$

that is, each observed variable $x_{o,i}$ is a linear function of x_h with additive noise $w_{o,i}$ independent of each $w_{o,j}$ for $j \neq i$. Taking covariances we have that

$$\mathbb{E}[x_o x_o^T] = \Sigma_o = ARA^T + Q_o \tag{1.2}$$

where Q_o is diagonal and ARA^T is low rank, with rank equal to the dimension of the subspace $U^\star$. From (1.2) we see that factor analysis can be thought of, concretely, as the problem of decomposing a positive semidefinite matrix Σ_o into the sum of a positive semidefinite diagonal matrix Q_o and a positive semidefinite matrix ARA^T with the smallest rank possible. Clearly, if we can correctly decompose Σ_o into its diagonal and low-rank constituents, we can identify the subspace $U^\star$, as we originally set out to achieve.

Despite our additional assumption on the noise, this matrix decomposition problem is not well-posed in general—it is quite possible that more than one decomposition exists with the smallest possible rank for the low-rank term. Furthermore, even if the problem is identifiable, there is no known general, computationally tractable, procedure for performing this matrix decomposition. In Chapter 3 of this thesis we analyze a convex optimization-based heuristic, called minimum trace factor analysis, for factor analysis, showing that this heuristic is often very successful.

The noise model Component-wise independence of the noise is a strong assumption, but it arises naturally in a number of applications. For example, if the $w_{o,i}$ are the measurements made at physically well separated sensors in a sensor network, one would expect the associated measurement

[2] When we refer to factor analysis in this thesis we mean precisely this model, as distinct from principal components analysis, which makes quite different assumptions about the noise.

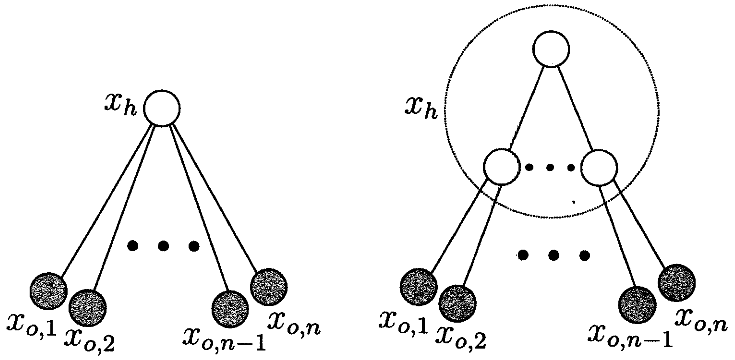

Figure 1-1: On the left is the Gaussian latent tree corresponding to the factor analysis model. If the hidden variable x_h in the factor analysis model is given more structure, as in the diagram on the right, we obtain a more general Gaussian latent tree model.

noise at any sensor to be independent of the measurement noise at any other sensor. Note that factor analysis does *not* assume that we know the noise variances, or that the noise variances are the same. Also, by changing basis, we need only assume that the noise is component-wise independent in *some* basis that is known *a priori*, rather than the measurement basis.

Covariance information Unless otherwise stated, throughout this thesis we assume that we have access to the population covariance of the observed random variable x_o. In terms of the matrix decomposition problem (1.2), this assumption means that the matrix Σ_o can be exactly expressed as the sum of a diagonal and a low rank positive semidefinite matrix, and so our aim is to exactly decompose Σ_o into these pieces. In this thesis we analyze only this exact decomposition problem. In practice we only expect to have access to samples of x_o and hence would use a sample covariance matrix. We discuss this issue further in Section 4.5 and in the further work section of Chapter 5.

1.1.2 Gaussian Latent Trees

In the factor analysis model, we try to identify a (potentially) arbitrary subspace $U^\star$ from measurements with component-wise independent additive noise. In Chapter 4 of this thesis, we consider refinements of this model where the subspace $U^\star$ has more structure.

The way in which we impose more structure on $U^\star$ is most easily seen by considering the representation of factor analysis as a linear Gaussian model in (1.1). This representation can be interpreted as a linear state-space model with states indexed by the vertices of a star shaped tree (see Figure 1-1). The hidden state x_h corresponds to the root of the tree and the observed states

$x_{o,i}$ for $i = 1, 2, \ldots, n$ correspond to the leaves of the tree, with each $x_{o,i}$ being a linear function (with additive noise) of its parent x_h. The dimension of the state x_h is precisely the dimension of the subspace $U^\star$ we aim to identify.

Generalizing this picture, we can consider a state-space model driven by Gaussian noise with states indexed by the vertices of a more general tree. In this case, the state at each vertex is a linear function of the state at its parent with additive noise that is independent of the noise at any other vertex. As in the factor analysis model, we assume we only observe the leaf variables of the process. Such models are referred to as Gaussian latent tree models in the literature on probabilistic graphical models [35], and multiscale autoregressive models [5] in the multiscale modeling literature. Given the index tree and the covariance among the leaf variables, the problem of determining the dimensions of the state spaces at each vertex, and the linear maps defining the relationships between the states at each vertex is the focus of Chapter 4.

1.2 A Motivating Application—Direction of Arrival Estimation in Sensor Networks

One motivating problem where a factor analysis model naturally arises, but is not typically used, is the direction of arrival estimation problem in array signal processing. Suppose we have N sensors at locations $x_1, x_2, \ldots, x_N \in \mathbb{R}^2$. Suppose these sensors are passively 'listening' for waves (electromagnetic or acoustic) from $K < N$ sources in the far field. Let the wave vectors of the sources be $k_1, k_2, \ldots, k_K \in \mathbb{R}^3$. When can we determine the number of sources K and their directions of arrival (DOA) given the sensor measurements? This problem has enjoyed great attention since the early days of wireless communications, in the early 20th century, and has associated with it a vast literature. The recent book of Tuncer and Friedlander [57] as well as the survey of Krim and Viberg [36] discuss many aspects of the direction of arrival estimation problem.

The standard mathematical model for this problem (see [36] for a derivation) is to model the vector of measurements $x(t)$ made by the sensors as

$$x(t) = As(t) + n(t) \tag{1.3}$$

where $s(t)$ is the vector of baseband signal waveforms from the sources $n(t)$ is the vector of sensor noise, and

$$[A]_{i\ell} = e^{-j\langle x_i, k_\ell \rangle} \tag{1.4}$$

where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product on $\mathbb{R}^3$ and $j = \sqrt{-1}$ is the complex unit. Typically $s(t)$ and $n(t)$ are modeled as a stationary white Gaussian processes with covariances $\mathbb{E}[s(t)s(t)^H] = P$ and $\mathbb{E}[n(t)n(t)^H] = D$ respectively (where A^H denotes the Hermitian transpose of A). Furthermore $s(t)$ and $n(t)$ are typically assumed to be independent. It is often natural to assume that the noise at each sensor is independent of the noise at every other sensor. As such we assume that D is diagonal.

Subspace-based approaches to DOA estimation Under these assumptions, one approach to estimating the k_ℓ involves decomposing the covariance of the measurements

$$\Sigma = \mathbb{E}[y(t)y(t)^H] = APA^H + D$$

into the contribution from the signal, APA^H, and the contribution from the noise, D, and then trying to estimate the k_ℓ from an estimate of APA^H. Observe that the rank of APA^H is the number of sources, which is assumed to be smaller than the number of sensors. Under our assumption on the sensor noise, we are faced with the problem of decomposing Σ into a diagonal and a low-rank matrix.

Subspace based approaches to direction of arrival estimation rose to prominence in the 1980s with the MUSIC [9, 50] and ESPRIT [49] algorithms, to name two more celebrated approaches. Both of these, and the many other approaches along the same lines, consist of two rather distinct steps:

1. *Estimating the signal subspace:* Given $\Sigma = APA^H + D$ estimate the column space of A.
2. *Estimating the source directions:* Use the estimate of the column space of A to estimate the wave vectors k_ℓ.

These classical subspace-based algorithm for direction of arrival estimation make the additional assumptions that the variance of the noise at each sensor is the same, so that $\mathbb{E}[n(t)n(t)^H] = \sigma^2 I$ and the number of sources K is known (or has already been estimated). Under these additional assumptions, it is reasonable to take the span of the K eigenvectors of Σ corresponding to the K largest eigenvalues as the signal subspace. This is the method of subspace estimation used in all of these classical subspace methods.

Suppose we are in the setting where we have a large number of sensors, perhaps of different types operating under different operating conditions. It is likely that the noise variances at all of the sensors are quite different. It is also unlikely that the number of sources is known *a priori*. As such, the assumptions, on which the classical method of estimating the signal subspace using eigenvector computations is based, are not appropriate. However, in this setting directly using a factor analysis model, the central model of interest in Chapter 3, is quite natural. Indeed in Chapter 3 we illustrate how convex optimization-based heuristics for factor analysis can provide an alternative method of identifying the signal subspace in the setting where the sensor variances are all different and potentially large.

1.3 Summary of Contributions

Our first main contribution is the development of a new analysis of minimum trace factor analysis, an existing convex optimization-based heuristic for factor analysis. In particular, our analysis focuses on understanding how well minimum trace factor analysis works as a heuristic in the high-dimensional setting where the ambient dimension is large *and* the dimension of the subspace we are trying to identify is also large. This setting is becoming increasingly important in applications—modern sensor networks, for example, consist of large numbers of cheap sensors spread our over a large area, aiming to detect as many sources as possible. From an analytical point of view, focusing on this setting allows us to perform analysis that is coarse enough to be tractable, and also coarse enough to provide useful intuition about the performance of the heuristic we analyze.

More specifically the two main aspects of our analysis are the following.

- We establish simple deterministic conditions on a subspace that ensure minimum trace factor analysis can identify that subspace. These conditions are expressed in terms of incoherence parameters that can easily be translated into more problem specific information (as we do for the direction of arrival estimation problem).
- We establish conditions on the rank of a random subspace that ensure that subspace can be identified by minimum trace factor analysis with high probability.

The key intuition to be extracted from these results is that minimum trace factor analysis is a good heuristic for factor analysis in the high-dimensional regime.

The second main contribution of the thesis is that we describe a way to generalize the computational methods and deterministic analysis of minimum trace factor analysis to the problem of learning the parameters and state dimensions of Gaussian latent tree models. In particular we formulate a semidefinite program that, given the covariance among the leaf variables and the tree structure, attempts to recover the parameters and state dimensions of an underlying Gaussian latent tree model.

We then analyze this semidefinite program, giving two deterministic conditions on the underlying tree model that ensure our semidefinite program correctly recovers the parameters and state dimensions of the model. The first condition characterizes when recovery is possible in the case where the underlying tree model has all scalar-valued variables. The second condition directly generalizes our deterministic conditions for recovery in the case of diagonal and low-rank decompositions. In the course of our analysis, we show how to reduce the analysis of our semidefinite program to the analysis of block diagonal and low-rank decomposition problems.

1.4 Organization of the Thesis

The remainder of the thesis is organized in the following way. Chapter 2 summarizes some of the technical background required for the subsequent chapters. Chapter 3 considers the diagonal and low rank decomposition problem, including a discussion of previous work on the problem and our analysis of the semidefinite programming-based heuristic, minimum trace factor analysis, for this problem. Chapter 4 extends these methods and analysis to the problem of learning the parameters and state dimensions of a Gaussian latent tree model given the index tree and observations of the leaf-indexed variables. Finally Chapter 5 summarizes our contributions and discusses potential future research directions that arise naturally from the content of this thesis.

Chapter 2

Background

In this chapter we summarize some technical background required for the remainder of the thesis. In Section 2.1 we briefly introduce the notation we use for the linear algebra and matrix analysis that is present throughout the thesis. In Section 2.2 we discuss convex heuristics for non-convex optimization problems, with a focus on rank minimization problems, before reviewing key results about semidefinite programs, the class of convex optimization problems that arise in our work. Finally in Section 2.3 we briefly consider the concentration of measure phenomenon, with a focus on functions of Gaussian random variables, reviewing results that play an important role in our randomized analysis in Chapter 3.

2.1 Notation

2.1.1 Basic Notions

Most of the linear algebra we do will be over either the n-dimensional real vector space $\mathbb{R}^n$ or the $\binom{k+1}{2}$-dimensional vector space of $k \times k$ symmetric matrices, which we denote by $\mathcal{S}^k$. We equip $\mathbb{R}^n$ with the standard Euclidean inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ and $\mathcal{S}^k$ with the trace inner product $\langle X, Y\rangle = \operatorname{tr}(X^T Y) = \sum_{i=1}^k \sum_{j=1}^k X_{ij} Y_{ij}$.

On two occasions we use complex scalars and so work in $\mathbb{C}^n$ and the set of $k \times k$ Hermitian matrices. In this case we use $\langle x, y\rangle = \sum_{i=1}^n \bar{x}_i y_i$ and $\langle X, Y\rangle = \operatorname{tr}(X^H Y) = \sum_{i=1}^k \sum_{j=1}^k \bar{X}_{ij} Y_{ij}$ where $\bar{x}$ is the complex conjugate of $x \in \mathbb{C}$ and A^H denotes the hermitian transpose of a matrix.

We will almost always think of linear maps $A : \mathbb{R}^n \to \mathbb{R}^m$ as $m \times n$ matrices with respect to the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, and will concretely denote the adjoint of A by A^T. By contrast, we often prefer to think of linear maps $\mathcal{A} : \mathcal{S}^k \to \mathbb{R}^n$ abstractly. In this case we denote the adjoint map by $\mathcal{A}^* : \mathbb{R}^n \to \mathcal{S}^k$.

We denote the eigenvalues of a $k \times k$ symmetric matrix X by

$$\lambda_{max}(X) = \lambda_1(X) \geq \cdots \geq \lambda_k(X) = \lambda_{min}(X)$$

and the singular values of an $n \times k$ matrix A by

$$\sigma_{max}(A) = \sigma_1(A) \geq \cdots \geq \sigma_{\min\{n,k\}}(A) = \sigma_{min}(A) \geq 0.$$

The (Moore-Penrose) pseudoinverse of a matrix A is denoted $A^\dagger$. Note that singular values and the pseudoinverse of a matrix are basis-independent and so are really properties of the underlying linear map.

Finally, the column space or range of a matrix A will be denoted by $\mathcal{R}(A)$.

Projections If V is a subspace of $\mathbb{R}^n$ then we denote by $P_V : \mathbb{R}^n \to \mathbb{R}^n$ the orthogonal (with respect to the Euclidean inner product) projection onto V, that is the unique map that satisfies $P_V(v) = v$ for all $v \in V$ and $P_V(w) = 0$ for all $w \in V^\perp$. If V is an r dimensional subspace of $\mathbb{R}^n$ we use the notation $\pi_V : \mathbb{R}^n \to \mathbb{R}^r$ to indicate some (fixed) choice of map that satisfies $\pi_V^T\pi_V = P_V$ and $\pi_V\pi_V^T = I$. We call such a map a *partial isometry*.

If V is a subspace of $\mathcal{S}^k$ we use calligraphic letters to denote the corresponding orthogonal projection, i.e. $\mathcal{P}_V : \mathcal{S}^k \to \mathcal{S}^k$. In the particular case where we are projecting onto the subspace of diagonal matrices we write $\text{diag} : \mathcal{S}^k \to \mathbb{R}^k$ and $\text{diag}^* : \mathbb{R}^k \to \mathcal{S}^k$ for the maps such that $\text{diag}^*\text{diag} : \mathcal{S}^k \to \mathcal{S}^k$ is the orthogonal projection onto the subspace of diagonal matrices.

Convex Cones A *convex cone* is a convex subset $\mathcal{K}$ of a real vector space V that is closed under multiplication by non-negative real scalars. The two convex cones of most interest to us in this thesis are the non-negative orthant

$$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for } i = 1, 2, \ldots, n\},$$

and the positive semidefinite cone

$$\mathcal{S}^k_+ = \{X \in \mathcal{S}^k : u^TXu \geq 0 \text{ for all } u \in \mathbb{R}^k\}.$$

We denote by $\leq$ and $\preceq$ respectively, the partial order on $\mathbb{R}^n$ induced by $\mathbb{R}^n_+$ and the partial order on $\mathcal{S}^k$ induced by $\mathcal{S}^k_+$.

Given a vector space V and a convex cone $\mathcal{K} \subset V$ we say a linear map $\mathcal{A} : V \to V$ is *cone preserving* if $\mathcal{A}(\mathcal{K}) \subset \mathcal{K}$. Clearly compositions of cone preserving maps are cone preserving. For example, the cone preserving linear maps on $\mathbb{R}^n_+$ correspond to $n \times n$ matrices with non-negative entries.

Hadamard Products We denote the Hadamard (or Schur or entry-wise) product of matrices A and B of compatible dimensions by $A \circ B$. The matrix all entries of which are one is the identity for this product and is denoted J. Perhaps the most basic among the many interesting spectral properties enjoyed by the Hadamard product is that if A and B are positive semidefinite matrices, then $A \circ B$ is also positive semidefinite.

2.1.2 Norms

We employ a variety of matrix and vector norms throughout this thesis.

Vector norms The norms on $\mathbb{R}^n$ of principal interest are the ℓ_p norms for $p \geq 1$ defined by $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ and $\|x\|_\infty = \max_i |x_i|$. Note that if $x \in \mathbb{R}^n$ we often use $\|x\|$ instead of $\|x\|_2$ to denote the Euclidean norm. Given a norm $\|\cdot\|$ on $\mathbb{R}^n$ its dual norm $\|\cdot\|_*$ with respect to the inner product is defined by

$$\|x\|_* = \sup_{\|y\|=1} \langle x, y\rangle.$$

If $1 \leq p, q \leq \infty$ and $1/p + 1/q = 1$ then $\|x\|_p$ and $\|x\|_q$ are dual norms. Hölder's inequality follows from the duality of $\|\cdot\|_p$ and $\|\cdot\|_q$, giving a nice way to relate norms and the standard inner product:

$$\langle x, y\rangle \leq \|x\|_p \|y\|_q$$

where $1 \leq p, q \leq \infty$ and $1/p + 1/q = 1$.

Matrix norms The principal norms on $\mathcal{S}^k$ of interest to us are the Frobenius norm

$$\|X\|_F = \langle X, X\rangle^{1/2} = \left(\sum_{i=1}^k \sum_{j=1}^k X_{ij}^2\right)^{1/2} = \left(\sum_{i=1}^k \sigma_i(X)^2\right)^{1/2}$$

the spectral norm, $\|X\|_{\mathrm{sp}} = \sigma_1(X)$ (which we usually denote by $\|X\|$) and the trace norm

$$\|X\|_* = \sum_{i=1}^k \sigma_i(X).$$

Note that each of these are examples of Schatten p-norms (the ℓ_p norms applied to the singular values of the matrix) and enjoy remarkably analogous behaviour to the corresponding ℓ_p norms on $\mathbb{R}^n$ in certain settings. For example, if $1/p + 1/q = 1$ then the Schatten p and q norms are dual and so obey Hölder's inequality with respect to the trace inner product.

Operator norms Given a linear map A between two normed spaces with norms $\|\cdot\|_a$ and $\|\cdot\|_b$, we denote the corresponding induced norm by

$$\|A\|_{a\to b} = \sup_{x\neq 0}\frac{\|Ax\|_b}{\|x\|_a}.$$

For example, if $\mathcal{A}:\mathbb{R}^n\to\mathcal{S}^k$ where $\mathbb{R}^n$ is equipped with the ℓ_∞ norm and $\mathcal{S}^k$ with the spectral norm, we denote the induced norm by $\|\mathcal{A}\|_{\infty\to\mathrm{sp}}$.

2.2 Convex Relaxations and Semidefinite Programming

At the core of this thesis is the well-known idea that one can often approach intractable, (typically non-convex) optimization problems by solving a related tractable *convex* optimization problem. Furthermore, one can often give conditions on the original intractable problem instance under which the convex relaxation actually solves the original problem. This general principal is of central importance in combinatorial optimization, and is of increasing significance in signal processing [14,42], statistics [55], machine learning [60], coding theory [24], and many other areas.

In this thesis the generally intractable non-convex problems we wish to solve are rank minimization problems of the form

$$\min \operatorname{rank}(X) \quad \text{s.t.} \quad X\in\mathcal{C}, \quad X\succeq 0$$

where $\mathcal{C}$ is some convex set and X is a symmetric matrix. A long-employed heuristic for such problems is to minimize the trace of X rather than the rank of X, yielding the convex optimization problem

$$\min \operatorname{tr}(X) \quad \text{s.t.} \quad X\in\mathcal{C}, \quad X\succeq 0.$$

In the factor analysis literature, the idea of minimizing trace as a surrogate for rank dates to Ledermann [37]. This heuristic has been put on firm theoretical ground more recently [43, 47]. One explanation for why it is reasonable to minimize trace as a surrogate for rank is that the trace function is the best pointwise convex approximation to the rank function when restricted to positive semidefinite matrices with spectral norm at most one [47].

2.2.1 Semidefinite Programming

When we employ the relaxation outlined in the previous section, all of the optimization problems we consider in this thesis can be cast as *semidefinite programs* (SDPs). (See [58] as a reference for the material in this section.) These are convex optimization problems of the form

$$\begin{aligned} \min_X\ & \langle C, X\rangle \\ \text{s.t. } & \mathcal{A}(X) = b \\ & X \succeq 0 \end{aligned} \tag{2.1}$$

where X, C and $n \times n$ symmetric matrices, $b \in \mathbb{R}^m$ and $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ is a linear map. The dual of this problem is

$$\begin{aligned} \max_y\ & \langle b, y\rangle \\ \text{s.t. } & C - \mathcal{A}^*(y) \succeq 0 \end{aligned} \tag{2.2}$$

where $\mathcal{A}^* : \mathbb{R}^m \to \mathcal{S}^n$ is the adjoint of $\mathcal{A}$. It is straightforward to verify that (2.1) and (2.2) satisfy weak duality, that is

$$\langle C, X\rangle \geq \langle b, y\rangle$$

whenever X is a feasible point for (2.1) and y is a feasible point for (2.2). In fact under quite mild conditions, strong duality holds. General sufficient conditions for strong duality in convex optimization can be specialized to this case to give the following result.

Theorem 1. *Suppose* (2.1) *and* (2.2) *are strictly feasible, that is there is an* $X \succ 0$ *such that* $\mathcal{A}(X) = b$ *and there is some* y *such that* $C - \mathcal{A}^*(y) \succ 0$*, then the optimal primal and dual costs coincide and are finite.*

The following summarizes the optimality conditions for semidefinite programming.

Theorem 2. *Under the conditions of Theorem 1,* $X^\star$ *is optimal for for* (2.1) *if and only if there exists some* $y^\star$ *such that*

1. $X^\star$ *is primal feasible*
2. $y^\star$ *is dual feasible*
3. $X^\star(C - \mathcal{A}^*(y^\star)) = 0$.

2.3 Concentration of Measure

In this thesis we focus on gaining qualitative understanding of the problems of interest in a high-dimensional setting. As such we are interested in understanding questions like, "For what problem parameters does a particular convex heuristic succeed on typical problem instances?". Questions of this type are often quite tractable because of the so-called 'blessing of dimensionality', that is the observation that in high dimensions, many quantities of interest typically behave in a very regular manner.

These notions are often formalized through the notion of concentration inequalities in probability theory. Suppose we are given a random variable X taking values in $\mathbb{R}^m$ (for concreteness) and a real valued function $f : \mathbb{R}^m \to \mathbb{R}$ on that space. By 'typical' behaviour of f, we mean the values it takes except on a set of rather small measure. To fix terminology we say that an event $\mathcal{E} \subset \mathbb{R}^m$ holds with *high probability* if there is a constant $\alpha > 0$ such that $\Pr[\mathcal{E}^c] = O(e^{-m^\alpha})$, and that $\mathcal{E}$ holds with *overwhelming probability* if we can take $\alpha \geq 1$. The informal notion of f typically being well behaved is formalized by bounding expressions of the form $\Pr[|f(X) - \mathbb{E}[f(X)]| \geq t]$ to show that f is very close to its mean with high or overwhelming probability.

There is a considerable literature on concentration inequalities. The book of Ledoux [39] covers many modern techniques for obtaining such inequalities. In this thesis the main concentration inequality we use is Borell's celebrated result on the concentration of Lipschitz functions of i.i.d. Gaussian random variables.

Theorem 3 (Borell [10]). *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz with respect to the Euclidean norm, that is $|f(x) - f(y)| \leq L\|x - y\|$ for all $x, y \in \mathbb{R}^n$. Then*

$$\Pr[f(X) \geq \mathbb{E}[f(X)] + t] \leq e^{-\frac{(t/L)^2}{2}}.$$

This tells us that any L-Lipschitz function of i.i.d. Gaussian random variables has tails like that of a scalar Gaussian with variance L. Examples of 1-Lipschitz functions that can be controlled in this way include:

- any ℓ_p norms with $p \geq 2$ of a standard Gaussian vector;
- any singular value of a random matrix with i.i.d. Gaussian entries;
- restricted singular values such as $\max_{x \in \mathcal{S}} \|Ax\|$ where $\mathcal{S}$ is a subset of the unit sphere and A is a matrix with i.i.d. Gaussian entries.

As an example of applying this concentration result, consider the problem of obtaining tail-bounds on the largest and smallest singular values of an $n \times k$ matrix ($n > k$) with i.i.d. Gaussian entries. As stated above, $f : \mathbb{R}^{n \times k} \to \mathbb{R}$ given by $f(V) = \sigma_{max}(V)$ is a 1-Lipschitz function of V, that is,

$$|\sigma_{max}(V) - \sigma_{max}(W)| \leq \|V - W\|_F.$$

Similarly $\sigma_{min}(V)$ is a 1-Lipschitz function of V. Hence Theorem 3 tells us that σ_{max} and σ_{min} concentrate around their respective means. The remaining challenge is to estimate the means

$\mathbb{E}[\sigma_{min}(V)]$ and $\mathbb{E}[\sigma_{max}(V)]$. For Gaussian matrices, techniques involving comparison inequalities [28] can be used to show that

$$\mathbb{E}[\sigma_{max}(V)] \leq \sqrt{n} + \sqrt{k} \quad \text{and} \quad \mathbb{E}[\sigma_{min}(V)] \geq \sqrt{n} - \sqrt{k}.$$

Combining Theorem 3 with these bounds gives sharp tail bounds for the largest and smallest singular values of rectangular Gaussian matrices.

Theorem 4. *If V is a $n \times k$ matrix (with $n \geq k$) with i.i.d. standard Gaussian entries then*

$$\Pr[\sqrt{n} - \sqrt{k} - t \leq \sigma_{min}(V) \leq \sigma_{max}(V) \leq \sqrt{n} + \sqrt{k} + t] \leq 2e^{-t^2/2}.$$

This example illustrates a common pattern of analysis to establish concentration inequalities for a function: first show that the function concentrates about some central value, and then estimate that central value using separate arguments. We repeatedly use the specific result in Theorem 4, as well as the pattern of controlling the mean and separately obtaining a concentration result, in our analysis in Chapter 3.

Chapter 3

Diagonal and Low-Rank Matrix Decompositions

3.1 Introduction

In this chapter we study a convex optimization-based heuristic for decomposing a matrix Σ, formed as the sum of a diagonal matrix $D^\star$ and a low-rank matrix $L^\star$, into these constituents. In particular, we focus on the case where $\Sigma, L^\star$ and $D^\star$ are all positive semidefinite, the situation that arises in factor analysis. We are particularly interested in this problem in high dimensional settings where the problem size n is large and the rank of the low-rank matrix is perhaps comparable to n.

To decompose Σ, the ideal problem we might wish to solve is known as the minimum rank factor analysis problem [51]

$$\begin{aligned} &\min \operatorname{rank}(L) \\ &\text{s.t. } \Sigma = D + L \\ &\quad L \succeq 0, D \succeq 0 \\ &\quad D \text{ diagonal.} \end{aligned} \tag{3.1}$$

This non-convex optimization is generally intractable. Our aim is to show that the heuristic of minimizing trace instead of rank very often allows us to successfully solve the minimum rank problem. This heuristic involves solving the following optimization problem, known as minimum trace factor analysis

$$\begin{aligned} &\min \operatorname{tr}(L) \\ &\text{s.t. } \Sigma = D + L \\ &\quad L \succeq 0, D \succeq 0 \\ &\quad D \text{ diagonal.} \end{aligned} \tag{3.2}$$

We are interested in two subtly different questions about minimum trace and minimum rank factor analysis.

1. If $\Sigma = L^\star + D^\star$ then when is $(L^\star, D^\star)$ the unique optimal point of minimum trace factor analysis?
2. When can we, in addition, assert that the optimal point of minimum trace factor analysis coincides with the optimal point of minimum rank factor analysis?

Our primary focus is on the first of these questions, although we briefly touch on the second when we discuss identifiability in Section 3.3. The relationship between minimum trace and minimum rank factor analysis has been studied before, most notably by Della Riccia and Shapiro [18]. However, many of the previous efforts related to understanding minimum rank and minimum trace factor analysis have focused on exact characterizations of the solutions of these problems in low dimensions. For example, in the system identification literature considerable effort has been put into the problem of determining the maximum corank in the Frisch scheme, that is characterizing the set of matrices Σ that can be expressed as the sum of a diagonal matrix and a positive semidefinite matrix of rank r. The case $r = n - 1$, resolved by Kalman [32], is the only case where a simple characterization is known.

We take a more coarse view in this chapter, focusing on the following two problems that, while quantitative in nature, also give a clear sense of how good minimum trace factor analysis is as a heuristic when applied to 'typical' problems.

1. Determine simple deterministic sufficient conditions on $L^\star$ that ensure minimum trace factor analysis correctly decomposes $\Sigma = L^\star + D^\star$.
2. Suppose $L^\star$ is chosen to have 'random' row/column space. For what values of r, the rank of $L^\star$, and n, the dimension of $L^\star$, does minimum trace factor analysis correctly decompose Σ with high probability?

(Here, and elsewhere in this chapter, by a random r-dimensional subspace of $\mathbb{R}^n$ we always mean a random subspace distributed according to the invariant (Haar) measure on the set of r dimensional subspaces of $\mathbb{R}^n$.)

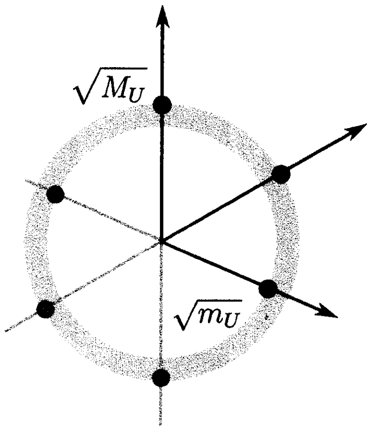

Figure 3-1: Consider the page as a two-dimensional subspace $U \subset \mathbb{R}^3$. Then $\sqrt{m_U}$ and $\sqrt{M_U}$ are the radii of the inner and outer circles shown, corresponding to the smallest and largest norms of projections of the standard basis vectors of $\mathbb{R}^3$ onto the page.

3.1.1 Incoherence parameters

The conditions we impose on $L^\star$ to ensure minimum trace factor analysis correctly decomposes $\Sigma = L^\star + \text{diag}^*(d^\star)$ are stated in terms of incoherence parameters. These parameters are (closely related to) parameters used in the matrix completion literature [13] which are, in turn, inspired by the work of Donoho and Huo [21]. These parameters depend only on the row/column space $U^\star$, of $L^\star$, and aim to capture the extent to which $U^\star$ aligns with the coordinate directions. As such for a subspace $U^\star$ of $\mathbb{R}^n$ we define

$$m_{U^\star} = \min_{i=1,\ldots,n} \|P_{U^\star} e_i\|^2 \quad \text{and} \quad M_{U^\star} = \max_{i=1,\ldots,n} \|P_{U^\star} e_i\|^2 \tag{3.3}$$

where e_i is the ith standard basis vector in $\mathbb{R}^n$. Note that if $U^\star$ has dimension r then $0 \leq m_{U^\star} \leq r/n \leq M_{U^\star} \leq 1$ (see [13]). These parameters are illustrated in Figure 3-1. Note that the closer $M_{U^\star}$ is to r/n, the less $U^\star$ lines up with any of the coordinate axes, and so the easier we expect it might be to distinguish between a matrix $L^\star$ with column space $U^\star$ and a diagonal matrix.

3.1.2 Main Results

We now state the main results established in this chapter. The first result confirms our intuition that as long as the column space of the low rank matrix is sufficiently incoherent with respect to the standard basis vectors, then minimum trace factor analysis succeeds.

Theorem 5. *Suppose $\Sigma = L^\star + D^\star$ where $D^\star$ is diagonal and positive semidefinite, $L^\star \succeq 0$, and the column space of $L^\star$ is $U^\star$. If $M_{U^\star} < 1/3$, minimum trace factor analysis has $(L^\star, D^\star)$ as its unique solution.*

In the setting where we solve a random instance of minimum trace factor analysis, we have the following result.

Theorem 6. *Suppose $\Sigma = L^\star + D^\star$ where $D^\star$ is diagonal and positive semidefinite, $L^\star \succeq 0$ and the column space of $L^\star$ is according to the Haar measure on the set of r-dimensional subspaces of $\mathbb{R}^n$. Then for any $\alpha \in (0, 1/6)$ there are positive constants $C, \tilde{c}, \bar{c}$, such if $r \leq n - Cn^{\frac{5}{6(1-\alpha)}}$ and n is sufficiently large then $(L^\star, D^\star)$ is the unique solution of minimum trace factor analysis with probability at least $1 - \bar{c}ne^{-\tilde{c}k^{3\alpha}}$.*

The key point of this theorem is that we can, for large n, decompose sums of diagonal and 'typical' matrices with ranks r that satisfy $r/n = 1 - o(1)$.

3.1.3 Related Results

The diagonal and low-rank decomposition problem can be viewed as a special case of two related matrix recovery problems in the recent literature. The first is the work of Chandrasekaran et al. [16] on decomposing matrices into sparse and low-rank components. The second is the work on low-rank matrix completion by convex optimization, initiated by Candes and Recht [13].

Comparison with sparse and low-rank decompositions Since a diagonal matrix is also sparse, we could try to solve the diagonal and low-rank decomposition problem using the convex program for sparse and low-rank decomposition outlined in [16]. In that work, deterministic conditions on the underlying sparse and low rank matrices under which the convex program successfully performs the desired decomposition are stated in terms of quantities μ and ξ. The parameter μ essentially measures the degree to which the non-zero entries in the sparse matrix are concentrated in a few rows and columns. For diagonal matrices $\mu = 1$. The parameter ξ essentially measures how much a low-rank matrix 'looks like' a sparse matrix. In fact the parameter ξ is very closely related to the incoherence parameters we use. Indeed if $U^\star$ is the column space of a symmetric matrix $L^\star$ then

$$M_{U^\star} \leq \xi^2 \leq 4M_{U^\star}.$$

The main deterministic result in the work of Chandrasekaran et al. is that given a matrix made up of a sparse and a low-rank matrix such that $\mu \cdot \xi < 1/6$ then their convex program exactly decomposes the matrix into its constituents. Since $\mu = 1$ in the diagonal and low rank case, this result specializes to $\xi < 1/6$ being sufficient for exact decomposition. In terms of incoherence parameters these results require $M_{U^\star} < 1/144$ for exact recovery, compared with the much milder condition of $M_{U^\star} < 1/3$ given in Theorem 5. This difference is hardly surprising as Chandrasekaran et al. deal with a much more general problem.

Relation to low-rank matrix completion We can also view the diagonal and low-rank decomposition problem as a low-rank matrix completion problem—given the off-diagonal entries of a low-rank matrix $L^\star$ we aim to fill in the diagonal entries correctly to obtain $L^\star$. In all of the recent work on low-rank matrix completion [13,29,34] the pattern of entries of the matrix that are revealed is random, whereas in our setting, the revealed entries consist of all of the off-diagonal entries. This makes it difficult to directly compare our results with exiting work on that problem. Furthermore this difference affects both the natural questions posed and answered, and the analysis. In the literature on low-rank matrix completion, one typically tries to determine, for a fixed rank, the smallest number of randomly sampled entries that are required to complete matrices of that rank correctly. On the other hand, we fix the pattern of known entries and essentially aim to determine the largest rank of a 'typical' matrix that can be completed correctly given these entries. We note, also, that having a random sampling procedure makes some aspects of the analysis of the associated convex program easier.

3.1.4 Outline of the Chapter

In Section 3.2 we consider some of the implications of our new results on minimum trace factor analysis for the direction of arrival estimation problem introduced in Section 1.2. We then proceed, in Section 3.3 to describe identifiability issues related to factor analysis, including presenting a simple new sufficient condition for local identifiability in terms of incoherence parameters. In Section 3.4 we discuss the key properties of minimum trace factor analysis. In Section 3.5 we provide a geometric interpretation of the optimality conditions for minimum trace factor analysis in terms of the problem of fitting an ellipsoid to a collection of points. The proofs of our main results on diagonal and low-rank decompositions then arise as corollaries of ellipsoid fitting results, the proofs of which are in Sections 3.5 and 3.9.

3.2 Some Implications for DOA Estimation

Let us revisit the direction of arrival estimation problem discussed in Section 1.2 of the introduction to see what our main deterministic result (Theorem 5) implies in this setting. Our discussion is at a fairly high level, with the aim to establish some intuition rather than a catalog of results.

Recall from (1.4) that the subspaces we are trying to identify in the direction of arrival problem are the column spaces of $N \times K$ matrices A with entries of the form

$$A_{i\ell} = e^{-j\langle x_i, k_\ell \rangle}$$

where x_i $(i = 1, 2, \ldots, N)$ denotes the location of the ith sensor and k_ℓ $(\ell = 1, 2, \ldots, K)$ the wave vector of the ℓth source. The first point to note is that in the direction of arrival problem, we are interested in identifying a subspace of $\mathbb{C}^N$, rather than a subspace of $\mathbb{R}^N$. This is not a problem as Theorem 5 and its proof are unchanged when we switch from real to complex scalars. The basic reason for this is that the optimality conditions for complex semidefinite programming have the same essential structure as those for real semidefinite programming.

Since Theorem 5 is stated in terms of the incoherence parameter M_U of the subspace $U = \mathcal{R}(A)$, to understand the implications of our analysis for the DOA problem, we need a way to translate between the geometry of the DOA problem and incoherence parameters. A first approach to this is to relate incoherence to the conditioning of the matrix $\frac{1}{N}A^H A$. Indeed suppose that $\frac{1}{N}A^H A$ has smallest and largest eigenvalues given by λ_{min} and λ_{max} respectively. Then

$$[P_U]_{ii} = \left[\frac{1}{N}A\left(\frac{1}{N}A^H A\right)^{-1} A^H\right]_{ii}$$

and so since

$$\frac{1}{\lambda_{max}}I \preceq \left(\frac{1}{N}A^H A\right)^{-1} \preceq \frac{1}{\lambda_{min}}I$$

and the rows of A each have squared Euclidean norm K it follows that

$$\frac{K}{N\lambda_{max}} \leq [P_U]_{ii} \leq \frac{K}{N\lambda_{min}}$$

for $i = 1, 2, \ldots, N$. Hence

$$\frac{K}{N\lambda_{max}} \leq m_U \leq \frac{K}{N} \leq M_U \leq \frac{K}{N\lambda_{min}}$$

giving us a method to translate between incoherence parameters and the largest and smallest eigenvalues of $A^H A$. In this setting Theorem 5 implies the following Corollary.

Corollary 1. *If λ_{min} is the smallest eigenvalue of $\frac{1}{N}A^H A$ and $\lambda_{min} > \frac{3K}{N}$ then minimum trace factor analysis correctly decomposes Σ into the noise and signal contributions.*

Let us now apply Corollary 1 to two different scenarios.

One source case Suppose there is only one source, that is $K = 1$. Then $\frac{1}{N}A^H A = 1$ and so $\lambda_{min} = 1$. Then as long as $N > 3$ we have $\lambda_{min} > 3K/N$ and so minimum trace factor analysis correctly decomposes Σ.

Random sensor case Consider, now, a situation where we have a fixed number of sources K, and a potentially large number of randomly located sensors. More precisely let $X_1, \ldots, X_N$ be independent and identically distributed with some distribution taking values in $\mathbb{R}^2$. (For concreteness, one could think of the X_i as being uniformly distributed on the unit square in $\mathbb{R}^2$.)

For a fixed set of sources $k_1, \dots, k_K \in \mathbb{R}^3$ consider the random vectors

$$Y_i = \begin{bmatrix} e^{j\langle X_i, k_1 \rangle} \\ \cdots \\ e^{j\langle X_i, k_K \rangle} \end{bmatrix} \quad \text{for } i = 1, 2, \dots N.$$

Note that the Y_i are independent and identically distributed. Furthermore, the Y_i are bounded as $\|Y_i\|^2 = K$ for all i. In general the (common) covariance of the Y_i has the form

$$\mathbb{E}[Y_i Y_i^H]_{a,b} = \varphi_{X_i}(P_{sensor}(k_a - k_b)) \tag{3.4}$$

where φ_{X_i} is the (common) characteristic function of the X_i and P_{sensor} is the Euclidean projection onto the plane in which the sensors lie. We expect that if the projection of the source locations onto the sensor plane are well-separated, then $\mathbb{E}[Y_i Y_i^H]$ will be positive definite, with smallest eigenvalue depending on the geometry of the source locations, and the distribution of the sensor locations.

Now notice that

$$\frac{1}{N} A^H A = \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^H.$$

So $\frac{1}{N} A^H A$ is the sum of independent identically distributed rank one positive semidefinite matrices. The following non-commutative Chernoff bound for bounded rank-one positive semidefinite matrices due to Oliveira [44] allows us to relate the smallest eigenvalue of $\frac{1}{N} A^H A = \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^H$ to that of $\mathbb{E}[Y_1 Y_1^H]$.

Theorem 7. *Suppose $Y_1, Y_2, \dots, Y_N$ are i.i.d. column vectors in $\mathbb{C}^K$ with $\|Y_1\|^2 \leq K$ almost surely. Then for any $t > 0$,*

$$\Pr\left[\left\| \frac{1}{N} \sum_{i=1}^{N} Y_i Y_i^H - \mathbb{E}[Y_1 Y_1^H] \right\| > t \|\mathbb{E}[Y_1 Y_1^H]\| \right] \leq (2N)^2 e^{-\frac{Nt^2}{8K(2+t)}}.$$

We specialize this result to the situation we are interested in.

Corollary 2. *With probability at least $1 - 1/N$*

$$\lambda_{min}(\frac{1}{N} A^H A) \geq \lambda_{\min}(\mathbb{E}[Y_1 Y_1^H]) - O\left(\sqrt{\frac{\log(N)}{N}}\right).$$

Note that $\lambda_{min}(\mathbb{E}[Y_1 Y_1^H])$ depends only on the source geometry and the sensor distribution, but is independent of N. So if $\mathbb{E}[Y_1 Y_1^H]$ is non-singular, we can conclude from Corollary 2 that for large enough N

$$\lambda_{min}(\frac{1}{N}A^HA) > \frac{3K}{N}$$

with high probability. Hence it follows from Corollary 1 that minimum trace factor analysis identifies the signal subspace for large enough N with high probability.

The settings considered here are a simple illustration of the sort of non-trivial implications Theorem 5 has in the particular example of the direction of arrival estimation problem.

3.3 Identifiability

Let us begin by focusing on the minimum rank factor analysis problem (3.1). The factor analysis model is, in general, not identifiable, in the sense that given $\Sigma = L^\star + \text{diag}^*(d^\star)$ where $L^\star$ has low rank, we cannot always hope to unambiguously identify $L^\star$ and $d^\star$. The identifiability of factor analysis has been studied by a number of authors [2,3,7,52] in work dating back to the 1940s. In this section we review some key results from that body of work. We also provide a new, simple, sufficient condition for identifiability in terms of the incoherence parameters (3.3) introduced in Section 3.1.

Definition 1. Minimum rank factor analysis is *globally identifiable* at $(d^\star, L^\star)$ if $(d^\star, L^\star)$ is the unique solution of (3.1). Minimum rank factor analysis is *locally identifiable* at $(d^\star, L^\star)$ if $(d^\star, L^\star)$ is the unique solution of (3.1) when restricted to some neighborhood of $(d^\star, L^\star)$.

3.3.1 Local identifiability

The following algebraic characterization of local identifiability for factor analysis appears in Theorem 3.1 of [52].

Theorem 8. *Factor analysis is locally identifiable at $(d^\star, L^\star)$ if and only if the Hadamard product $P_{(U^\star)^\perp} \circ P_{(U^\star)^\perp}$ is invertible, where $U^\star$ is the column space of $L^\star$.*

Other, more geometric characterizations along the lines of those considered by Chandrasekaran et al. [16] can also be given for this problem, and can be shown to be equivalent to the characterization in Theorem 8.

We now establish a new local identifiability result that gives a simple sufficient condition for local identifiability in terms of incoherence parameters.

Proposition 1. *Factor analysis if locally identifiable at $(d^\star, L^\star)$ if $M_{U^\star} < 1/2$ where $U^\star = \mathcal{R}(L^\star)$.*

Proof. Note that since $P_{U^\star} \circ P_{U^\star} \succeq 0$ it follows that

$$P_{(U^\star)^\perp} \circ P_{(U^\star)^\perp} = (I - P_{U^\star}) \circ (I - P_{U^\star}) = I - 2I \circ P_{U^\star} + P_{U^\star} \circ P_{U^\star} \succeq I - 2I \circ P_{U^\star}. \qquad (3.5)$$

So if $[P_{U^\star}]_{ii} < 1/2$ for all $i = 1, 2, \ldots, n$ (that is $M_{U^\star} < 1/2$) then $I - 2I \circ P_{U^\star} \succ 0$. Then (3.5) implies that $P_{(U^\star)^\perp} \circ P_{(U^\star)^\perp} \succ 0$ and so is invertible, as we set out to establish. □

Since we consider random instances of factor analysis, we are interested in generic identifiability properties (where we say that a property holds generically if it holds except on a set of Lesbegue measure zero). Factor analysis is generically locally identifiable as long as $n \leq \binom{n-r+1}{2}$ [51]. This bound is known in the factor analysis literature as the Ledermann bound [38].

Remark. Throughout this chapter we focus on generically identifiable problem instances, so we always assume that the bound $n \leq \binom{n-r+1}{2}$ holds.

3.3.2 Global identifiability

One approach to solving the minimum rank problem (3.1) is somehow to produce a feasible decomposition $\Sigma = \hat{L} + \text{diag}^*(\hat{d})$ and then *certify* that the pair $(\hat{L}, \hat{d})$ is a solution of the minimum rank problem (3.1), by certifying that the minimum rank problem is globally identifiable at $(\hat{L}, \hat{d})$.

From a computational point of view we would like such global identifiability conditions to be 'small'—polynomial in the problem size. The following result of Anderson and Rubin gives an explicit sufficient condition for global identifiability of factor analysis.

Theorem 9 (Anderson and Rubin [3]). *Factor analysis is globally identifiable at $(d^\star, L^\star)$ if*

- *$\text{rank}(L^\star) < n/2$ and*
- *if R is an $n \times r$ matrix such that $L^\star = RR^T$ then if we delete any row of R, there remain two distinct $r \times r$ submatrices with full rank.*

Furthermore this condition is necessary and sufficient in the cases $\text{rank}(L^\star) = 1$ and $\text{rank}(L^\star) = 2$.

Note that the condition in Theorem 9 requires us to exhibit only $2n$, $r \times r$ minors of R that are non-zero to certify global identifiability of factor analysis. Weaker sufficient conditions for global identifiability of factor analysis are known (for example, see Theorem 5 of [18]), but these tend to involve certificates that are much larger, such as the non-vanishing of *all* $r \times r$ minors of R.

The following generic global identifiability result essentially tells us that, except for the border case where $n = \binom{n-r+1}{2}$, any generically locally identifiable instance of factor analysis is also generically globally identifiable.

Theorem 10 (Bekker, ten Berge [7]). *If $n < \binom{n-r+1}{2}$ then factor analysis is generically globally identifiable at $(d^\star, L^\star)$ as long as $d^\star \geq 0$ and $\text{rank}(L^\star) = r$.*

As such, whenever we consider random instances of factor analysis problems, if we can construct a decomposition of $\Sigma = L^\star + \text{diag}^*(d^\star)$ into $\hat{L} + \text{diag}^*(\hat{d})$ with $\text{rank}(\hat{L}) = r$ satisfying $n < \binom{n-r+1}{2}$

then with probability 1 (over the randomness in the problem instance) we have that $(\hat{L},\hat{d}) = (L^\star, d^\star)$. Hence under the assumptions of our main randomized result, Theorem 6, we can, in addition, conclude that the optima of minimum trace and minimum rank factor analysis coincide with high probability.

3.4 Minimum Trace Factor Analysis

We now turn our attention to minimum trace factor analysis (3.2), a convex relaxation of the non-convex minimum rank factor analysis problem (3.1). Throughout this section we always assume that $\Sigma = L^\star + \operatorname{diag}^*(d^\star)$ for some $n \times n$ matrix $L^\star \succeq 0$ and some $d^\star \geq 0$.

Let us first rewrite (3.2) in an equivalent form that is slightly easier to analyze.

$$\begin{aligned} \hat{d}(\Sigma) = \arg\max_{d}\ & \langle 1_n, d\rangle \\ \text{s.t. } & \Sigma - \operatorname{diag}^*(d) \succeq 0 \\ & d \geq 0 \end{aligned} \tag{3.6}$$

where here and throughout the remainder of this chapter, 1_n denotes the vector in $\mathbb{R}^n$ all entries of which are one. Note that (3.6) is a conic program in standard form (see Section 2.2) and so its dual is given by

$$\begin{aligned} \min_{Y}\ & \langle \Sigma, Y\rangle \\ \text{s.t. } & \operatorname{diag}(Y) \geq 1_n \\ & Y \succeq 0. \end{aligned} \tag{3.7}$$

As long as $\Sigma \succ 0$, strong duality holds for this pair of semidefinite programs. This is because the point $d = \Sigma - \lambda_{min}(\Sigma)I/2$ is strictly feasible for the primal problem, $Y = 2I$ is strictly feasible for the dual, and the dual cost is clearly bounded below by zero.

Our main interest is in understanding conditions on $L^\star$ and $d^\star$ such that minimum trace factor analysis can correctly decompose Σ into its constituents. For convenience we formalize this notion.

Definition 2. We say that minimum trace factor analysis *correctly decomposes* Σ if $\hat{d} = d^\star$ is the *unique* optimal point of the semidefinite program (3.6).

Applying the usual optimality conditions for semidefinite programming we obtain necessary and sufficient conditions under which minimum trace factor analysis correctly decomposes Σ. This result, albeit in a rather different form, was first established as Theorems 3 and 4 of [18].

Proposition 2. *Minimum trace factor analysis correctly decomposes Σ if and only if there exists $Y \succeq 0$ such that $Y_{ii} = 1$ if $d_i^\star > 0$, $Y_{ii} \geq 1$ if $d_i^\star = 0$ and $YL^\star = 0$.*

Remark. It is a straightforward application of the optimality conditions for semidefinite programming (Theorem 2 of Section 2.2) to see that the existence of a Y with the stated properties certifies that $d^\star$ is an optimal point. We focus, here, on showing that under these conditions $d^\star$ is the unique optimal point. We supply a slightly different proof to that in [18] because we want an argument that generalizes easily as we need a more general version of this result in Proposition 9 in Chapter 4.

Proof. Suppose d^1, d^2 are optimal points and let $L^1 = \Sigma - \text{diag}^*(d^1)$ and $L^2 = \Sigma - \text{diag}^*(d^2)$. Then by convexity $(d^1 + d^2)/2$ is also optimal. Hence there is some $Y \succeq 0$ such that $\text{diag}(Y) \geq 1_n$ and $Y(L^1 + L^2) = 0$. Since $Y \succeq 0$ and $L^1 + L^2 \succeq 0$ it follows that the column space of Y is contained in the intersection of the nullspaces of L_1 and L_2. Hence $Y(L^1 - L^2) = 0$ and so $Y(\text{diag}^*(d^2) - \text{diag}^*(d^1)) = 0$. Since all of the diagonal elements of Y are non-zero this implies that $d^1 - d^2 = 0$ as we require. □

Remark. We are usually interested in the case where $d^\star > 0$ (note that in a factor analysis setting $d_i = 0$ indicates no noise in a particular entry). In this case the conditions on Y in Proposition 2 become $Y \succeq 0$, $Y_{ii} = 1$, $YL^\star = 0$. In general requiring that $Y_{ii} = 1$ gives a sufficient condition for optimality of $(d^\star, L^\star)$. We use this sufficient condition in all of our analysis.

3.5 Ellipsoid Fitting

In this section we interpret the optimality conditions for minimum trace factor analysis in a geometric way in terms of the problem of finding an ellipsoid that interpolates a given set of points. We then proceed to describe the intuition behind, and high level strategy of, the proofs of the main technical results of this chapter. While in this section we state, and think of, these main results in terms of the ellipsoid fitting problem, we also show how to translate them into the form of the results stated in the introduction to this chapter. The details of some of the proofs in this section appear in the appendix to this chapter, Section 3.9.

3.5.1 Optimality Conditions and Ellipsoid Fitting

Recall from the optimality conditions for minimum trace factor analysis, Proposition 2, that to decompose the sum of a diagonal matrix and a $n \times n$ positive semidefinite matrix $L^\star$ we needed to be able to find an $n \times n$ matrix Y satisfying $Y \succeq 0$, $\text{diag}(Y) = 1_n$ and $YL^\star = 0$. Suppose $W = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$ is a $k \times n$ matrix (where k is the corank of $L^\star$) whose column space is the orthogonal complement of the column space of $L^\star$. Then any $Y \succeq 0$ satisfying $YL^\star$ can be expressed as $Y = WZW^T$ for some $k \times k$ positive semidefinite matrix Z. The additional condition that $\text{diag}(Y) = 1_n$ can be expressed in terms of Z as the condition $v_i^T Z v_i = 1$ for $i = 1, 2, \ldots, n$.

To summarize, the problem of finding an $n \times n$ matrix $Y \succeq 0$ such that $\text{diag}(Y) = 1_n$ and $YL^\star = 0$ is equivalent to the problem of finding a $k \times k$ positive semidefinite matrix Z such that $v_i^T Z v_i = 1$ for $i = 1, 2, \ldots, n$. Since any positive semidefinite matrix can be interpreted as an origin-symmetric ellipsoid, this problem is equivalent to the following ellipsoid fitting problem.

Problem 1 (Ellipsoid fitting). Given n points $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ find an origin symmetric ellipsoid that passes through $v_1, v_2, \ldots, v_n$.

We say that a collection of points $v_1, v_2, \ldots v_n \in \mathbb{R}^k$ has the *ellipsoid fitting property* if there is $Z \succeq 0$ such that $v_i^T Z v_i = 1$ for $i = 1, 2, \ldots, n$. Clearly this property is invariant under changing basis in $\mathbb{R}^k$. Hence if, as before, $W = \begin{bmatrix} v_1 & \cdots v_n \end{bmatrix}^T$ then the ellipsoid fitting property of the points $v_1, \ldots, v_n$ depends only on the column space of W, a subspace of $\mathbb{R}^n$. This observation allows us to give a more abstract definition of the ellipsoid fitting property that we will use throughout the remainder of this chapter.

Definition 3. A k-dimensional subspace $V \subset \mathbb{R}^n$ has the ellipsoid fitting property if the points $\pi_V e_i \in \mathbb{R}^k$ for $i = 1, 2, \ldots, n$ have the ellipsoid fitting property.

We consider two flavors of the ellipsoid fitting problem.

1. Determine simple deterministic sufficient conditions on a subspace V to ensure V has the ellipsoid fitting property.
2. For what values of k and n does a random k dimensional subspace of $\mathbb{R}^n$ have the ellipsoid fitting property with high probability?

Monotonicity It is useful to observe that the ellipsoid fitting property enjoys certain monotonicity relations. In particular, it is clear that if a subspace V has the ellipsoid fitting property then any subspace $\bar{V} \supset V$ also has the ellipsoid fitting property. This simple observation implies the following monotonicity property in the randomized setting.

Proposition 3. *If $k_1 \leq k_2 \leq n$ and V_1 is a random k_1 dimensional subspace of $\mathbb{R}^n$ and V_2 is a random k_2 dimensional subspace of $\mathbb{R}^n$ then*

$$\Pr[V_1 \textit{ has the ellipsoid fitting property}] \leq \Pr[V_2 \textit{ has the ellipsoid fitting property}].$$

We also note that there are two different, but equivalent, ways to think of the randomized ellipsoid fitting problem. If $v_1, \ldots, v_n \in \mathbb{R}^k$ are i.i.d. Gaussian vectors and $W = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^T$ then there is an ellipsoid passing through $v_1, \ldots, v_n$ if and only if the column space of W has the ellipsoid fitting property. But the column space of W is uniformly distributed on the set of k-dimensional subspaces of $\mathbb{R}^n$ by the rotation invariance of the Gaussian measure. As such if V is a random k-dimensional subspace of $\mathbb{R}^n$ and $v_1, \ldots, v_n \sim \mathcal{N}(0, I_k)$ are i.i.d. then

$$\Pr[v_1, \ldots, v_n \text{ have the ellipsoid fitting property}] = \Pr[V \text{ has the ellipsoid fitting property}].$$

So when we analyze the randomized ellipsoid fitting problem we can choose either to think in terms of a random subspace having the ellipsoid fitting property, or in terms of fitting an ellipsoid to a collection of i.i.d. Gaussian vectors.

3.5.2 Least Squares-based constructions

In this section we give a sufficient condition for a set of points $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ to have the ellipsoid fitting property. The sufficient condition is based on choosing a *quadratic form* on which all of the points $v_1, v_2, \ldots, v_n$ lie, and then checking that the chosen quadratic form is positive definite. Since the set of quadratic forms on which a given set of points lie is a subspace of the set of all quadratic forms, the first part of the construction is purely linear-algebraic in nature. Showing that this quadratic form is positive definite is the challenging part of the construction.

A similar construction has been used in the analysis of a number of convex optimization problems. The basic idea was at the heart of the proofs in Candes, Romberg, and Tao's work on 'compressive sensing' [14]. The construction has been adapted and extended to assist in the analysis of a number of other convex heuristics for non-convex problems such as low-rank matrix completion [13], and sparse and low-rank decompositions [12], referred to by the name of 'Robust PCA' in that work. We first introduce the least-squares sufficient condition in a rather algebraic way and then appeal to geometric intuition to understand when we might expect it to be useful.

Given points $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$ we define a map $\mathcal{A} : \mathcal{S}_+^k \to \mathbb{R}^n$ by

$$[\mathcal{A}(Z)]_i = v_i^T Z v_i.$$

Observe that the collection of points $v_1, v_2, \ldots, v_n$ has the ellipsoid fitting property if and only if $1_n \in \mathcal{A}(\mathcal{S}_+^k)$. The following proposition establishes a sufficient condition for ellipsoid fitting based on this observation.

Proposition 4. *If $\mathcal{A}^\dagger(1_n) \in \mathcal{S}_+^k$ then $v_1, v_2, \ldots, v_n$ have the ellipsoid fitting property.*

Proof. Since $\mathcal{A}\mathcal{A}^\dagger(1_n) = 1_n$ it follows that $\mathcal{A}^\dagger(1_n) \in \mathcal{S}_+^k$ implies $1_n = \mathcal{A}\mathcal{A}^\dagger(1_n) \in \mathcal{A}(\mathcal{S}_+^k)$. □

Our choice of basis for $\mathbb{R}^k$ affect the quadratic form $\mathcal{A}^\dagger(1_n)$ that we construct, and also affects whether or not $\mathcal{A}^\dagger(1_n) \succeq 0$.

Due to this non-uniqueness, to establish that a subspace V of $\mathbb{R}^n$ has the ellipsoid fitting property, we often define $\mathcal{A}$ in a canonical way with respect to V as

$$[\mathcal{A}(Z)]_i = e_i^T(\pi_V^T X \pi_V)e_i \tag{3.8}$$

and establish that $\mathcal{A}^\dagger(1_n) \succeq 0$.

While proving that $\mathcal{A}^\dagger(1_n) \succeq 0$ suffices to prove that the ellipsoid fitting property holds for the associated set of points, it is not obvious that this should be a good sufficient condition. By interpreting $\mathcal{A}^\dagger(1_n)$ geometrically, we can gain some understanding of when we expect this sufficient condition to work well.

First we note that for any fixed $c \in \mathbb{R}^n$, $\mathcal{A}^\dagger(1_n)$ is the optimal point of the following optimization problem

$$\begin{aligned} &\min_Z \|Z - \mathcal{A}^*(c)\|_F^2 \\ &\text{s.t. } \mathcal{A}(Z) = 1_n. \end{aligned}$$

Suppose there exists some $c \in \mathbb{R}^n$ such that we know, *a priori* that

- $\mathcal{A}^*(c) \succ 0$ and $\mathcal{A}^*(c)$ is far from the boundary of the positive semidefinite cone and
- $\mathcal{A}\mathcal{A}^*(c)$ is close to 1_n.

Then since $\mathcal{A}\mathcal{A}^*(c)$ is close to 1_n, we expect that $\mathcal{A}^\dagger(1_n) = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}(1_n)$ is close to $\mathcal{A}^*(c)$, and so might expect that $\mathcal{A}^\dagger(1_n)$ is also positive definite. The canonical choice of $\mathcal{A}$ given by (3.8) satisfies $\mathcal{A}^*((n/k)1_n) = (n/k)\pi_V\pi_V^T = (n/k)I$. With the choice $c = (n/k)1_n$ our construction chooses the quadratic form (scaled to that it has the correct trace) passing through the points $\pi_V e_i$ for $i = 1, 2, \ldots, n$ and also 'closest' in some sense to a sphere, an intuitively appealing construction.

The map $\mathcal{A}$ corresponding to a set of points $v_1, \ldots, v_n$ plays an important role in the following discussion. The next lemma summarizes some easily verifiable properties of $\mathcal{A}$ that we use in the sequel.

Lemma 1. *If $v_1, \ldots, v_n \in \mathbb{R}^k$ and W is the $n \times k$ matrix whose ith row is v_i then*

1. *$\mathcal{A}(Z) = diag(WZW^T)$ and $\mathcal{A}^*(x) = W^T diag^*(x)W$*
2. *$\mathcal{A}\mathcal{A}^* : \mathbb{R}^n \to \mathbb{R}^n$ has the matrix representation $WW^T \circ WW^T$.*

The canonical choice of $\mathcal{A}$ defined with respect to a k dimensional subspace V satisfies

1. *$\mathcal{A}\mathcal{A}^* = P_V \circ P_V$*
2. *$\mathcal{A}^*(1_n) = I$*

3. $[\mathcal{A}\mathcal{A}^*(1_n)]_i = \|\pi_V e_i\|^2$ *for* $i = 1, 2, \ldots, n$.

Remark. We note that in light of Proposition 8, $\mathcal{A}\mathcal{A}^*$ is invertible if and only if minimum rank factor analysis is locally identifiable.

3.6 Deterministic Ellipsoid-Fitting Results

In this section we state and prove our main deterministic results for ellipsoid fitting, indicating how the deterministic results for minimum trace factor analysis stated in the introduction of this chapter follow from these results. We begin, however, with the main previously known deterministic result for ellipsoid-fitting. The result, due to Delorme and Poljak [19] was established in the context of analyzing the properties of an approximation algorithm for the MAXCUT problem.

Theorem 11 (Delorme, Poljak [19]). *If* $u \in \mathbb{R}^n$ *then there is a matrix* $Y \succeq 0$ *with* $diag(Y) = 1_n$ *such that* $Yu = 0$ *if and only if* u *is* balanced, *that is*

$$|u_i| \leq \sum_{j \neq i} |u_j| \quad \textit{for all } i = 1, 2, \ldots, n.$$

The corresponding ellipsoid-fitting result is the following.

Theorem 12. *A collection of points* $v_1, \ldots, v_n \in \mathbb{R}^{n-1}$ *has the ellipsoid fitting property if and only if no* v_i *is in the interior of the convex hull of the* v_j *for* $j \neq i$.

Remark. It is clear that for an ellipsoid to pass through n points in $\mathbb{R}^k$, it is necessary that none of these points is in the interior of the convex hull of the points (otherwise we could not fit *any* convex body to the points). Delorme and Poljak's result can be interpreted as saying that this obvious necessary condition is also sufficient when $k = n - 1$.

As a corollary of Theorem 11 we obtain a result about minimum trace factor analysis that plays an important role in Chapter 4.

Corollary 3. *If* $\Sigma = diag(d^\star) + u^\star u^{\star T}$ *for some* $u^\star \in \mathbb{R}^n$ *then minimum trace factor analysis correctly decomposes* Σ *if and only if* $u^\star$ *satisfies*

$$|u_i^\star| \leq \sum_{j \neq i} |u_j^\star| \quad \textit{for all } i = 1, 2, \ldots, n.$$

This result, however, only applies in the case where the low rank matrix has rank one. No such characterizations are known for higher ranks, except in the case of $n = 4$ and $r = 2$ [4]. Our next result establishes a simple sufficient condition for a subspace to have the ellipsoid fitting property without explicit assumptions on the dimension of the subspace V.

Theorem 13. *If V is a subspace of $\mathbb{R}^n$ with $m_V > 2/3$ then V has the ellipsoid fitting property.*

Let us first explain why we might expect a result like this to hold. The condition $m_V > 2/3$ tells us that all of the points $\pi_V e_i$ to which we are trying to fit an ellipsoid lie outside a sphere of radius $\sqrt{2/3}$. Also, since the $\pi_V e_i$ are orthogonal projections of unit vectors, these points lie inside a sphere of radius 1. Hence the assumptions of Theorem 13 confine the points to which we are trying to fit an ellipsoid to lie in a spherical shell, which might be expected to make it 'easier' to fit an ellipsoid to those points. Of course, we cannot take *any* points in this spherical shell and hope to fit an ellipsoid to them, it is important that these points arise as projections of the standard basis vectors onto a subspace.

Before proving Theorem 13, we show how it implies our main deterministic result about minimum trace factor analysis.

Proof of Theorem 5. Observe that if V is the orthogonal complement of $U^\star$ then $M_{U^\star} = 1 - m_V$. Hence if $M_{U^\star} < 1/3$, the assumption in Theorem 5, then $m_V > 2/3$ and so V has the ellipsoid fitting property. Then by Proposition 2 it follows that minimum trace factor analysis correctly decomposes any Σ of the form $\operatorname{diag}^*(d^\star) + L^\star$ where $L^\star \succeq 0$ and the column space of $L^\star$ is $U^\star$. □

Proof of Theorem 13. We use the least squares approach outlined in Section 3.5.2 with the canonical choice of $\mathcal{A}$ corresponding to V, that is $[\mathcal{A}(X)]_i = e_i^T(\pi_V^T X \pi_V)e_i$. Instead of showing that $\mathcal{A}^\dagger(1_n) \succeq 0$ we establish the sufficient condition that $(\mathcal{A}\mathcal{A}^*)^{-1}(1_n) \geq 0$. This is sufficient because $\mathcal{A}^*$ maps the non-negative orthant into the positive semidefinite cone.

Recall from Lemma 1 that $\mathcal{A}\mathcal{A}^*$ can be represented as $P_V \circ P_v$ which can be decomposed as

$$P_V \circ P_V = (I - P_U) \circ (I - P_U) = I - 2I \circ P_U + P_U \circ P_U = (2I \circ P_V - I) + P_U \circ P_U = A + B$$

where U is the orthogonal complement of V and $A = (2I \circ P_V - I)$ and $B = P_U \circ P_U$. Note that if $m_V > 2/3$ then

$$A = 2I \circ P_V - I \succeq (2m_V - 1)I \succeq (1/3)I.$$

Hence A is diagonal, non-negative, and so has a non-negative inverse.

We now expand $(P_V \circ P_V)^{-1} = (A+B)^{-1}$ as a Neumann series. This is valid as, using properties of $P_U \circ P_U$ in Lemma 1, we see that

$$\|A^{-1}B\| \leq \|A^{-1}\| \|P_U \circ P_U\|_{\infty \to \infty} = \frac{1}{2m_V - 1} \max_i [P_U \circ P_U 1_n]_i = \frac{M_U}{2m_V - 1} = \frac{1 - m_V}{2m_V - 1}$$

which is strictly less than one as long as $m_V > 2/3$. Then

$$(P_V \circ P_V)^{-1}(1_n) = A^{-1}(1_n - BA^{-1}1_n) + A^{-1}BA^{-1}B\left[(P_V \circ P_V)^{-1}(1_n)\right] \tag{3.9}$$

so that

$$(P_V \circ P_V)^{-1}1_n = (I - A^{-1}BA^{-1}B)^{-1}\left[A^{-1}(1_n - BA^{-1}1_n)\right]$$
$$= \sum_{i=0}^{\infty}(A^{-1}BA^{-1}B)^i\left[A^{-1}(1_n - BA^{-1}1_n)\right].$$

As such, to show that $(P_V \circ P_V)^{-1}1_n \geq 0$ it suffices to show that

1. $A^{-1}BA^{-1}B$ preserves the non-negative orthant and
2. $A^{-1}(1_n - BA^{-1}1_n) \geq 0$.

The first of these properties holds because A^{-1} and B are entrywise non-negative matrices. The second property holds because

$$BA^{-1}1_n \leq P_U \circ P_U\left(\frac{1}{2m_V - 1}1_n\right) \leq \frac{M_U}{2m_V - 1}1_n = \frac{1 - m_V}{2m_V - 1}1_n < 1_n$$

since $m_V > 2/3$. □

Remark. The proof technique we use here could potentially be strengthened in a number of ways.

First, we could modify our decomposition of $P_V \circ P_V$ into any positive diagonal part A and non-negative remainder B, and carry out the same analysis. Slightly better results can be obtained this way, but none are as clean as the result stated here.

Second, we could try to work directly with the cone $\mathcal{K} = \{x \in \mathbb{R}^n : \mathcal{A}^*(x) \succeq 0\}$ instead of the non-negative orthant. The same idea could work in this setting. We would need to choose a decomposition of $P_V \circ P_V = A + B$ so that $A^{-1}BA^{-1}B$ preserves the cone $\mathcal{K}$ and so that $A^{-1}(1_n - BA^{-1}1_n) \in \mathcal{K}$.

Importantly, this method of proof generalizes nicely to other situations, such as the case of block diagonal and low-rank decompositions that we consider in Chapter 4.

3.7 Randomized Ellipsoid-Fitting Results

We now consider the randomized version of the ellipsoid-fitting problem introduced in Section 3.5. In particular, for given integers $k \leq n$ we seek lower bounds on the probability that there is an origin-symmetric ellipsoid passing through n standard Gaussian vectors in $\mathbb{R}^k$. First we establish that if $2/3 < c \leq 1$ and $k \geq cn$ then the deterministic conditions of Theorem 13 hold with high probability, and so under these conditions 'most' k-dimensional subspaces of $\mathbb{R}^n$ have the ellipsoid fitting property. We then present empirical evidence that subspaces of much smaller dimension also have the ellipsoid fitting property with high probability. We conclude the section with our main result—that there is a constant C such that if $k \geq Cn^{5/6-\epsilon}$ (for some small ϵ) then a random k-dimensional subspace of $\mathbb{R}^n$ has the ellipsoid fitting property with high probability.

3.7.1 Theorem 13 in the randomized setting

Theorem 13 specifies conditions under which a subspace of $\mathbb{R}^n$ has the ellipsoid fitting property. We now provide conditions on n and k such that a random k dimensional subspace of $\mathbb{R}^n$ satisfies the conditions of Theorem 13 with overwhelming probability.

Theorem 14. *Let $2/3 < c \leq 1$ be a constant. Then there are positive constants $\bar{c}$, $\tilde{c}$, and K (depending only on c) such that if V is a random k dimensional subspace of $\mathbb{R}^n$ with $k \geq \max\{K, cn\}$ then*

$$\Pr[V \textit{ has the ellipsoid fitting property}] \geq 1 - \bar{c}n^{1/2}e^{-\tilde{c}n}.$$

Proof. Since $m_V > 2/3$ implies that V has the ellipsoid fitting property, it suffices to show that $\|\pi_V e_i\|^2 \geq \frac{1}{2}(c + \frac{2}{3}) > 2/3$ for all i with overwhelming probability. The main observation we use is that if V is a random k dimensional subspace of $\mathbb{R}^n$ and x is any fixed vector with $\|x\| = 1$ then $\|\pi_V x\|^2 \sim \beta(k/2, (n-k)/2)$ where $\beta(p, q)$ denotes the beta distribution [26]. In the case where $k = cn$, using a tail bound for β random variables [26] we see that if $x \in \mathbb{R}^n$ is fixed and $0 < \epsilon < 1$ and $k > 12/\epsilon^2$ then

$$\Pr[\|\pi_V x\|^2 \leq (1-\epsilon)c] \leq \frac{2}{(\epsilon^2 \pi c(1-c))^{1/2}} n^{-1/2} e^{-\frac{\epsilon^2}{2}k}.$$

Taking $\epsilon = \frac{1}{2} - \frac{1}{3c} > 0$ and a union bound over n events, as long as $k > K = 12/\epsilon^2$

$$\begin{aligned} \Pr\left[m_U \leq 2/3\right] &\leq \Pr\left[\|\pi_V e_i\|^2 < (c + 2/3)/2 \text{ for some } i\right] \\ &\leq n \cdot \frac{2}{(\epsilon^2 \pi c(1-c))^{1/2}} n^{-1/2} e^{-\frac{\epsilon^2}{2}k} = \bar{c}n^{1/2}e^{-\tilde{c}n} \end{aligned}$$

for appropriate constants $\bar{c}$ and $\tilde{c}$.

Finally we note that by the monotonicity of the ellipsoid fitting property (Proposition 3), the result also holds for any $k \geq cn$. □

3.7.2 Numerical Results

We now investigate, numerically, for which pairs (n, k) a random k-dimensional subspace of $\mathbb{R}^n$ has the ellipsoid fitting property with high probability.

To test the ellipsoid fitting property we perform the following experiment. For each (n, k) with $50 \leq n \leq 550$ and $0 \leq k \leq n$ we sample 10 independent $k \times n$ matrices with i.i.d. $\mathcal{N}(0, 1)$ entries, thinking of the rows $v_1, \ldots, v_k$ as vectors in $\mathbb{R}^k$ to which we are trying to fit an ellipsoid. We test

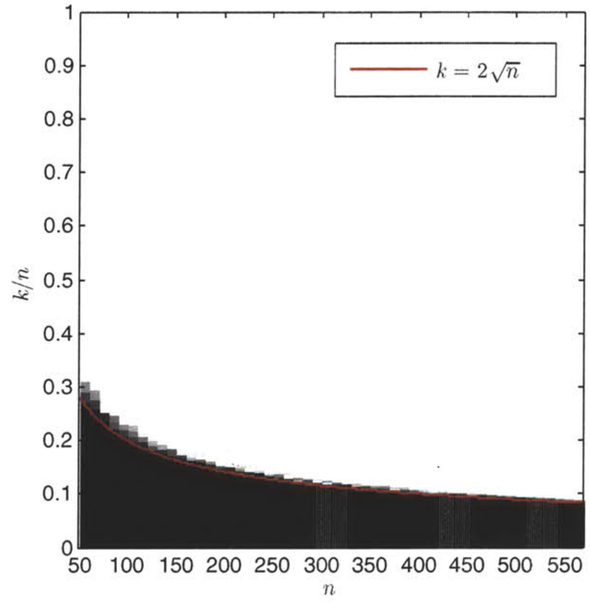

Figure 3-2: For each (n, k) with $50 \leq n \leq 550$ and $0 \leq k \leq n$ we repeat the following experiment 10 times. We sample a random k-dimensional subspace of $\mathbb{R}^n$ and check whether it has the ellipsoid fitting property. Cells are shaded according to the number of successful trials, with white corresponding to success on all trials, and black corresponding to failure on all trials. The red line is the line $k = 2\sqrt{n}$, our conjectured form for the threshold between success and failure of the convex program.

the ellipsoid fitting property in each case by checking whether there exists a positive semidefinite $Z \in \mathcal{S}^k$ such that $v_i^T Z v_i = 1$ for $i = 1, 2, \ldots, n$. We carry out this experiment using a combination of YALMIP [41] and SDPT3 [56]. The results are shown in Figure 3-2.

Recall from Theorem 14 that the deterministic conditions of Theorem 5 only hold with high probability if $k/n > 2/3$. It is evident from the results in Figure 3-2 that with overwhelming probability one can fit an ellipsoid to many more than $n = 3k/2$ points in $\mathbb{R}^k$ (i.e. the phase transition happens for much smaller values of k/n). This provides us with motivation to analyze the randomized ellipsoid fitting problem directly, as we do in Section 3.7.3. A comparison of the phase transition in Figure 3-2 with a plot of $k = 2\sqrt{n}$, suggests the following conjecture.

Conjecture 1. *If n and k are large and $n \leq k^2/4$ then we can fit an ellipsoid to n i.i.d. standard Gaussian vectors in $\mathbb{R}^k$ with overwhelming probability.*

3.7.3 Improved Randomized Results

In light of the numerical results presented in the previous section, we directly analyze the randomized ellipsoid fitting problem, with the aim of closing the gap between the results we obtain via the incoherence conditions in Theorem 14 and the numerical results. While our main result does not quite achieve the scaling suggested by the numerical results of Section 3.7.2, it is a significant qualitative improvement on the results of Theorem 14 in the randomized setting.

Theorem 15. *Fix $\alpha \in (0,1/6)$. There are absolute positive constants $C, \bar{c}, \tilde{c}$ such that for sufficiently large n and k, if $n \leq Ck^{6(1-\alpha)/5}$ and $v_1, v_2, \ldots, v_n \sim \mathcal{N}(0, I_k)$ are i.i.d. then*

$$\Pr[v_1, \ldots, v_n \textit{ have the ellipsoid fitting property}] \geq 1 - \bar{c}e^{-\tilde{c}k^{3\alpha}}.$$

Our main randomized result about minimum trace factor analysis (Theorem 6) follows directly from Theorem 15. We simply need to observe that the orthogonal complement of a random subspace is also a random subspace and so substitute $n - r$ for k in Theorem 15.

We only give a high level outline of the proof of Theorem 15 here and defer proofs of a number of technical lemmas to Section 3.9. The proof, like that of our incoherence based deterministic result—uses the least squares construction as explained in Section 3.5.2. In this case, rather than thinking in terms of a random subspace and using the canonical choice of $\mathcal{A}$, we think directly in terms of fitting a ellipsoid to standard Gaussian vectors $v_1, v_2, \ldots, v_n \in \mathbb{R}^k$. As such, throughout this section we take $\mathcal{A} : \mathcal{S}^k \to \mathbb{R}^n$ to be defined by $[\mathcal{A}(X)]_i = v_i^T X v_i$. We do so because in this case many random quantities of interest have a good deal of independence.

Proof. Recall that our overall aim is to show that $\mathcal{A}^\dagger(1_n) \succeq 0$.

Let $n = \delta k^{6(1-\alpha)/5}$ for some sufficiently small constant δ to be determined later. Since the probability that $v_1, \ldots, v_n$ have the ellipsoid fitting property is monotonically increasing with k, establishing the result for this choice of n implies it holds for all $n \leq \delta k^{6(1-\alpha)/5}$.

The key idea of our proof is that $\mathcal{A}\mathcal{A}^*$ is close to a deterministic matrix $M = (k^2 - k)I + kJ$ with high probability, and the matrix M has 1_n as an eigenvector. As such, we expand the term $(\mathcal{A}\mathcal{A}^*)^{-1}$ of $\mathcal{A}^\dagger = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}$ as a Neumann series around M. Defining $\Delta = M - \mathcal{A}^*\mathcal{A}$ and choosing $\eta > 0$

such that $M^{-1}(\eta 1_n) = 1_n$, as long as $\|M^{-1}\Delta\| < 1$ we can write

$$\mathcal{A}^{\dagger}(\eta 1_n) = \mathcal{A}^* \left[M^{-1}\eta 1_n + M^{-1}\Delta M^{-1}\eta 1_n + \sum_{i=1}^{\infty} (M^{-1}\Delta)^i (M^{-1}\Delta M^{-1}\eta 1_n) \right] \tag{3.10}$$

$$= \mathcal{A}^*(1_n) + \mathcal{A}^*(M^{-1}\Delta 1_n) + \mathcal{A}^* \left[\sum_{i=1}^{\infty} (M^{-1}\Delta)^i M^{-1}\Delta 1_n \right] \tag{3.11}$$

$$\succeq \left(\sigma_{min}(\mathcal{A}^*(1_n)) - \|\mathcal{A}^*(M^{-1}\Delta 1_n) + \mathcal{A}^* \left[\sum_{i=1}^{\infty} (M^{-1}\Delta)^i M^{-1}\Delta 1_n \right] \| \right) I. \tag{3.12}$$

So by the triangle inequality it suffices to show that

$$\sigma_{min}(\mathcal{A}^*(1_n)) \geq \|\mathcal{A}^*\|_{\infty\to\mathrm{sp}} \|M^{-1}\Delta 1_n\|_\infty + \|\mathcal{A}^*\|_{2\to\mathrm{sp}} \left(\sum_{i=1}^{\infty} \|M^{-1}\Delta\|^i \right) \|M^{-1}\Delta 1_n\|_2 \tag{3.13}$$

which certainly holds if $\|M^{-1}\|\|\Delta\| < 1$ and

$$\sigma_{min}(\mathcal{A}^*(1_n)) \geq \left(\|\mathcal{A}^*\|_{\infty\to\mathrm{sp}} + \sqrt{n}\|\mathcal{A}^*\|_{2\to\mathrm{sp}} \frac{\|M^{-1}\|\|\Delta\|}{1 - \|M^{-1}\|\|\Delta\|} \right) \|M^{-1}\|_{\infty\to\infty} \|\Delta 1_n\|_\infty. \tag{3.14}$$

Since $M = (k^2 - k)I + kJ$ it is straightforward to check that $M^{-1} = \frac{1}{k(k-1)}\left[I - \frac{1}{n+k-1}J\right]$ and so that $\|M^{-1}\| \leq \frac{2}{k^2}$ and $\|M^{-1}\|_{\infty\to\infty} \leq \frac{6}{k^2}$.

So we need only bound below $\sigma_{min}(\mathcal{A}^*(1_n))$ with high probability and bound above $\|\Delta 1_n\|_\infty$, $\|\mathcal{A}^*\|_{\infty\to\mathrm{sp}}$, $\|\mathcal{A}^*\|_{2\to\mathrm{sp}}$ and $\|\Delta\|$ with high probability. The following sequence of lemmas establish such bounds. The proofs are given in Sections 3.9.1, 3.9.2, 3.9.4, and 3.9.3 respectively.

Lemma 2. *If $k = o(n)$, there are positive constants $\tilde{c}_1$ and c_1 and c_1' such that with probability at least $1 - 2e^{-\tilde{c}_1 n}$,*

$$\sigma_{min}(\mathcal{A}^*(1_n)) \geq c_1 n \quad \text{and} \quad \|\mathcal{A}^*\|_{\infty\to sp} \leq c_1' n.$$

Lemma 3. *With probability at least $1 - 2e^{-\frac{1}{2}(k+\sqrt{n})}$, $\|\mathcal{A}^*\|_{2\to sp} \leq 8(k+\sqrt{n})$.*

Lemma 4. *If $k \geq \sqrt{n}$, there are positive constants $\tilde{c}_3, c_3$ and $\bar{c}_3$ such that with probability at least $1 - \bar{c}_3 n e^{-\tilde{c}_3\sqrt{n}}$*

$$\|\Delta\| \leq c_3 k n^{3/4}.$$

Lemma 5. *If $0 < \alpha < 1/6$ there are positive constants $\tilde{c}_4, c_4$ and $\bar{c}_4$ such that with probability at least $1 - \bar{c}_4 n e^{-\tilde{c}_4 k^{3\alpha}}$*

$$\|\Delta 1_n\|_\infty \leq c_4 n k^{1/2+3\alpha/2}.$$

We now prove the result assuming these bounds. We take a union bound over the complements of the events implicitly defined by Lemmas 2, 3, 4, and 5. Then since $n = \delta k^{6(1-\alpha)/5}$, for suitable constants $\bar{c}$ and $\tilde{c}$ and with probability at least

$$1 - (2e^{-\bar{c}_1 k} + 2e^{-\bar{c}_2(k+\sqrt{n})} + \bar{c}_3 n e^{-\bar{c}_3\sqrt{n}} + \bar{c}_4 n e^{-\bar{c}_4 k^{3\alpha}}) \geq 1 - \bar{c} n e^{-\tilde{c} k^{3\alpha}}$$

we have that $\sigma_{min}(\mathcal{A}^*(1_n)) \geq c_1 n$, $\|\mathcal{A}^*\|_{\infty\to\mathrm{sp}} = O(n)$, $\|\mathcal{A}^*\|_{2\to\mathrm{sp}} = O(k)$, $\|\Delta\| = O(n^{3/4}k)$, and $\|\Delta 1_n\|_\infty = O(nk^{(1+3\alpha)/2})$. As a result $\|M^{-1}\|\|\Delta\| = O(n^{3/4}k^{-1}) = o(1)$ under our assumption that $n = \delta k^{6(1-\alpha)/5}$, so the Neumann series converges and $\sum_{i=1}^{\infty} \|M^{-1}\|^i\|\Delta\|^i = O(n^{3/4}k^{-1})$. Combining these bounds with equation (3.14) we see that

$$\begin{aligned} T &:= \left(\|\mathcal{A}^*\|_{\infty\to\mathrm{sp}} + \sqrt{n}\|\mathcal{A}^*\|_{2\to\mathrm{sp}}\frac{\|M^{-1}\|\|\Delta\|}{1-\|M^{-1}\|\|\Delta\|}\right)\|M^{-1}\|_{\infty\to\infty}\|\Delta 1_n\|_\infty \\ &= O\left((n + \sqrt{n}\cdot k \cdot n^{3/4}k^{-1})k^{-2}\cdot nk^{(1+3\alpha)/2}\right) \\ &= O\left(n^{9/4}k^{-3(1-\alpha)/2}\right). \end{aligned}$$

So there exists a constant C such that for large enough n, $T \leq Cn^{9/4}k^{-3(1-\alpha)/2}$. Since $\sigma_{min}(\mathcal{A}^*(1_n)) \geq c_1 n$ and $n = \delta k^{6(1-\alpha)/5}$, by (3.14) we need to choose δ so that

$$\frac{T}{\sigma_{min}(\mathcal{A}^*(1_n))} \leq \frac{Cn^{9/4}k^{-3(1-\alpha)/2}}{c_1 n} = \frac{C}{c_1}n^{5/4}k^{-3(1-\alpha)/2} = \frac{C}{c_1}\delta^{5/4} \leq 1.$$

Clearly it suffices to choose $\delta \leq (c_1/C)^{4/5}$. With this choice, for sufficiently large n we have shown that $\mathcal{A}^\dagger(1_n) \succeq 0$, completing the proof. □

3.8 Discussion

In this chapter we have considered the problem of decomposing a positive semidefinite matrix Σ made up of the sum of a positive semidefinite low-rank matrix and a diagonal matrix into these constituents. In particular, we provided a new analysis of a semidefinite programming-based heuristic for the decomposition problem. Our analysis showed that under a natural incoherence condition on the column space of the low-rank matrix, the convex heuristic, known as minimum trace factor analysis, correctly decomposes Σ. We also analyzed the decomposition problem under the assumption that the column space of the low-rank matrix is random. We showed that minimum trace factor analysis can decompose Σ correctly with high probability if the column space of the low-rank matrix is random and has rank at most $n - cn^{5/6-\epsilon}$ for some constant c and some small ϵ.

Future research directions based on the work in this chapter are considered in Chapter 5.

3.9 Proofs for Chapter 3

We begin by collecting a number of standard estimates that we use for subsequent proofs. The first result is closely related to Lemma 36 of [59].

Lemma 6. *Let X be a real-valued random variable. Then for any $a \in \mathbb{R}$*

$$\Pr[|X^2 - a^2| \geq t] \leq 2\Pr[|X - a| \geq \min\{t/(3a), \sqrt{t/3}\}].$$

Proof. For any $a, b \in \mathbb{R}$, $|b^2 - a^2| \leq 3\max\{|b-a|^2, a|b-a|\}$. Hence

$$\begin{aligned}\Pr[|X^2 - a^2| \geq t] &\leq \Pr[\max\{|X-a|^2, a|X-a|\} \geq t/3] \\ &\overset{(a)}{\leq} \Pr[|X-a| \geq \sqrt{t/3}] + \Pr[|X-a| \geq t/(3a)] \\ &\leq 2\Pr[|X-a| \geq \min\{t/(3a), \sqrt{t/3}\}]\end{aligned}$$

where the inequality marked (a) is the result of taking a union bound. □

The following is a standard tail-bound for chi-squared random variables can be easily deduced from Proposition 16 of [59].

Lemma 7. *If $v \sim \mathcal{N}(0, I_k)$ is a k-dimensional standard Gaussian vector then*

$$\Pr\left[|\|v\|^2 - \mathbb{E}[\|v\|^2]| \geq t\right] \leq 2\exp\left(-\frac{1}{8}\min\left\{\frac{t^2}{k}, t\right\}\right)$$

Combining Lemma 6 and Lemma 7 yields the following tail bound that we use repeatedly.

Corollary 4. *If $v \sim \mathcal{N}(0, I_k)$ is a k dimensional standard Gaussian vector then*

$$\Pr[|\|v\|^4 - k^2| \geq t] \leq 4\exp\left(-\frac{1}{8}\min\left\{\frac{(t/3)^2}{k^3}, \sqrt{t/3}\right\}\right).$$

3.9.1 Proof of Lemma 2

Proof. Since $\mathcal{A}^*(1) = \sum_{i=1}^n v_i v_i^T$ the first statement is a standard result about Gaussian matrices (see Lemma 36 and Corollary 35 of [59]). For the second statement we observe that if $x, y \in \mathbb{R}^n$ and $x \geq y$ then $\mathcal{A}^*(x) \succeq \mathcal{A}^*(y)$. Hence for any $x \in \mathbb{R}^n$ with $\|x\|_\infty \leq 1$, $\mathcal{A}^*(-1_n) \preceq \mathcal{A}^*(x) \preceq \mathcal{A}^*(1_n)$. Hence

$$\lambda_{min}(\mathcal{A}^*(-1_n)) \leq \lambda_{min}(\mathcal{A}^*(x)) \leq \lambda_{max}(\mathcal{A}^*(x)) \leq \lambda_{max}(\mathcal{A}^*(1_n))$$

and so if $\|\mathcal{A}^*(1_n) - nI\| \leq c_1 k$,

$$\|\mathcal{A}^*\|_{\infty\to\text{sp}} = \sup_{\|x\|_\infty \leq 1} \|\mathcal{A}^*(x)\| \leq \max\{-\lambda_{min}(\mathcal{A}^*(-1_n)), \lambda_{max}(\mathcal{A}^*(1_n))\} \leq n + c_1 k.$$

□

3.9.2 Proof of Lemma 3

Lemma 3 follows from Proposition 5, a slightly more general tail bound, by putting $p = 2$ (and so $q = 2$) and noting that $c_2 = 1$.

Proposition 5. *Let $1 \leq p \leq \infty$ and let q be such that $p^{-1} + q^{-1} = 1$ (taking, as usual, $\infty^{-1} = 0$). Then there is an absolute constant $c > 0$ and a constant $c_q > 0$ depending only on q, such that*

$$\Pr[\|\mathcal{A}^*\|_{p\to sp} \geq 8(k + c_q^2 n^{1/q})] \leq 2\exp\left(-\frac{1}{2}(k + c_q^2 n^{1/q})\right).$$

Proof. Let $S^{k-1} = \{x \in \mathbb{R}^k : \|x\| = 1\}$ be the Euclidean unit sphere in $\mathbb{R}^k$. We first observe that

$$\|\mathcal{A}^*\|_{p\to\text{sp}} = \sup_{\substack{\|x\|_p \leq 1 \\ y \in S^{k-1}}} |\langle \mathcal{A}^*(x), yy^T\rangle| = \sup_{\substack{\|x\|_p \leq 1 \\ y \in S^{k-1}}} |\langle x, \mathcal{A}(yy^T)\rangle| \leq \sup_{y\in S^{k-1}} \|\mathcal{A}(yy^T)\|_q \tag{3.15}$$

by Hölder's inequality. Since $[\mathcal{A}(yy^T)]_i = \langle v_i, y\rangle^2$ it follows that

$$\sup_{y\in S^{k-1}} \|\mathcal{A}(yy^T)\|_q = \sup_{y\in S^{k-1}} \left(\sum_{i=1}^n \langle v_i, y\rangle^{2q}\right)^{1/q} = \sup_{y\in S^{k-1}} \|Vy\|_{2q}^2 = \|V\|_{2\to 2q}^2 \tag{3.16}$$

where V is an $n \times k$ matrix with standard Gaussian entries. So to establish the stated tail bound for $\|\mathcal{A}^*\|_{p\to\text{sp}}$ it suffices to establish the corresponding tail bound for $\|V\|_{2\to 2q}^2$.

Since $q \geq 1$, the map $V \mapsto \|V\|_{2\to 2q}$ is 1-Lipschitz with respect to the Euclidean norm on $\mathbb{R}^{n\times k}$. If we let $\mu = \mathbb{E}[\|V\|_{2\to 2q}]$ then by concentration of measure for Lipschitz functions of Gaussians (Theorem 3),

$$\Pr[\|V\|_{2\to 2q} \geq \mu + t] \leq e^{-\frac{t^2}{2}}$$

and so by Lemma 6

$$\Pr[\|V\|_{2\to 2q}^2 \geq \mu^2 + t] \leq 2\exp\left(-\frac{1}{2}\min\left\{t^2/(3\mu)^2, t/3\right\}\right). \tag{3.17}$$

It remains to compute an upper bound on $\mu = \mathbb{E}[\|V\|_{2\to 2q}]$. By a straightforward modification of Gordon's application of the Fernique-Sudakov comparison inequality to bounding the expectation of the largest singular value of an $n \times k$ Gaussian matrix [28], it can be shown that

$$\mu = \mathbb{E}[\|V\|_{2\to 2q}] \leq k^{1/2} + c_q n^{1/2q} \tag{3.18}$$

where the constant $c_q = \mathbb{E}[X^{2q}]^{1/2q}$ for $X \sim \mathcal{N}(0,1)$. Putting $t = 3\mu^2$ in (3.17) and using the inequality $a^2 + b^2 \leq (a+b)^2 \leq 2(a^2+b^2)$ for $a, b \geq 0$, gives

$$\Pr[\|V\|_{2\to 2q}^2 \geq 8(k + c_q^2 n^{1/q})] \leq \Pr[\|V\|_{2\to 2q}^2 \geq 4\mu^2] \leq 2\exp\left(-\frac{\mu^2}{2}\right) \leq 2\exp\left(-\frac{1}{2}(k + c_q^2 n^{1/q})\right).$$

□

3.9.3 Proof of Lemma 5

Lemma 8. *If* $0 < \alpha < 1/6$ *there are positive constants* $c_4, \bar{c}_4, \tilde{c}_4$ *such that*

$$\Pr[\|(W - M)1_n\|_\infty \geq c_4 n k^{1/2+3\alpha/2}] \leq \bar{c}_4 n e^{-\tilde{c}_4 k^{3\alpha}}.$$

for all $t \geq 0$.

Proof. Fix $1 \leq i \leq n$ and write $[M1_n]_i$ and $[W1_n]_i$ for the ith component of $M1_n$ and $W1_n$ respectively. Note that

$$|[M1_n]_i - [W1_n]_i| = |(\|v_i\|^4 - k^2) + \sum_{j\neq i}(\langle v_j, v_i\rangle^2 - k)| \tag{3.19}$$

$$\leq |\|v_i\|^4 - k^2| + \left|\sum_{j\neq i}(\langle v_j, v_i\rangle^2 - \|v_i\|^2)\right| + (n-1)|\|v_i\|^2 - k|. \tag{3.20}$$

Hence

$$\Pr[|[M1_n]_i - [W1_n]_i| \geq t] \leq \Pr[|\|v_i\|^4 - k^2| \geq t/3] + \Pr[(n-1)|\|v_i\|^2 - k| \geq t/3] + \Pr\left[\|v_i\|^2\left|\sum_{j\neq i}(\langle v_j, v_i/\|v_i\|\rangle^2 - 1)\right| \geq t/3\right] \tag{3.21}$$

The first term and second terms can be controlled by Corollary 4 and the chi-squared tail bound (Lemma 7) respectively, so we focus on the third term. Conditioned on v_i, if $j \neq i$ then $\langle v_j, v_i/\|v_i\|\rangle \sim \mathcal{N}(0,1)$ for $j \neq i$. Then let $Z = \sum_{j\neq i}\langle v_j, v_i/\|v_i\|\rangle^2$ and note that conditioned on v_i, Z is a chi-squared random variable with $n-1$ degrees of freedom. Let $\mathcal{E}$ be the event

$$\mathcal{E} = \{v_i : (n-1)|\|v_i\|^2 - k| \leq t/3 \quad \text{and} \quad |\|v_i\|^4 - k^2| \leq t/3\}.$$

Then if $\mathbb{I}[v \in \mathcal{E}]$ denotes the function that takes value 1 if $v \in \mathcal{E}$ and 0 otherwise,

$$\begin{aligned}
\Pr\left[|Z-\mathbb{E}[Z]| \geq \frac{t}{3\|v_i\|^2}\right] &= \mathbb{E}_{v_i}\left[\Pr\left[|Z-\mathbb{E}[Z]| \geq \frac{t}{3\|v_i\|^2}\middle| v_i\right]\right] \\
&\leq \mathbb{E}_{v_i}\left[2e^{-\frac{1}{8}\min\left\{\frac{(t/3)^2}{\|v_i\|^4(n-1)}, \frac{t}{3\|v_i\|^2}\right\}}\right] \\
&\leq \mathbb{E}_{v_i}\left[\mathbb{I}[v_i \in \mathcal{E}] \cdot 2e^{-\frac{1}{8}\min\left\{\frac{(t/3)^2}{\|v_i\|^4(n-1)}, \frac{t}{3\|v_i\|^2}\right\}}\right] + \mathbb{E}_{v_i}[\mathbb{I}[v_i \notin \mathcal{E}]] \\
&\leq 2e^{-\frac{1}{8}\min\left\{\frac{(t/3)^2}{(k^2+t/3)(n-1)}, \frac{t(n-1)}{3(k(n-1)+t/3)}\right\}} + \\
&\qquad \Pr[|\|v_i\|^4 - k^2| \geq t/3] + \Pr[|(n-1)\|v_i\|^2 - k| \geq t/3] \qquad (3.22)
\end{aligned}$$

where the last inequality holds by the definition of $\mathcal{E}$ and a union bound. Combining (3.21) and (3.22) we have that

$$\begin{aligned}
\Pr\left[|[M1_n]_i - [W1_n]_i| \geq t\right] \leq 2\Pr\left[|\|v_i\|^4 - k^2| \geq t/3\right] + 2\Pr\left[(n-1)|\|v_i\|^2 - k| \geq t/3\right] + \\
2\exp\left(-\frac{1}{8}\min\left\{\frac{(t/3)^2}{(k^2+t/3)(n-1)}, \frac{t(n-1)}{3(k(n-1)+t/3)}\right\}\right). \quad (3.23)
\end{aligned}$$

Putting $t = nk^{1/2+3\alpha/2}$ and using Lemma 7 and Corollary 4 to bound the first two terms gives

$$\begin{aligned}
\Pr[|[M1_n]_i - [W1_n]_i| \geq 3nk^{1/2+3\alpha/2}] \leq 4\exp\left(-\frac{1}{8}\min\left\{(n/3)^2k^{3\alpha}, \sqrt{n/3}k^{1/4+3\alpha/4}\right\}\right) + \\
4\exp\left(-\frac{1}{8}\min\left\{(n/(n-1))^2k^{3\alpha}, (n/(n-1))k^{1/2+3\alpha/2}\right\}\right) + \\
2\exp\left(-\frac{1}{8}\min\left\{\frac{(nk^{1/2+3\alpha/2})^2}{(k^2+nk^{1/2+3\alpha/2})(n-1)}, \frac{n(n-1)k^{1/2+3\alpha/2}}{3(k(n-1)+nk^{1/2+3\alpha/2})}\right\}\right).
\end{aligned}$$

Since $0 < \alpha < 1/6$, the right hand side is bounded by $\bar{c}_4e^{-\tilde{c}_4k^{3\alpha}}$ for suitable constants $\bar{c}_4$ and $\tilde{c}_4$. Then, taking a union bound we have that

$$\Pr[\|\Delta 1_n\|_\infty \geq 3nk^{1/2+3\alpha/2}] \leq n\Pr[|[M1_n]_i - [W1_n]_i| \geq 3nk^{1/2+3\alpha/2}] \leq n\bar{c}_4e^{-\tilde{c}_4k^{3\alpha}}.$$

□

3.9.4 Proof of Lemma 4

We now bound the spectral norm of $\Delta = M - \mathcal{A}\mathcal{A}^* = ((k^2-k)I - kJ) - VV^T \circ VV^T$ where V is a $n \times k$ matrix with $\mathcal{N}(0,1)$ entries. We show in Proposition 7 that the function $f : \mathbb{R}^{n\times k} \to \mathbb{R}$ defined by $f(V) = \|M - VV^T \circ VV^T\|$ concentrates around its mean. We then proceed to estimate the expected value of $f(V)$ by dealing with the diagonal and off-diagonal parts separately (in Lemmas 9 and 10).

Since f is not (globally) Lipschitz, we need a slight modification of the usual concentration of measure for functions of Gaussians to show that f concentrates around its expected value. It turns out that f is Lipschitz (with a small enough Lipschitz constant) on a set of large measure, so we first establish a simple variation on the usual concentration of measure for Lipschitz functions to this setting. Our result relies on a special case of Kirszbaum's theorem.

Theorem 16 (Kirszbaum). *Suppose $f : R^{m_1} \to \mathbb{R}^{m_2}$ is a function and $\mathcal{S} \subset \mathbb{R}^{m_1}$ is a subset such that $f|_{\mathcal{S}} : \mathcal{S} \to \mathbb{R}^{m_2}$ is L-Lipschitz with respect to the Euclidean metric. Then there is a function $\tilde{f} : \mathbb{R}^{m_1} \to \mathbb{R}^{m_2}$ that is L-Lipschitz with respect to the Euclidean metric such that $\tilde{f}(x) = f(x)$ for all $x \in \mathcal{S}$ and $\tilde{f}(\mathbb{R}^{m_1}) \subset conv(f(\mathcal{S}))$ (where $conv(A)$ is the convex hull of $A \subset \mathbb{R}^{m_2}$).*

Proposition 6. *Consider a non-negative real-valued function $f : \mathbb{R}^m \to \mathbb{R}$ and a subset $\mathcal{S} \subset \mathbb{R}^m$ such that $f|_{\mathcal{S}} : \mathcal{S} \to \mathbb{R}$ is L-Lipschitz with respect to the Euclidean norm and bounded by R. Then if $X \sim \mathcal{N}(0, I_m)$,*

$$\Pr\left[f(X) \geq \mathbb{E}[f(X)] + R\Pr[\mathcal{S}^c] + t\right] \leq e^{-\frac{1}{2}(t/L)^2} + \Pr[\mathcal{S}^c].$$

Proof. Since f is L-Lipschitz on $\mathcal{S}$ it follows from Kirszbaum's theorem (with $m_2 = 1$) that there is an L-Lipschitz function $\tilde{f} : \mathbb{R}^m \to \mathbb{R}$ such that $\tilde{f}(x) = f(x)$ for all $x \in \mathcal{S}$ and $|\tilde{f}(x)| \leq \sup_{y \in \mathcal{S}} |f(y)| \leq R$ for all $x \in \mathbb{R}^m$. Then by concentration of measure for Lipschitz functions of Gaussians (Theorem 3)

$$\Pr[\tilde{f}(X) \geq \mathbb{E}[\tilde{f}(X)] + t] \leq e^{-\frac{1}{2}(t/L)^2}. \tag{3.24}$$

Note that

$$\mathbb{E}[\tilde{f}(X)] = \mathbb{E}[f(X)\mathbb{I}[X \in \mathcal{S}]] + \mathbb{E}[\tilde{f}(X)\mathbb{I}[X \notin \mathcal{S}]] \leq \mathbb{E}[f(X)] + R\Pr[X \notin \mathcal{S}]$$

where the inequality is valid because f is a non-negative function. Then

$$\begin{aligned}\Pr[f(X) \geq \mathbb{E}[f(X)] + R\Pr[X \notin \mathcal{S}] + t] &\leq \Pr[f(X) \geq \mathbb{E}[\tilde{f}(X)] + t \text{ and } X \in \mathcal{S}] + \Pr[X \notin \mathcal{S}]\\ &\leq \Pr[\tilde{f}(X) \geq \mathbb{E}[\tilde{f}(X)] + t \text{ and } X \in \mathcal{S}] + \Pr[X \notin \mathcal{S}]\\ &\leq \Pr[\tilde{f}(X) \geq \mathbb{E}[\tilde{f}(X)] + t] + \Pr[X \notin \mathcal{S}]\end{aligned}$$

which, when combined with (3.24), yields the result. □

Proposition 7. *There is a universal constant $c > 0$ such that if $V \in \mathbb{R}^{n \times k}$ has i.i.d. standard*

Gaussian entries and M is any fixed $n \times n$ matrix with $\|M\| \leq 2nk$

$$\Pr[f(V) \geq \mathbb{E}[f(V)] + 36kn^2e^{-k/8} + t] \leq e^{-\frac{t^2}{32nk^2}} + 2ne^{-k/8}.$$

Proof. We first show that f is L-Lipschitz on a subset of $\mathbb{R}^{n\times k}$ of large measure and use Proposition 6 to complete the argument. Let V, W be two elements of $\mathbb{R}^{n\times k}$ with rows $v_1, \ldots v_n$ and $w_1, \ldots, w_n$ respectively. Let $\mathcal{S} \subset \mathbb{R}^{n\times k}$ be given by

$$\mathcal{S} = \{V : \|v_i\|^2 \leq 2k \text{ for } i = 1, 2, \ldots, n \text{ and } \|V\| \leq \sqrt{n} + 2\sqrt{k}\}.$$

Then if $V, W \in \mathcal{S}$,

$$\begin{aligned}
|f(V) - f(W)| &\leq \|VV^T \circ VV^T - WW \circ WW^T\| \\
&= \|(VV^T - WW^T) \circ (VV^T + WW^T)\| \\
&\overset{(a)}{\leq} \max_i(\|v_i\|^2 + \|w_i\|^2)\|VV^T - VW^T + VW^T - WW^T\| \\
&\leq \max_i(\|v_i\|^2 + \|w_i\|^2)(\|V\| + \|W\|)\|V - W\| \\
&\leq 8k(\sqrt{n} + 2\sqrt{k})\|V - W\|_F
\end{aligned}$$

where the inequality marked (a) follows from the fact that for positive semidefinite A and symmetric B, $\|A \circ B\| \leq (\max_i A_{ii})\|B\|$ (see Theorem 5.3.4 of [30]), and the final inequality invokes the definition of $\mathcal{S}$. It then follows that since $k \leq n$, f is $16k\sqrt{n}$-Lipschitz when restricted to $\mathcal{S}$. Furthermore, restricted to $\mathcal{S}$, f is bounded in the following way

$$f(V) \leq \|M\| + \|VV^T \circ VV^T\| \leq \|M\| + 2k(\sqrt{n} + 2\sqrt{k})^2 = 2nk + 4k(n + 4k) \leq 18nk$$

since we assume that $\|M\| \leq 2nk$. Furthermore

$$\Pr[\mathcal{S}^c] \leq \Pr[\max_i \|v_i\|^2 \leq 2k] + \Pr[\|V\| \geq \sqrt{n} + 2\sqrt{k}] \leq n\exp\left(-\frac{k}{8}\right) + \exp\left(-\frac{\sqrt{k}^2}{2}\right) \leq 2ne^{-k/8}$$

where the first term follows from Lemma 7 and a union bound, and the second follows from Theorem 4 in Section 2.3 on the spectral norm of matrices with i.i.d. Gaussian entries. Finally we apply Proposition 6 with $L = 16k\sqrt{n}$ and $R = 17nk$ to conclude that

$$\Pr[f(V) \geq \mathbb{E}[f(V)] + 36kn^2e^{-k/8} + t] \leq \Pr[f(V) \geq \mathbb{E}[f(V)] + \Pr[\mathcal{S}^c]R + t] \leq \exp\left(-c\frac{t^2}{nk^2}\right) + 2ne^{-k/8}$$

where $c = 2^9$, for example. □

Having established that f concentrates about its mean, we now need to bound $\mathbb{E}[f(V)]$. We do so by noting that

$$\mathbb{E}[f(V)] \leq \mathbb{E}[\|\mathrm{diag}(VV^T \circ VV^T) - k^2 1_n\|_\infty] + \mathbb{E}[\|\tilde{\Delta}\|] = \mathbb{E}[\max_i |\|v_i\|^4 - k^2|] + \mathbb{E}[\|\tilde{\Delta}\|]$$

where $\tilde{\Delta}$ is the off-diagonal part of Δ. Explicitly, the entries of $\tilde{\Delta}$ are given by $\tilde{\Delta}_{ii} = 0$ for $i = 1, 2, \ldots, n$ and $\tilde{\Delta}_{ij} = k - \langle v_i, v_j \rangle^2$ for $i \neq j$ (where the v_i are i.i.d. $\mathcal{N}(0, I_k)$ random vectors and are the rows of V).

Lemma 9. *If $k \geq 8\log(n)$ then for some positive constant c',*

$$\mathbb{E}[\max_{1\leq i\leq n} |\|v_i\|^4 - k^2|] \leq c' k^{3/2}\sqrt{\log(n)}.$$

Proof. The proof uses a standard technique, exemplified in [40] where it is used to bound the expectation of the maximum of finitely many Gaussian random variables. Let $\delta = 6k^{3/2}\sqrt{2\log(n)} \leq 3k^2$ and X be a scalar Gaussian random variable with zero mean and variance $36k^3$.

$$\begin{aligned}
\mathbb{E}[\max_i |\|v_i\|^4 - k^2|] &= \int_0^\infty \Pr[\max_i |\|v_i\|^4 - k^2| \geq t]\, dt \\
&\overset{(a)}{\leq} \int_0^\delta \Pr[\max_i |\|v_i\|^4 - k^2| \geq t]\, dt + n\int_\delta^\infty \Pr[|\|v_1\|^4 - k^2| \geq t]\, dt \\
&\overset{(b)}{\leq} \delta + n\int_\delta^{3k^2} 4e^{-\frac{t^2}{2(6k^{3/2})^2}}\, dt + n\int_{3k^2}^\infty e^{-\frac{1}{8}\sqrt{\frac{t}{3}}}\, dt \\
&\leq \delta + \bar{c}nk^{3/2}\Pr[X \geq \delta] + nc(k/8+1)e^{-k/8} \quad \text{(for some constants } c, \bar{c} \geq 0) \\
&\leq \delta + \bar{c}nk^{3/2}e^{-\frac{1}{2}\left(\frac{\delta}{6k^{3/2}}\right)^2} + nc(k/8+1)e^{-k/8} \\
&\leq 6k^{3/2}\sqrt{2\log(n)} + \bar{c}k^{3/2} + c(k/8+1) \quad (\text{since } k \geq 8\log(n))
\end{aligned}$$

which gives the desired result for a suitable choice of constant c'. Note that the inequality marked (a) holds by taking a union bound and the inequality marked (b) follows from the tail bound in Corollary 4. □

Lemma 10. *If $n \leq k^2$, there is a constant c such that*

$$\mathbb{E}[\|\tilde{\Delta}\|] \leq ckn^{3/4}.$$

Proof. Our proof follows rather closely the general strategy of the proof of Theorem 1 of [23], which deals in much more generality with the behaviour of random matrices of the form $X_{ij} = f(\langle v_i, v_j \rangle)$ for random vectors v_i. Since our assumptions are much stronger than the assumptions in that work, things will simplify considerably, and it will be fairly straightforward to perform explicit computations.

We proceed by using the moment method, that is by using the observation that for a symmetric random matrix X, $\mathbb{E}[\|X\|] \leq \mathbb{E}[\mathrm{tr}(X^{2p})]^{1/2p}$ for any positive integer p. In particular, we bound $\mathbb{E}[\mathrm{tr}(\tilde{\Delta}^4)]^{1/4}$. Computing higher moments gives slightly better estimates, but without a systematic way to compute these moments, the computations soon become rather unwieldy.

We make repeated use of the elementary inequalities $(a-b)^2 \leq a^2+b^2$ and $(a+b)^2 \leq 2(a^2+b^2)$ as well as the non-central moments of chi-squared random variables

$$\mathbb{E}[\|v_i\|^{2p}] = k(k+2)\cdots(k+2p-2).$$

Note, also, that if we condition on v_i, $\langle v_i, v_j\rangle \sim \mathcal{N}(0, \|v_i\|^2)$. Hence if $i \neq j$,

$$\begin{aligned}\mathbb{E}[\langle v_i, v_j\rangle^{2p}] &= \mathbb{E}[\mathbb{E}[\langle v_i, v_j\rangle^{2p}|v_i]] \\ &= (2p-1)(2p-3)\cdots(1)\mathbb{E}[\|v_i\|^{2p}] \\ &= (2p-1)(2p-3)\cdots(1)(k)(k+2)\cdots(k+2p-2) \qquad (3.25)\end{aligned}$$

where we have used the fact that if $X \sim \mathcal{N}(0,1)$ then $\mathbb{E}[X^{2p}] = (2p-1)(2p-3)\cdots(3)(1)$.

Note that $\tilde{\Delta}_{ii} = 0$ so the terms appearing in $\mathrm{tr}(\tilde{\Delta}^4)$ correspond to cycles of length four in the complete graph on n nodes. In particular, there are three different types of non-zero terms:

1. terms of the form $\tilde{\Delta}_{ij}^4$ (where $i \neq j$) of which there are fewer than n^2
2. terms of the form $\tilde{\Delta}_{ij}^2\tilde{\Delta}_{jk}^2$ (where $\neq j \neq k$) of which there are fewer than n^3 and
3. terms of the form $\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}\tilde{\Delta}_{k\ell}\tilde{\Delta}_{\ell i}$ (where $i \neq j \neq k \neq \ell$) of which there are fewer than n^4.

We analyze each of these terms separately, and then combine the results to give our estimate of $\mathbb{E}[\mathrm{tr}(\tilde{\Delta}^4)]$.

The case $i \neq j$

$$\mathbb{E}[\tilde{\Delta}_{ij}^4] = \mathbb{E}[(k - \langle v_i, v_j\rangle^2)^4] \leq k^4 + \mathbb{E}[\langle v_i, v_j\rangle^8] = O(k^4) \qquad (3.26)$$

The case $i \neq j \neq k$ The basic strategy is to note that conditioned on v_j, $\tilde{\Delta}_{ij}$ and $\tilde{\Delta}_{jk}$ are independent.

$$\begin{aligned}\mathbb{E}[\tilde{\Delta}_{ij}^2\tilde{\Delta}_{jk}^2] &= \mathbb{E}[(k - \langle v_i, v_j\rangle^2)^2(k - \langle v_j, v_k\rangle^2)^2] \\ &= \mathbb{E}[\mathbb{E}[(k - \langle v_i, v_j\rangle^2)^2|v_j]\mathbb{E}[(k - \langle v_j, v_k\rangle^2)^2|v_j]] \\ &= \mathbb{E}[\mathbb{E}[(k - \langle v_i, v_j\rangle^2)^2|v_j]^2] \\ &\leq \mathbb{E}[\mathbb{E}[k^2 + \langle v_i, v_j\rangle^4|v_j]^2] \\ &= \mathbb{E}[(k^2 + 3\|v_j\|^4)^2] = O(k^4).\end{aligned}$$

The case $i \neq j \neq k \neq \ell$ If we condition on v_i and v_k, $\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}$ and $\tilde{\Delta}_{k\ell}\tilde{\Delta}_{\ell i}$ are independent. Hence

$$\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}\tilde{\Delta}_{k\ell}\tilde{\Delta}_{\ell i}] = \mathbb{E}[\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}|v_i, v_k]\mathbb{E}[\tilde{\Delta}_{k\ell}\tilde{\Delta}_{\ell i}|v_i, v_k]] = \mathbb{E}[(\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}|v_i, v_k])^2].$$

As such we first compute $\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}|v_i, v_k]$.

$$\begin{aligned}\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}|v_i, v_k] &= \mathbb{E}[v_i^T v_j v_j^T v_i v_k^T v_j v_j^T v_k - k(\langle v_i, v_j\rangle^2 + \langle v_k, v_j\rangle^2) + k^2|v_i, v_k] \\ &= v_i^T\mathbb{E}[(v_j v_j^T)v_i v_k^T(v_j v_j^T)|v_i, v_k]v_k - k(\|v_i\|^2 + \|v_k\|^2) + k^2 \\ &\overset{(a)}{=} v_i^T(v_k v_i^T + v_i v_k^T + \langle v_i, v_k\rangle I_k)v_k - k(\|v_i\|^2 + \|v_k\|^2) + k^2 \\ &= 2\langle v_i, v_k\rangle^2 + (\|v_i\|^2 - k)(\|v_k\|^2 - k)\end{aligned}$$

where the equality marked (a) follows from a straightforward computation that can be found in Lemma A.1 of [23]. Then we see that

$$\begin{aligned}\mathbb{E}[\tilde{\Delta}_{ij}\tilde{\Delta}_{jk}\tilde{\Delta}_{k\ell}\tilde{\Delta}_{\ell i}] &= \mathbb{E}[(2\langle v_i, v_k\rangle^2 + (\|v_i\|^2 - k)(\|v_k\|^2 - k))^2] \\ &\leq 8\mathbb{E}[\langle v_i, v_k\rangle^4] + 2\mathbb{E}[(\|v_i\|^2 - k)^2]\mathbb{E}[(\|v_k\|^2 - k)^2] \\ &= 8\mathbb{E}[\langle v_i, v_k\rangle^4] + 2(\mathbb{E}[\|v_i\|^4] - k^2)^2 \\ &= 8\mathbb{E}[\langle v_i, v_k\rangle^4] + 2(2k)^2 = O(k^2).\end{aligned}$$

The proof of the lemma follows from combining these results and observing that $n \leq k^2$ to conclude that

$$\mathbb{E}[\mathrm{tr}(\tilde{\Delta}^4)]^{1/4} = O\left((n^2k^4 + n^3k^4 + n^4k^2)^{1/4}\right) = O(kn^{3/4}).$$

□

We now assemble these pieces to give our overall proof of Lemma 4. Combining Lemmas 9 and 10, and noting that $k^{3/2}\sqrt{\log(n)} = O(kn^{3/4})$ and, since $k \geq \sqrt{n}$, $kn^2e^{-k/8} = O(kn^{3/4})$, we have that $\mathbb{E}[f(V)] = O(kn^{3/4})$. Hence putting $t = O(kn^{3/4})$ in Proposition 7 we see that for some positive constants $c_3, \tilde{c}_3$

$$\Pr[f(V) \geq c_3 k n^{3/4}] \leq e^{-\sqrt{n}/32} + 2ne^{-k/8} \leq \bar{c}_3 n e^{-\tilde{c}_3 \sqrt{n}}$$

where we have again used the assumption that $k \geq \sqrt{n}$.

Chapter 4

Gaussian Latent Tree Modeling

4.1 Introduction

In this chapter we consider the problem of learning the parameters and state dimensions of a Gaussian latent tree model given the tree structure and the covariance matrix among the leaf variables. Our approach is based on the observation, described in Section 4.3, that the covariance among the leaf variables of such a model admits a decomposition as a sum of block diagonal low-rank positive semidefinite matrices with nested column spaces. In Section 4.4 we formulate an SDP to decompose a given covariance matrix into these constituents and in Section 4.4.1 give conditions on an underlying Gaussian latent tree model that ensures our SDP-based decomposition method succeeds. Once we have performed this decomposition we provide a method to construct an explicit parametrization of a Gaussian latent tree model.

In Section 4.5 we propose another convex program that approximately decomposes a covariance matrix in the required way. This can then be used for modeling purposes, where we would like to construct a parsimonious Gaussian latent tree model that fits a given covariance matrix well. We evaluate this approximate covariance decomposition convex program using synthetic experiments in Section 4.6, demonstrating that given only sufficiently many i.i.d. samples of the leaf variables of certain Gaussian latent tree models, the method can correctly estimate the state dimensions of the latent variables in the underlying model.

The problem of constructing a Gaussian latent tree model with a fixed index tree and covariance among the leaf variables that (approximately) realizes a given covariance has been considered by a number of authors. Irving et al. [31] developed a method for this problem based on the notion of canonical correlations, an approach initially developed for the corresponding realization problem for time-series by Akaike [1]. Frakt et al. [25] proposed a computationally efficient method for learning internal Gaussian latent tree models, an important subclass of these models. Both of these methods operate one vertex at a time, in a computationally greedy fashion, and require prior assumptions (such as hard upper bounds) on the dimensions of the state spaces at each vertex.

A standard approach to choosing parameters for any statistical model with latent variables is to use the expectation-maximization (EM) algorithm [20] which has been specialized to the case of learning parameters of Gaussian latent tree models [33]. The EM algorithm, however, does not offer any consistency guarantees, and does not (in its most basic form) learn the state dimensions of the latent variables along with the parameters.

Both the problem and the proposed solution methods in this chapter are a non-trivial generalization of the basic problem of factor analysis and the semidefinite programming-based method, minimum trace factor analysis, considered in Chapter 3. One of the key points of this chapter is that the analysis of our SDP-based method essentially reduces to the analysis of a number of instances of a problem that is a slight generalization of minimum trace factor analysis. As such some of the results of Chapter 3 play a role in the sequel.

Finally let us emphasize that we assume we are given a tree structure for the purpose of identifying model parameters and state dimensions. The problem of learning the tree structure from data is an interesting and challenging one that has received attention in the phylogenetics [22] and machine learning communities [45], for example. Many different techniques have been proposed for the problem of learning the tree structure. For a recent review and new techniques see [17]. We could use any of these techniques to come up with a tree structure for our formulation of the problem.

4.2 Preliminaries

We introduce notation and terminology related to trees and Gaussian tree models. In particular we discuss some particular parametrizations of Gaussian tree models that are convenient later in the chapter.

4.2.1 Trees

Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be a tree with a distinguished vertex $r \in \mathcal{V}$ called the *root*. We divide the vertices into *scales* depending on their distance from the root. Explicitly, $\mathcal{V}_i$ denotes the set of vertices at distance i from the root.

When it is convenient we can think of $\mathcal{T}$ as a directed tree with edges oriented away from the root. Given a vertex $v \in \mathcal{V}$ let $\mathcal{P}(v)$ be the parent of v, the (unique) vertex such that $(\mathcal{P}(v), v)$ is a directed edge in $\mathcal{T}$. Similarly the children of v, denoted $\mathcal{C}(v)$, are those vertices whose (common) parent is v. The *leaves* of the tree are those vertices with no children. The *descendants* of a vertex v are the vertices connected to v by a directed path. Generalizing the notation for children, if v is

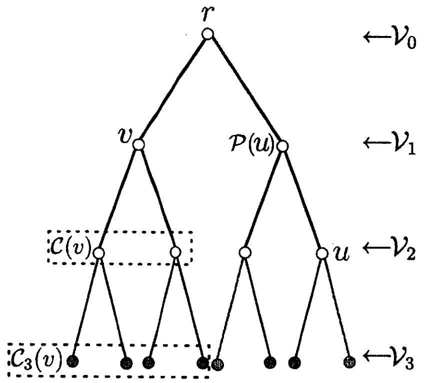
r
$\leftarrow \mathcal{V}_0$
v
$\mathcal{P}(u)$
$\leftarrow \mathcal{V}_1$
$\mathcal{C}(v)$
u
$\leftarrow \mathcal{V}_2$
$\mathcal{C}_3(v)$
$\leftarrow \mathcal{V}_3$

Figure 4-1: Summary of notation related to trees. Note that $\mathcal{V}_2$, for example, refers to all of the vertices at distance 2 from the root r. Note that all of the leaves of this tree are at the same distance from the root, and so this tree satisfies our standing assumption.

at scale $s(v)$ then for $n \geq s(v)$ we denote the descendants of v that are at scale n by $\mathcal{C}_n(v) \subset \mathcal{V}_n$. Finally we use the notation $\mathcal{V}_{\backslash r}$ instead of $\mathcal{V} \setminus \{r\}$ for the set of vertices excluding the root. We summarize some of these notational conventions in Figure 4-1.

We restrict ourselves to a particular class of trees in this chapter. We assume that trees are rooted and have *all of their leaves at the same scale* with respect to the root.

4.2.2 Model Parametrization

Throughout this chapter we always use 'directed' parametrizations of Gaussian latent tree models, thinking of such models as linear state space models indexed by trees and driven by white noise. It turns out that the exposition is cleaner and more intuitive from this point of view. As we deal only with distributions that are Markov with respect to trees, there is no loss of generality in focusing on directed parametrizations as for such distributions it is always possible to convert between directed and undirected parametrizations [35]. When thought of as as state space models, Gaussian tree models are often referred to as multiscale autoregressive models [5], although we will not use that terminology here.

Given a tree $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ we define a zero-mean Gaussian process $(x_v)_{v\in\mathcal{V}}$ where each x_v takes values in $\mathbb{R}^{n_v}$ for some n_v. We call the space in which x_v takes values the *state space* at v. The generative process defining $(x_v)_{v\in\mathcal{V}}$ is the following. If r denotes the root of the tree then $x_r \sim \mathcal{N}(0, R)$ and if $v \in \mathcal{V}_{\backslash r}$,

$$x_v = A_v x_{\mathcal{P}(v)} + w_v \tag{4.1}$$

where A_v is an $n_v \times n_{\mathcal{P}(v)}$ matrix, $w_v \sim \mathcal{N}(0, Q_v)$, w_v and w_u are independent if $u \neq v$, and for

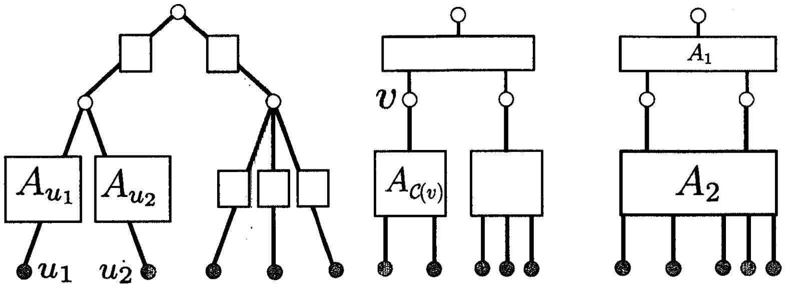

Figure 4-2: Three different abstractions of the tree that are present in our notation. On the left is the edge-level view of the tree and corresponding notation, in the center is the parent-children-level view of the tree and corresponding notation, and on the right is the scale-level abstraction of the tree and corresponding notation.

each $v \in \mathcal{V}_{\backslash r}$, w_v is independent of x_r. We always assume that the leaf variables are at scale n, and that only the leaf variables are observed.

We write $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ for the Gaussian latent tree model parametrized by $\mathcal{T}$, R, and Q_v and A_v for $v \in \mathcal{V}_{\backslash r}$. To avoid certain non-identifiability issues we assume, throughout, that R and each A_v and Q_v have full rank.

Since we do not specify the dimensions n_v of the state spaces *a priori*, almost all of our discussion is at the level of block matrices, where each block is indexed by a pair of vertices (u, v) and has dimension $n_u \times n_v$ as a matrix. If X is a block matrix indexed by subsets $\mathcal{U}$ and $\mathcal{W}$ of vertices, we abuse notation and terminology slightly and call X a $|\mathcal{U}| \times |\mathcal{W}|$ matrix. For example we call A_v, in (4.1) a $|\mathcal{P}(v)| \times |v|$ matrix. When we write such block matrices, we always assume that the vertices are ordered in such a way that vertices with a common parent are consecutively ordered. This is not an essential assumption, it just makes for more convenient notation.

Abstractions of the tree It will be useful to introduce notation that allows us to look at the tree at three levels of abstraction: the edge-level, the parent-children-level, and the scale-level. This notation is illustrated in Figure 4-2. We have already introduced the edge-level notation in (4.1).

Parent-Children Level At this level of abstraction (see Figure 4-2) we consider the influence of a parent on all of its children simultaneously. We show in Section 4.4.1 that, in some sense, this is the level at which the semidefinite program we formulate in Sec. 4.4 *operates*. Given a non-leaf vertex v define

$$A_{\mathcal{C}(v)} = \begin{bmatrix} A_{u_1} \\ \cdots \\ A_{u_m} \end{bmatrix}$$

where $\mathcal{C}(v) = \{u_1, \ldots, u_m\}$. Then (4.1) can be reformulated as

$$x_{\mathcal{C}(v)} = A_{\mathcal{C}(v)} x_v + w_{\mathcal{C}(v)}. \tag{4.2}$$

where if $\mathcal{U} \subset \mathcal{V}$ we write $x_{\mathcal{U}}$ and $w_{\mathcal{U}}$ for the appropriate sub-processes indexed by $\mathcal{U}$.

Scale-Level At this level of abstraction the tree is just a Markov chain (see Figure 4-2). This is the level of abstraction at which the SDP we formulate in Section 4.4 is *defined.* Given $i \geq 1$ define

$$A_i = \operatorname{diag}(A_{\mathcal{C}(v_1)}, \ldots, A_{\mathcal{C}(v_m)}) \tag{4.3}$$

where $\mathcal{V}_{i-1} = \{v_1, \ldots, v_m\}$ and $\operatorname{diag}(B_1, \ldots, B_m)$ is the block diagonal matrix with blocks $B_1, \ldots, B_m$. Then (4.1) can be reformulated as

$$x_i = A_i x_{i-1} + w_i \tag{4.4}$$

where we write x_i and w_i instead of $x_{\mathcal{V}_i}$ and $w_{\mathcal{V}_i}$ respectively.

Equivalence Since we assume we only observe the leaf variables of the Gaussian tree model defined by (4.1), we cannot distinguish models that realize the same covariance among the leaf variables. This gives rise to the following notion of equivalence for Gaussian latent tree models.

Definition 4. We say that two Gaussian latent tree models (indexed by the same tree) are *equivalent* if the covariance Σ_n of the leaf variables x_n is the same for both models.

4.3 Covariance Decompositions

As for time-indexed linear state space models we can solve for the leaf variables in terms of the $(w_v)_{v \in \mathcal{V}}$ as

$$x_n = (A_n \cdots A_1)x_0 + (A_n \cdots A_2)w_1 + \cdots + A_n w_{n-1} + w_n. \tag{4.5}$$

Let Σ_n be the covariance of x_n and Q_i be the covariance of w_i, noting that Q_i is diagonal as a block matrix. Taking covariances of (4.5) yields a decomposition of Σ_n that will play an important role in this chapter.

$$\Sigma_n = (A_n \cdots A_1)R(A_n \cdots A_1)^T + (A_n \cdots A_2)Q_1(A_n \cdots A_2)^T + \cdots + A_n Q_{n-1} A_n^T + Q_n \tag{4.6}$$

This decomposition is illustrated in Figure 4-3.

Block Diagonal Structures In a sense (4.6) abstracts away the tree structure, leaving only the chain structure of the scale-level view of a tree shown in Figure 4-2. The tree structure can be

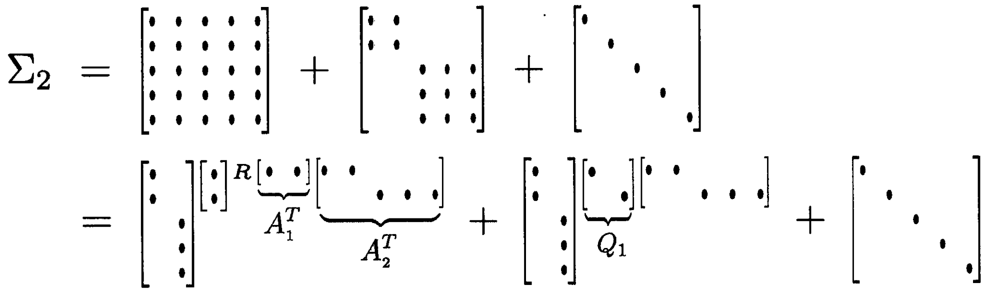

Figure 4-3: An illustration of the leaf-covariance decomposition described by Proposition 8 for the tree in Figure 4-1. The first equality represents the block diagonal structure of the terms, the second equality represents the low-rank and nested column space structures of the terms.

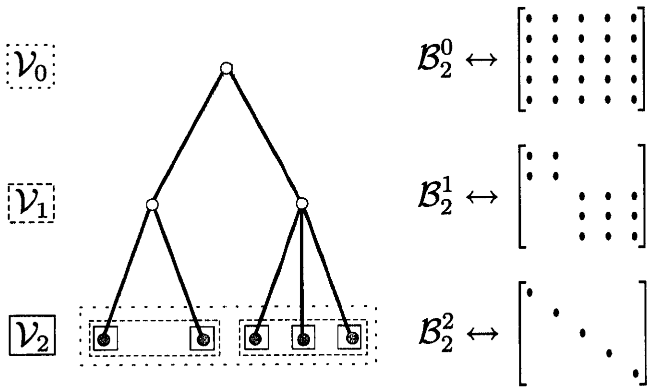
$\mathcal{V}_0$
$\mathcal{V}_1$
$\mathcal{V}_2$
$\mathcal{B}_2^0 \leftrightarrow$
$\mathcal{B}_2^1 \leftrightarrow$
$\mathcal{B}_2^2 \leftrightarrow$

Figure 4-4: An illustration of the block diagonal projections $\mathcal{B}_t^s$ defined in (4.7). For example, the vertices $\mathcal{V}_1$ induce a partition of $\mathcal{V}_2$ given by $\{\mathcal{C}_2(v_1), \mathcal{C}_2(v_2)\}$ shown by the dashed boxes. This partition defines the block pattern of $\mathcal{B}_2^1$.

captured by the block diagonal structure of the terms in the covariance decomposition.

Since each of the A_i and Q_j are block diagonal, it follows that all of the terms in (4.6) are block diagonal as illustrated in Figure 4-3. This structure arises statistically because if $v \in \mathcal{V}_i$ then the only variables at scale n that depend on w_v are those indexed by $\mathcal{C}_n(v)$, the descendants of v at scale n. As such $(A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T$ is block diagonal with blocks indexed by $\mathcal{C}_n(v)$ for $v \in \mathcal{V}_i$.

As is illustrated in Figure 4-3, the tree structure ensures that the block diagonal support patterns of each of the terms in (4.6) are nested. Specifically

$$\mathrm{supp}(Q_n) \subset \mathrm{supp}(A_n Q_{n-1} A_n^T) \subset \cdots \subset \mathrm{supp}((A_n \cdots A_1)R(A_n \cdots A_1)^T).$$

This is the case because if variable at scale n depends on w_v for some vertex v, then that variable necessarily depends on w_u for all vertices u which have v as a descendant.

Since this block diagonal structure plays a prominent role in this chapter, we introduce the notation $\mathcal{B}_n^i$ for the map that given a symmetric $|\mathcal{V}_n| \times |\mathcal{V}_n|$ matrix X is defined by

$$\mathcal{B}_n^i(X) = \mathrm{diag}(X_{\mathcal{C}_n(v_1)}, \ldots, X_{\mathcal{C}_n(v_m)}) \tag{4.7}$$

where $\mathcal{V}_i = \{v_1, \ldots, v_m\}$ and $X_{\mathcal{C}_n(i)}$ indexes the appropriate submatrix of X. Note that this is precisely the orthogonal projection onto the set of $|\mathcal{V}_n| \times |\mathcal{V}_n|$ matrices that are block diagonal with blocks indexed by $\mathcal{C}_n(v)$ for $v \in \mathcal{V}_i$. This definition is illustrated in Figure 4-4. Using this notation we can express the block diagonal structure of terms of the form $(A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T$ in a compact way by writing

$$\mathcal{B}_n^i((A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T) = (A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T.$$

Column Space and Low Rank Structures The nature of the factorized structure of each of the terms in the covariance decomposition (4.6) implies that these terms have nested column spaces. This structure arises because the tree is simply a Markov chain when viewed at the scale-level of abstraction.

Although it is not explicit in our formulation so far, we are always interested in *parsimonious* Gaussian latent tree models. As such, we are particularly interested in identifying models with low-dimensional state spaces at each vertex v. Since each of the terms $(A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T$ in the covariance decomposition (4.6) has rank equal to the sum of the dimensions of the state spaces corresponding to vertices in $\mathcal{V}_i$, we expect the terms in the decomposition to have low rank, with the exception of Q_n. Furthermore, as there are fewer vertices at coarser scales, we expect terms in the decomposition corresponding to coarser scales to have lower rank than those corresponding to finer scales.

A Covariance Decomposition Characterization of Gaussian Latent Trees We now formalize the salient features of the covariance decomposition (4.6). The following lemma shows that the essential structure of (4.6) is that the covariance at scale n of a Gaussian latent tree model can be expressed as the sum of $n+1$ block diagonal positive semidefinite matrices with nested column spaces and nested support.

Proposition 8. *Suppose Σ_n is the covariance at scale n of a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$. Then there exist positive semidefinite matrices $L_0, L_1, \ldots, L_n$ such that*

1. $\Sigma_n = L_0 + L_1 + \cdots + L_{n-1} + L_n$
2. *the column spaces $\mathcal{R}(L_i)$ of the L_i satisfy $\mathcal{R}(L_0) \subset \mathcal{R}(L_1) \subset \cdots \subset \mathcal{R}(L_{n-1})$*
3. *each L_i is block diagonal with $supp(L_n) \subset supp(L_{n-1}) \subset \cdots \subset supp(L_0)$*

Conversely, if $L_0, L_1, \ldots, L_n$ is a collection of positive semidefinite matrices satisfying properties 1–3 above, then there is a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ such that the covariance at scale n is Σ_n.

The proof of Proposition 8 is in Section 4.8.1. One direction of the proof follows directly from (4.6) and the subsequent discussion. Proving the converse requires showing how to map a decomposition of a covariance matrix Σ_n as $\Sigma_n = L_0 + \cdots + L_n$ with the properties stated in Proposition 8 to a Gaussian latent tree model. The algorithmic content of the proof is summarized in Algorithm 1.

Algorithm 1. Given a tuple $(L_0, \ldots, L_n)$ of symmetric positive semidefinite matrices satisfying properties 1–3 of Proposition 8, the following procedure produces a tree $\mathcal{T}$ and a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ with covariance at scale n given by $\sum_{i=0}^{n} L_i$.

1. Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be defined as follows. Associate a single vertex r with L_0. For each $i \geq 1$ associate a vertex v with each block in the block diagonal structure of L_i. There is an edge $(v, u) \in \mathcal{E}$ if and only if there is some i such that v corresponds to a block of L_i and u to a block of L_{i+1} with support contained in the block of L_i corresponding to v.
2. For $v \in \mathcal{V}_n$ let $Q_v = [L_n]_v$.
3. For $v \in \mathcal{V}_{n-1}$ let $A_{\mathcal{C}(v)}$ have columns given by an orthonormal set of eigenvectors of $[L_{n-1}]_{\mathcal{C}(v)}$ corresponding to non-zero eigenvalues. Let Q_v be the corresponding diagonal matrix of eigenvalues and define A_n in terms of the $A_{\mathcal{C}(v)}$ by (4.3).
4. For $i = n-1, n-2, \ldots, 0$, and for each $v \in \mathcal{V}_{i-1}$ choose $A_{\mathcal{C}(v)}$ to have columns given by an orthonormal set of eigenvectors of $[(A_n \cdots A_{i+1})^T L_{i-1} (A_n \cdots A_{i+1})]_{\mathcal{C}(v)}$ corresponding to non-zero eigenvalues. Let Q_v be the corresponding diagonal matrix of eigenvalues and define A_i in terms of the $A_{\mathcal{C}(v)}$ according to (4.3).

This algorithm also gives us a way to construct 'nice' parametrizations of Gaussian latent tree models. Suppose we are given a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$. We can set $L_0 = (A_n \cdots A_1)Q_0(A_n \cdots A_1)^T$, $L_n = Q_n$, and $L_i = (A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T$ for $i = 1, 2, \ldots, n-1$, and use Algorithm 1 to construct a new parametrization of the Gaussian latent tree model from the L_i. When we do this we are just choosing a nice basis for the state space of each of the hidden variables in our model. For future reference let us give a name to these 'nice' parametrizations.

Definition 5. A parametrization $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ of a Gaussian latent tree model is *normalized* if for all non-leaf vertices v, $A_{\mathcal{C}(v)}^T A_{\mathcal{C}(v)} = I$ and Q_v is diagonal and has full rank .

Note that Algorithm 1 always produces normalized parametrizations and Proposition 8 implies that every Gaussian latent tree model has a normalized parametrization. Normalized parametrizations are *not* unique, but different normalized parametrizations for equivalent (in the sense of Definition 4) Gaussian latent tree models have the same structural properties, such as state space dimensions.

In light of Proposition 8 we can state our problem of interest more abstractly as follows.

Problem 2. Suppose $L_0^\star, L_1^\star, \ldots, L_n^\star$ are block diagonal positive semidefinite matrices such that $\mathcal{R}(L_0^\star) \subset \cdots \subset \mathcal{R}(L_n^\star)$ for $0 \leq i \leq n$ and $\operatorname{supp}(L_n^\star) \subset \cdots \operatorname{supp}(L_0^\star)$. Given $\operatorname{supp}(L_i^\star)$ for $i = 0, 1, \ldots, n$ and the sum

$$\Sigma_n = L_0^\star + L_1^\star + \cdots + L_n^\star$$

recover the $L_i^\star$ for $0 \leq i \leq n$.

If we had a method to solve Problem 2 then, given the covariance at scale n of a Gaussian latent tree model, we could use this method to find the $L_i^\star$ and then use Algorithm 1 to reconstruct from them a normalized parametrization of a Gaussian latent tree model. In Section 4.4 we develop a method to partially solve Problem 2 based on semidefinite programming.

4.4 Covariance Decomposition SDP

In this section we propose a semidefinite programming-based heuristic that attempts to solve Problem 2. The SDP we formulate is a generalization of minimum trace factor analysis, the focus of our attention in Chapter 3. Our SDP explicitly addresses the assumptions that the terms of the decomposition should be low rank, positive semidefinite matrices with nested block diagonal support. Our SDP, however, does not explicitly enforce the subspace inclusion constraint. While the subspace inclusion constraint is semidefinite representable (see, for example, Lemma 2.1 of [46]), it is difficult to impose in practice using interior point solvers for semidefinite programs.

The constraints that each L_i is positive semidefinite and has a particular block diagonal structure are straightforward to incorporate into a semidefinite programming framework. We now address the assumption that the L_i (for $0 \leq i < n$) are low rank. As in Section 3.4 of Chapter 3, we again employ the heuristic that minimizing the trace of positive semidefinite matrices is a good convex surrogate for minimizing the rank. So we choose the objective of the SDP to be $\sum_{i=0}^{n-1} \lambda_i \mathrm{tr}(L_i)$ where the λ_i are non-negative scalars. It turns out that our analysis will require that if $i < j$ then $\lambda_i > \lambda_j$. This is intuitively appealing because, following the discussion in Section 4.3, if $i < j$ then we expect $\mathrm{rank}(L_i) < \mathrm{rank}(L_j)$ so it makes sense to penalize the term in the objective corresponding to $\mathrm{tr}(L_i)$ more than that corresponding to $\mathrm{tr}(L_j)$.

The Primal SDP Putting these pieces together we can write down an SDP-based heuristic to decompose Σ_n into its constituents.

$$
\begin{aligned}
& (\hat{L}_0, \hat{L}_1, \ldots, \hat{L}_n) \in \arg\min \sum_{i=0}^{n-1} \langle \lambda_i I, L_i \rangle \\
\text{subject to} \quad & \Sigma_n = \sum_{i=0}^{n} \mathcal{B}_n^i(L_i) \\
& L_i \succeq 0 \quad \text{for } i = 0, 1, \ldots, n
\end{aligned}
\tag{4.8}
$$

and the λ_i are non-negative parameters of the SDP satisfying $0 = \lambda_n < \lambda_{n-1} < \cdots < \lambda_1$. Without loss of generality we can take $\lambda_1 = 1$, as this serves to fix a normalization for the objective of (4.8).

The Dual SDP Observe that (4.8) is a conic program in standard form (see Section 2.2) so we can write down its dual by inspection.

$$
\begin{aligned}
& \max_Y \langle \Sigma_n, Y \rangle \\
& \text{s.t. } \lambda_i I - \mathcal{B}_n^i(Y) \succeq 0 \text{ for } i = 0, 1, \ldots, n.
\end{aligned}
\tag{4.9}
$$

where we have used the fact that $\mathcal{B}_n^i$ is self adjoint for $i = 0, 1, \ldots, n$.

We now establish that strong duality holds for this primal-dual pair of semidefinite programs under the assumption that $\Sigma_n \succ 0$.

Lemma 11. *If $\Sigma_n \succ 0$ then strong duality holds for the primal-dual pair* (4.8) *and* (4.9).

Proof. We establish this by verifying Slater's condition (see Section 2.2). If $\Sigma_n \succ 0$, let σ_{min} denote the smallest eigenvalue of Σ_n. Then take $L_0 = \Sigma_n - (\sigma_{min}/2)I$ and $L_i = (\sigma_{min}/(2n))I$ for $i = 1, 2, \ldots, n$. Then $L_i \succ 0$ for $i = 0, 1, \ldots, n$ and $\sum_{i=0}^{n} L_i = \Sigma_n$ so the primal problem is strictly feasible. To complete the proof we note that the primal objective function is bounded below by zero. □

From now on we will assume that $\Sigma_n \succ 0$, so that Lemma 11 ensures that strong duality holds.

4.4.1 Analysis of the Covariance Decomposition SDP

In this section we analyze the SDP formulated in Section 4.4. In particular we give conditions on the parameters of an underlying Gaussian latent tree model (or equivalently on the $L_i^\star$ arising from that model) under which the SDP (4.8) successfully solves Problem 2.

Definition 6. Suppose $\Sigma_n = L_0^\star + L_1^\star + \cdots + L_n^\star$ where the $L_i^\star$ satisfy the assumptions of Problem 2. If the SDP (4.8) has a unique optimal point $(\hat{L}_0, \ldots, \hat{L}_n)$ and $\hat{L}_i = L_i^\star$ for $i = 0, 1, \ldots, n$ then we say that the SDP (4.8) *correctly decomposes* Σ_n.

The following result establishes conditions under which the covariance decomposition SDP correctly decomposes Σ_n. The conditions are a specialization of the usual optimality conditions for semidefinite programming (see Section 2.2) to this context.

Proposition 9. *If there exists a dual certificate Y such that*

1. $\lambda_i I - \mathcal{B}_n^i(Y) \succeq 0$ *for* $i = 0, 1, \ldots, n$
2. $L_i^\star(\lambda_i I - \mathcal{B}_n^i(Y)) = 0$ *for* $i = 0, 1, \ldots, n$

then the SDP (4.8) *correctly decomposes* Σ_n.

The proof is in Section 4.8.2. The part of the proof that is somewhat involved is proving that the SDP has a unique solution.

It is not particularly obvious how to construct a Y with the properties stated in Proposition 9 as these properties are rather global in nature. It turns out that we can simplify the task of constructing Y by combining dual certificates that are defined locally—certificates that concern only the interactions between a parent and all of its children. This is the main technical lemma of this chapter.

Lemma 12. *Let Σ_n be the covariance at scale n of a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$. Suppose that for each non-leaf vertex v there is a $|\mathcal{C}(v)| \times |\mathcal{C}(v)|$ symmetric positive semidefinite matrix M_v such that*

1. $[M_v]_{uu} = I$ *for all* $u \in \mathcal{C}(v)$
2. $M_v A_{\mathcal{C}(v)} = 0$.

Then there exists Y with the properties stated in Proposition 9 and so the SDP (4.8) *correctly decomposes* Σ_n.

The proof of Lemma 12 is in Section 4.8.3. Lemma 12 allows us to consider only the apparently more simple situation of finding local dual certificates. In particular, any results about constructing matrices M_v satisfying the conditions of Lemma 12 translate into results about the success of the covariance decomposition SDP (4.8).

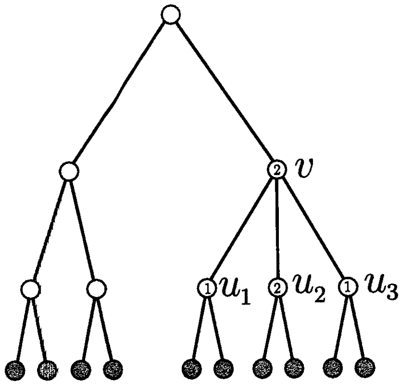

Figure 4-5: The tree $\mathcal{T}$ shown, and in particular the subtree consisting of v and its children $\mathcal{C}(v) = \{u_1, u_2, u_3\}$ is used to illustrate the notation in Lemma 12. We assume that there is some Gaussian latent tree model $\mathcal{M}_\mathcal{T}(R, Q_v, A_v)$ defined with respect to $\mathcal{T}$ with state space dimensions given by the numbers inside the vertices of the tree.

Example 1. Let us clarify our notation here with an example. Consider the tree in Figure 4-5 and, specifically, the vertex v and its children $\mathcal{C}(v)$. Note that the dimension of the state space at v is 2 and the dimensions of the state spaces at u_1, u_2, and u_3, the children of v, are 1, 2, and 1 respectively. Then $A_{\mathcal{C}(v)}$ is a 4×2 matrix, and also a 3×1 block matrix with the block rows indexed by u_1, u_2, u_3 and the block column indexed by v. So $A_{\mathcal{C}(v)}$ has the form

$$A_{\mathcal{C}(v)} = \left[\begin{array}{c} A_{u_1} \\ \hline A_{u_2} \\ \hline A_{u_3} \end{array}\right] = \left[\begin{array}{cc} * & * \\ \hline * & * \\ * & * \\ \hline * & * \end{array}\right].$$

Hence if there were an M_v satisfying the assumptions of Lemma 12, M_v would be a symmetric positive semidefinite 4×4 matrix, (and 3×3 as a block matrix) of the form

$$M_v = \left[\begin{array}{c|cc|c} 1 & * & * & * \\ \hline * & 1 & 0 & * \\ * & 0 & 1 & * \\ \hline * & * & * & 1 \end{array}\right]$$

satisfying $M_v A_{\mathcal{C}(v)} = 0$.

4.4.2 Success of the Covariance Decomposition SDP

Scalar variables In the case where the underlying Gaussian latent tree model has all scalar variables (that is, the state space at each vertex has dimension one) then we can combine Theorem 11, due to Delorme and Poljak [19], with Lemma 12 to characterize when the covariance decomposition SDP succeeds. The result essentially says that as long as no vertex is much more strongly influenced by its parent than all of its 'siblings', the covariance decomposition SDP succeeds.

Theorem 17. *Suppose $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ is a Gaussian latent tree model with $n_v = 1$ for each vertex v and leaf covariance Σ_n. If for all non-leaf vertices v*

$$|A_u| \leq \sum_{w \in \mathcal{C}(v)\setminus\{u\}} |A_w| \quad \textit{for all } u \in \mathcal{C}(v) \tag{4.10}$$

then the covariance decomposition SDP (4.8) *correctly decomposes Σ_n.*

Remark. The condition in (4.10) imposes an interesting structural restriction on the trees $\mathcal{T}$ with respect to which a Gaussian latent tree model is defined if we hope to identify the model using the SDP (4.8). Suppose $v \in \mathcal{V}$ has just two children, u_1 and u_2. Then the balance condition says that we must have

$$|A_{u_1}| = |A_{u_2}|$$

a condition that does not hold generically. As such, in order for the parameters of a Gaussian latent tree model to be generically balanced, we need every vertex in the tree $\mathcal{T}$ to have at least three children. Even at this qualitative level, Theorem 17 gives us insight about how we ought *not* to go about choosing our tree $\mathcal{T}$ when trying to solve a modeling problem as our procedure will clearly not be effective on trees having vertices with only two children.

Non-scalar variables In the case where the underlying Gaussian latent tree model has non-scalar variables, we cannot directly apply results from Chapter 3. Nevertheless, it is possible to generalize the main deterministic result, Theorem 5, from Chapter 3 so that it does apply in this setting.

Theorem 18. *Suppose $\mathcal{M}_{\mathcal{T}}(R, Q_v, A_v)$ is a Gaussian latent tree model with leaf covariance Σ_n. If, for all non-leaf vertices v,*

$$[P_{\mathcal{R}(A_v)}]_{uu} \prec (1/3)I \quad \textit{for all } u \in \mathcal{C}(v)$$

then the covariance decomposition SDP (4.8) *correctly decomposes Σ_n.*

Theorem 18 essentially says that as long as the column space $\mathcal{R}(A_v)$ of each of the matrices A_v is not too closely aligned with any of the coordinate subspaces indexed by $u \in \mathcal{C}(v)$, the covariance decomposition SDP will succeed. We defer the proof of this result to Section 4.8.4, as it is a direct generalization of the proof of Theorem 5 in Chapter 3. We note that unlike Theorem 17, Theorem 18 only provides a sufficient condition for the success of the covariance decomposition SDP.

While it would be possible to generalize the randomized results of Chapter 3 to apply in this case, these results do not make a great deal of sense in the present setting. This is because they would require the degree of the vertices in the tree to be growing to apply, which is quite an unnatural assumption.

4.4.3 Producing valid approximate decompositions

Recall from Problem 2 that we aim to decompose Σ_n into a sum of block diagonal positive semidefinite matrices L_i with *nested column spaces*. We did not, however, impose the constraint that the column spaces of the L_i be nested in our covariance decomposition SDP (4.8). If the conditions in Theorem 17 or Theorem 18 hold, then this is not a problem, as the covariance decomposition SDP correctly decomposes Σ_n and so the matrices $\hat{L}_i$ do satisfy the column space nesting constraints.

In the case when the covariance decomposition SDP fails to correctly decompose Σ_n, we have no guarantee that the column spaces of the $\hat{L}_i$ are nested, and in general they are not. This problem will also occur when we consider the 'noisy' version of the decomposition problem in Section 4.5. We now describe a method that takes a solution $(\hat{L}_0, \ldots, \hat{L}_n)$ of the covariance decomposition SDP and produces from it a new tuple of positive semidefinite matrices $(\tilde{L}_0, \ldots, \tilde{L}_n)$ that have the same support as $(\hat{L}_0, \ldots, \hat{L}_n)$ and also satisfy the subspace nesting constraint. The price we pay for this is that it is no longer the case that $\sum_{i=0}^{n} \tilde{L}_i = \Sigma_n$.

Algorithm 2. Given a tuple of symmetric matrices $(\hat{L}_0, \ldots, \hat{L}_n)$ the procedure produces a tuple of symmetric matrices $(\tilde{L}_0, \ldots, \tilde{L}_n)$ satisfying $\mathcal{R}(\tilde{L}_0) \subset \cdots \subset \mathcal{R}(\tilde{L}_n)$.

```
1. Initialize by setting $\tilde{L}_n \leftarrow \hat{L}_n$
2. For $j = n-1, n-2, \ldots, 0$
       $V \leftarrow \mathcal{R}(\tilde{L}_{j+1})$
       $\tilde{L}_j \leftarrow P_V \hat{L}_j P_V$ (where $P_V$ is the orthogonal projection onto the subspace $V$).
```

For convenience we refer to this procedure as a 'rounding' scheme, as it enforces a constraint that we omit from our convex program.

While we could modify this procedure to ensure that the resulting tuple of matrices also satisfies $\sum_i \tilde{L}_i = \Sigma_n$, such a modification would most likely cause an increase in the rank of the $\tilde{L}_i$. Given that we never, in practice, aim to exactly realize a given covariance matrix, from a modeling perspective it makes more sense to produce a valid parsimonious model that approximately realizes the given covariance, than to go to pains to produce an exact decomposition that is no longer parsimonious.

4.5 An Approximate Covariance Decomposition SDP

In practice we do not have access to the covariance among the leaf variables, only to some covariance matrix $\hat{\Sigma}_n$ that may be an approximation of the covariance among the leaf variables of a Gaussian latent tree model. A natural variation on the SDP (4.8) proposed in Section 4.4 to deal with this case is to replace the equality constraint $\Sigma_n = \sum_{i=0}^{n} \mathcal{B}_n^i(L_i)$ with the minimization of a convex loss function $f(\hat{\Sigma}_n, \mathcal{B}_n^0(L_0), \mathcal{B}_n^1(L_1), \ldots, \mathcal{B}_n^n(L_n))$. An example of such a function might be $\|\hat{\Sigma}_n - \sum_{i=0}^{n} \mathcal{B}_n^i(L_i)\|_F$ where $\|X\|_F = \left(\sum_{i,j} X_{ij}^2\right)^{1/2}$ is the Frobenius norm of a matrix. This is the example we use for our experiments in Section 4.6, but it is by no means a canonical choice.

We formulate the approximate covariance decomposition SDP as follows.

$$\begin{aligned} &\min f(\hat{\Sigma}_n, \mathcal{B}_n^0(L_0), \mathcal{B}_n^1(L_1), \ldots, \mathcal{B}_n^n(L_n)) + \gamma \left(\sum_{i=0}^{n-1} \langle \lambda_i I, L_i \rangle \right) \\ &\text{s.t. } L_i \succeq 0 \quad \text{for } i = 0, 1, \ldots, n \end{aligned} \tag{4.11}$$

where $\gamma > 0$ is a regularization parameter that balances the competing objectives of building a model that matches the observations (i.e. $\hat{\Sigma}_n$) and has low complexity in the sense of low total state dimension.

If we are given some covariance matrix $\hat{\Sigma}_n$ and want to approximate it by the scale-n covariance of a Gaussian latent tree model, we again have the problem that the estimates $\hat{L}_i$ produced by solving the convex program (4.11) do not satisfy the subspace containment constraints $\mathcal{R}(\hat{L}_0) \subset \cdots \subset \mathcal{R}(\hat{L}_n)$. If we apply Algorithm 2 to the $\hat{L}_i$, the output $\tilde{L}_i$ is a sum of positive semidefinite block diagonal matrices with nested column spaces and so corresponds to a valid Gaussian latent tree model.

4.6 Experiments

In this section we focus on demonstrating the consistency of the approximate covariance decomposition convex program (4.11) (followed by Algorithm 2) using synthetic experiments. We focus on the case where the loss function is $f(\hat{\Sigma}_n, \mathcal{B}_n^0(L_0), \ldots, \mathcal{B}_n^n(L_n)) = \|\hat{\Sigma}_n - (\mathcal{B}_n^0(L_0) + \mathcal{B}_n^1(L_1) + \ldots + \mathcal{B}_n^n(L_n))\|_F$.

In particular we assume we are given i.i.d. samples of the leaf-variables of the two Gaussian latent tree models shown in Figure 4-6 with state space dimensions given by the numbers next to the vertices in that figure. In each of the two models the matrices $A_{\mathcal{C}(v)}$ are chosen as follows. If $A_{\mathcal{C}(v)}$ has only one column then we take $A_{\mathcal{C}(v)} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T + N$ where $N \sim \mathcal{N}(0, \frac{1}{100}I)$. If $A_{\mathcal{C}(v)}$ has two columns, the first is chosen as in the previous sentence and the second is chosen to be $\begin{bmatrix} 1 & \cdots & 1 & -1 & \cdots & -1 \end{bmatrix}^T + N'$ where $N' \sim \mathcal{N}(0, \frac{1}{100}I)$ and is independent of N. These choices

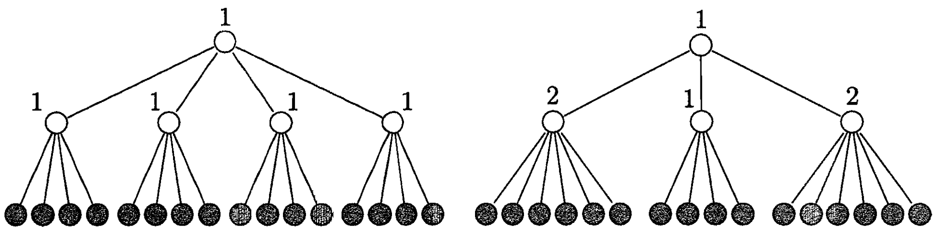

Figure 4-6: The trees with respect to which the Gaussian latent tree models in our experiments are defined. The numbers next to the unobserved vertices are the state dimensions at those vertices. All the leaves have state dimension one.

ensure that the columns of $A_{C(v)}$ satisfy the deterministic conditions of Theorems 17 and 18. We investigate the number of samples required for the output of the convex program (4.11) followed by Algorithm 2 to correspond to a model with the correct state dimensions at each vertex.

Explicitly for each of the two models and each value of N in a range (shown in Figure 4-7 we perform the following steps fifty times

1. Construct N i.i.d. samples from the leaf variables of the model and form the associated sample covariance matrix $\hat{\Sigma}_2^N$.

2. Solve the approximate covariance decomposition convex program with $\gamma = 0.6$ and $\lambda_0 = 1, \lambda_1 = 0.5, \lambda_2 = 0$ (using a combination of YALMIP [41] and SDPT3 [56]) with input $\hat{\Sigma}_2^N$ and the appropriate tree $\mathcal{T}$ shown in Figure 4-6.

3. Round the solution of the convex program by applying Algorithm 2.

4. Use Algorithm 1 to explicitly construct a normalized parametrization $\mathcal{M}_{\mathcal{T}}(\hat{R}, \hat{Q}_v, \hat{A}_v)$ of a Gaussian latent tree model from the rounded solution of the convex program.

5. Check if the dimensions of $\hat{R}$ and the $\hat{Q}_v$ match the state dimensions of the underlying models shown in Figure 4-6.

We note that the recovered structures were typically the same for a range of γ around the chosen value. In practice the regularization parameters γ and λ_1 could be chosen by cross-validation.

Figure 4-7 shows the results indicating that with sufficiently many samples, the method is successful in both cases in recovering the state dimensions of the model, demonstrating that this procedure is likely to be consistent under appropriate assumptions on the underlying model and on the choices of regularization parameter.

Computational Complexity It is not straightforward to give an accurate estimate of the computational complexity of interior point methods for solving SDPs as the complexity depends a great

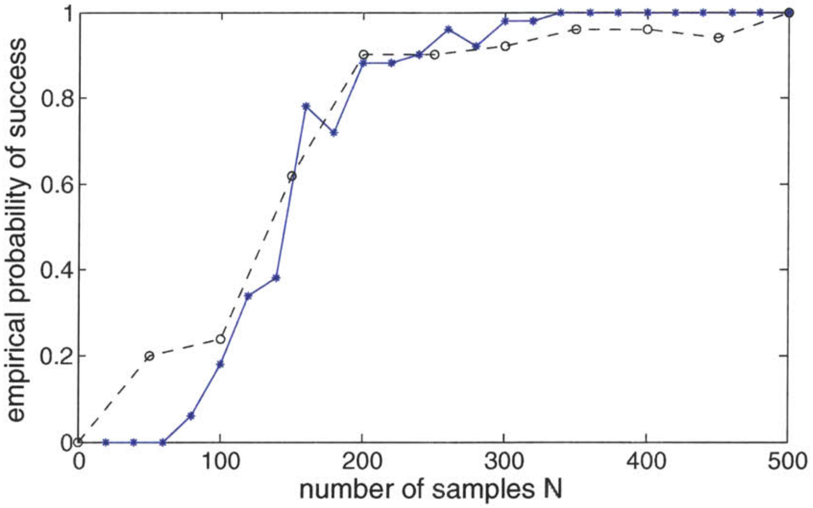

Figure 4-7: For each of the two trees shown in Figure 4-6 and each N we repeat the following procedure 50 times. We form a sample covariance matrix from N i.i.d. samples of the leaf variables of a Gaussian latent tree model defined with respect to the given tree with the given state dimensions. We use our method to learn the parameters and state dimensions of a Gaussian latent tree model, and check whether the state dimensions match those of the underlying model. On the vertical axis we plot the proportion of trials in which all state dimensions were identified correctly. The solid blue curve corresponds to the tree on the left in Figure 4-6 and the dashed black curve corresponds to the tree on the right in Figure 4-6.

deal on the extent to which problem structure is exploited in the solver. In particular, when using a high-level modeling language like YALMIP to interface with a lower-level solver like SDPT3, it is not always clear what transformations have been applied to exploit problem structure. Nevertheless, to give a sense of how our method scales with problem size, we give a conservative estimate of the complexity of solving the exact covariance decomposition SDP for a tree where each vertex has q children and the leaves of the tree are at scale d.

Let $n = q^d$ be the number of leaf variables. Then the covariance decomposition SDP (4.8) in has

$$\sum_{i=0}^{d} q^{d-i}\binom{q^i+1}{2} = \frac{q^d}{2}\left(1+q+\cdots+q^d+d+1\right) = O(n(n+d))$$

variables and $O(n^2)$ linear equality constraints, making the complexity per iteration of an interior point solver in the worst case $O(n^6)$ [11]. This analysis is very conservative, ignoring a good deal of the block diagonal structure in the problem. Since the problem we are solving is very structured, it is likely that emerging first-order methods for solving structured convex programs could be applied to this problem [6]. Such methods have lower the computational complexity, and, perhaps more importantly in practice, require much less memory than generic logarithmic barrier-based interior-point methods.

Another possible approach to reducing the complexity of semidefinite-programming-based methods for the problem considered in this chapter is to develop related methods that are less global, but more computationally tractable, than those considered in this chapter. We briefly discuss this point of view in Chapter 5.

4.7 Discussion

In this chapter we have considered the problem of learning the parameters and state dimensions of a Gaussian latent tree model given the tree and the covariance among the leaf variables. We formulated an SDP that, together with a 'rounding' scheme, approximately decomposes the covariance matrix among the leaf variables as a sum of block diagonal positive semidefinite matrices with nested column spaces. Given any such decomposition we can recover (by Algorithm 1) an explicit parametrization of a Gaussian latent tree model that approximately realizes the given covariance. We give conditions on the underlying latent tree model that ensure the decomposition procedure identifies the correct model (up to equivalence). Finally we formulate a variation on the covariance decomposition SDP that, together with a 'rounding' scheme, approximately decomposes a given covariance matrix as a sum of block diagonal positive semidefinite matrices with nested column spaces. We demonstrate the stability properties of this approximate decomposition method by simulation, showing that given sufficiently many i.i.d. samples of the leaf variables of an underlying Gaussian latent tree model, our method can correctly identify the state dimensions of the underlying model.

We discuss avenues of further research stemming from this work in Chapter 5.

4.8 Proofs for Chapter 4

In this appendix we provide proofs for main technical results in this chapter.

4.8.1 Proof of Proposition 8

Proof of Proposition 8. To prove the first part of Proposition 8 we take $L_0 = (A_n \cdots A_1)R(A_n \cdots A_1)^T$, $L_n = Q_n$ and for $1 \leq i \leq n-1$, $L_i = (A_n \cdots A_{i+1})Q_i(A_n \cdots A_{i+1})^T$. From the discussion prior to the statement of Proposition 8 it is clear that the L_i satisfy properties 1 and 2. Since $\mathcal{B}_n^i(L_i) = L_i$ the L_i are block diagonal. If $j < i$ then for every $v \in \mathcal{V}_i$ there is some $u \in \mathcal{V}_j$ such that $\mathcal{C}_n(v) \subset \mathcal{C}_n(u)$. Indeed u is the ancestor of v at scale j. Hence the support of L_j contains the support of L_i whenever $j < i$, verifying property 3.

Conversely, suppose we are given $L_0, L_1, \ldots, L_n$ satisfying properties 1–3. We argue by induction on n that there is a Gaussian latent tree model that has Σ_n as the covariance at scale n.

If $n = 0$ we choose the tree $\mathcal{T} = \mathcal{T}_0$ to have a single vertex and no edges, and take $R = L_0$. Suppose that if $n = k$ and $\bar{L}_0, \ldots, \bar{L}_k$ satisfy properties 1–3, then there is some Gaussian latent tree model $\mathcal{M}_{\mathcal{T}_k}(R, Q_v, A_v)$ that has $\sum_{i=0}^k \bar{L}_k$ as the covariance among its leaf variables. Now consider the case $n = k+1$. Since $\mathcal{R}(L_i) \subset \mathcal{R}(L_k)$ for $i \leq k$ we can write

$$\Sigma_{k+1} = L_{k+1} + A_{k+1}(\bar{L}_k + \cdots + \bar{L}_0)A_{k+1}^T$$

for some $\bar{L}_i$, $i = 0, 1, \ldots, k$ where A_{k+1} is a matrix the columns of which are an orthonormal basis for the space spanned by the eigenvectors corresponding to the non-zero eigenvalues of L_k. Since each of the L_i are block diagonal with nested support, the same is true of the $\bar{L}_i$. Since L_k is block diagonal, we can take A_{k+1} to be block diagonal (with corresponding block-diagonal structure). Similarly since $\mathcal{R}(L_j) \subset \mathcal{R}(L_i)$ for $j \leq i$ the same holds for the $\bar{L}_i$.

Applying the induction hypothesis there is a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}_k}(R, Q_v, A_v)$ such that the covariance Σ_k at scale k of the model is $\sum_{i=0}^k \bar{L}_k$. Let $\mathcal{T}_{k+1}$ be a tree constructed as follows. Take $\mathcal{T}_k$ and add a new vertex for each block in the block diagonal structure of L_{k+1}. Then if v is any of these new vertices, add an edge between v and the leaf of $\mathcal{T}_k$ that corresponds to a block of $A_{k+1}\bar{L}_k A_{k+1}^T$ that contains the block of L_{k+1} corresponding to v. Finally, for $v \in \mathcal{V}_k$ we take $A_{\mathcal{C}(v)}$ to be the relevant block of A_{k+1}, and for $v \in \mathcal{V}_{k+1}$ we take $Q_v = [L_{k+1}]_v$, the block of L_{k+1} corresponding to v. This specifies a Gaussian latent tree model $\mathcal{M}_{\mathcal{T}_{k+1}}(R, Q_v, A_v)$ with the desired properties. □

4.8.2 Proof of Proposition 9

Recall that Proposition 9 provides conditions under which we can certify that $L_0^\star, \ldots, L_n^\star$ is the unique solution of the covariance decomposition SDP.

Proof of Proposition 9. By assumption the $(L_0^\star, \ldots, L_n^\star)$ is a feasible point for (4.8). The first property in the statement of Proposition 9 simply states that Y is dual feasible. The second property is the complementary slackness condition for semidefinite programming. Since we have constructed primal and dual feasible points that satisfy the complementary slackness conditions it follows that $(L_0^\star, \ldots, L_n^\star)$ is *an* optimal point of the primal SDP (4.8).

It remains to show that under these conditions $(L_0^\star, \ldots, L_n^\star)$ is the unique optimal point of the primal SDP (4.8). Arguing by contradiction, assume this is not the case. Then there is some other optimal point $(\tilde{L}_0, \ldots, \tilde{L}_n)$ of the primal SDP. Since $(L_0^\star, \ldots, L_n^\star) \neq (\tilde{L}_0, \ldots, \tilde{L}_n)$ and yet $\sum_{i=0}^n L_i^\star = \sum_{i=0}^n \tilde{L}_i$ it follows that $L_i^\star \neq \tilde{L}_i$ for at least two indices i. Let $j_1 < j_2$ be the smallest two indices such that $L_{j_1}^\star \neq \tilde{L}_{j_1}$ and $L_{j_2}^\star \neq \tilde{L}_{j_2}$.

By convexity $((L_0^\star + \tilde{L}_0)/2, \ldots, (L_n^\star + \tilde{L}_n)/2)$ is also an optimal point for the primal SDP. Hence there exists Y such that $\lambda_i I - \mathcal{B}_n^i(Y) \succeq 0$ for $i = 0, 1, \ldots, n$ and $\mathcal{B}_n^i(Y)(L_i^\star + \tilde{L}_i) = 0$ for $i = 0, 1, \ldots, n$. Since $L_{j_1}^\star \succeq 0$ and $\tilde{L}_{j_1} \succeq 0$ and $(\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y))(L_{j_1}^\star + \tilde{L}_{j_1}) = 0$ it follows that $(\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y))L_{j_1}^\star = 0$ and $(\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y))\tilde{L}_{j_1} = 0$. Hence

$$(\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y))(L_{j_1}^\star - \tilde{L}_{j_1}) = 0. \tag{4.12}$$

Note that by our choice of j_1 and j_2, $L_{j_1}^\star - \tilde{L}_{j_1}$ satisfies $\mathcal{B}_n^{j_2}(L_{j_1}^\star - \tilde{L}_{j_1}) = L_{j_1}^\star - \tilde{L}_{j_1}$. It then follows from (4.12), the fact that $j_1 < j_2$ so $\mathcal{B}_n^{j_2}\mathcal{B}_n^{j_1} = \mathcal{B}_n^{j_2}$, and properties of block diagonal matrices that

$$\mathcal{B}_n^{j_2}\left[(\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y))(L_{j_1}^\star - \tilde{L}_{j_1})\right] = \mathcal{B}_n^{j_2}(\lambda_{j_1} I - Y)\mathcal{B}_n^{j_2}(L_{j_1}^\star - \tilde{L}_{j_1}) = 0. \tag{4.13}$$

To reach a contradiction, it suffices to show that $L_{j_1}^\star - \tilde{L}_{j_1} = 0$. To achieve this, we need only show that $\lambda_{j_1} I - \mathcal{B}_n^{j_1}(Y)$ is invertible as then we can solve (4.12) to obtain $L_{j_1}^\star - \tilde{L}_{j_1} = 0$. Since $j_1 < j_2$ it follows from our choice of the λ_i that $\lambda_{j_1} > \lambda_{j_2}$. Hence

$$\lambda_{j_1} I - \mathcal{B}_n^{j_2}(Y) \succ \lambda_{j_2} I - \mathcal{B}_n^{j_2}(Y) \succeq 0$$

since Y is dual feasible. This establishes that $\lambda_{j_1} I - \mathcal{B}_n^{j_2}(Y)$ is invertible, showing that $L_{j_1}^\star = \tilde{L}_{j_1}$, yielding a contradiction. □

4.8.3 Proof of Lemma 12

Lemma 12 shows that to construct a global dual certificate that proves that the covariance decomposition SDP correctly decomposes the scale-n covariance of a Gaussian latent tree model, it suffices to construct and combine local dual certificates corresponding to the subtrees consisting of just a parent and all of its children.

Before providing a proof of Lemma 12 we record a straightforward result that holds simply because any principal submatrix of a positive semidefinite matrix is itself positive semidefinite.

Lemma 13. *If $0 \leq i \leq n$ and $X \succeq 0$ is a $|\mathcal{V}_i| \times |\mathcal{V}_i|$ matrix then $\mathcal{B}_n^i(X) \succeq 0$.*

Furthermore, it will be useful to have notation for the state transition matrix [48] corresponding to the scale-level abstraction of a Gaussian latent tree model. For $i, j \geq 0$ define

$$\Phi_j^i = \begin{cases} I & \text{if } i = j \\ A_j \cdots A_{i+1} & \text{if } i < j \\ {\Phi_i^j}^T & \text{if } i \geq j. \end{cases}$$

Note that the state transition matrix depends on the parametrization of the Gaussian latent tree model. In the case where the model has a normalized parametrization, the state transition matrix has some nice properties that will be used a number of times in the sequel.

Lemma 14. *If $\mathcal{M}_\mathcal{T}(R, Q_v, A_v)$ is a normalized parametrization then if $k \geq j \geq i$*

1. $\Phi_i^j \; \Phi_j^i = I$
2. $\Phi_j^i \; \Phi_i^j \preceq I$
3. $\Phi_j^k \; \Phi_k^i = \Phi_j^i$.

Proof. Since the parametrization is normalized $A_\mathcal{C}(v)^T A_{\mathcal{C}(v)} = I$ for all non-leaf variables v. Hence for all $i = 1, 2, \ldots, n$, $A_i^T A_i = I$ and so the first statement holds by the definition of Φ_i^j. The second statement holds simply because if X is a matrix that satisfies $X^T X = I$ then its non-zero singular values are all one, hence the eigenvalues of XX^T are all bounded above by one. Finally the third statement follows from the definition of $\Phi_\bullet^\bullet$ and the first statement. □

Proof of Lemma 12. First, note that the assumptions in Lemma 12 are independent of the choice of basis for the state spaces at each non-leaf vertex v. Hence we can assume, without loss of generality, that the parametrization is normalized. As such we liberally make use of the results in Lemma 14.

We first take each of the certificates M_v and combine them to give a certificate for each scale. Indeed we define for each $1 \leq i \leq n$ the $|\mathcal{V}_i| \times |\mathcal{V}_i|$ matrix

$$M_i = I - \operatorname{diag}(M_{v_1}, \ldots, M_{v_m})$$

where $\mathcal{V}_{i-1} = \{v_1, \ldots, v_m\}$.

Claim 1. The matrices M_i have the following properties.

1. $M_i \preceq I$
2. If $k < i$ then $M_i \, \Phi_i^k = \Phi_i^k$.
3. If $k \geq j \geq i$ then $\mathcal{B}_k^j(\Phi_k^i \, M_i \, \Phi_i^k) = 0$.

Proof. Since each $M_v \succeq 0$, it follows that $\mathrm{diag}(M_{v_1}, \ldots, M_{v_m}) \succeq 0$ and so $M_i \preceq I$.

If $k < i$ then

$$M_i \, \Phi_i^k = (I - \mathrm{diag}(M_{v_1}, \ldots, M_{v_m})) \times \mathrm{diag}(A_{\mathcal{C}(v_1)}, \ldots, A_{\mathcal{C}(v_m)}) \, \Phi_{i-1}^k = \Phi_i^k \tag{4.14}$$

since $M_{v_1} A_{\mathcal{C}(v_1)} = \cdots = M_{v_m} A_{\mathcal{C}(v_m)} = 0$ where $\mathcal{V}_{i-1} = \{v_1, \ldots, v_m\}$.

If $k \geq j \geq i$ then since $[M_i]_{vv} = 0$ for all v it follows from the definition of Φ_k^i as a product of block diagonal matrices that $\mathcal{B}_k^i(\Phi_k^i \, M_i \, \Phi_i^k) = 0$. Since $j \geq i$ implies that $\mathcal{B}_k^j = \mathcal{B}_k^j \mathcal{B}_k^i$ it is the case that

$$\mathcal{B}_k^j(\Phi_k^i \, M_i \, \Phi_i^k) = \mathcal{B}_k^j \mathcal{B}_t^i(\Phi_k^i \, M_i \, \Phi_i^k) = 0,$$

completing the proof of the claim. □

With the claim established we now construct a 'global' dual certificate by taking

$$Y = \sum_{i=1}^{n} (\lambda_{i-1} - \lambda_i) \, \Phi_n^i \, M_i \, \Phi_i^n \, .$$

We now show that Y satisfies the conditions of Proposition 9 with $L_0^\star = \Phi_n^0 \, R \, \Phi_0^n$ and $L_i^\star = \Phi_n^i \, Q_i \, \Phi_i^n$ for $i = 1, 2, \ldots, n$.

First we show that Y is feasible for the dual semidefinite program (4.9) by showing that property

1 in Proposition 9 holds.

$$\begin{aligned}
\mathcal{B}^i_n(Y) &= \sum_{j=1}^{n}(\lambda_{j-1} - \lambda_j)\mathcal{B}^i_n(\Phi^j_n\ M_j\ \Phi^n_j) \\
&= \sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)\mathcal{B}^i_n(\Phi^j_n\ M_j\ \Phi^n_j) \quad \text{by Claim 1} \\
&\overset{(a)}{\preceq} \sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)\mathcal{B}^i_n(\Phi^j_n\ I\ \Phi^n_j) \\
&\overset{(b)}{\preceq} \sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)I \\
&= \lambda_i I \quad \text{as the sum telescopes and } \lambda_n = 0.
\end{aligned}$$

The inequality marked (a) holds because $\Phi^j_n\ (\cdot)\ \Phi^n_j$ and $\mathcal{B}^i_n$ both preserve the positive semidefinite cone and $M_i \preceq I$. The inequality marked (b) holds because $\Phi^j_n\ \Phi^n_j \preceq I$, and so by Lemma 14, $\mathcal{B}^i_n(\Phi^j_n\ \Phi^n_j) \preceq \mathcal{B}^i_n(I) = I$. Note that we used the fact that $i \leq j$ implies that $\lambda_i \geq \lambda_j$ in this argument to ensure all of the terms in the sum $\sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)$ are non-negative.

We use a very similar argument to establish the complementary slackness conditions (property 2 of Prop. 9) We first consider the case $i = 0$. Then

$$\begin{aligned}
YL^\star_0 &= \sum_{j=1}^{n}(\lambda_{j-1} - \lambda_j)\ \Phi^j_n\ M_j\ \Phi^n_j\ \Phi^0_n\ R\ \Phi^n_0 \\
&\overset{(a)}{=} \sum_{j=1}^{n}(\lambda_{j-1} - \lambda_j)\ \Phi^j_n\ M_j\ \Phi^0_j\ R\ \Phi^n_0 \\
&\overset{(b)}{=} \sum_{j=1}^{n}(\lambda_{j-1} - \lambda_j)\ \Phi^j_n\ \Phi^0_j\ R\ \Phi^n_0 \\
&\overset{(c)}{=} L^\star_0 \sum_{j=1}^{n}(\lambda_{j-1} - \lambda_j) \\
&= \lambda_n L^\star_0 = 0 \quad \text{as the sum telescopes and } \lambda_n = 0
\end{aligned}$$

where equalities (a) and (c) are applications of property 3 of Lemma 14 and (b) follows from Claim 1. In the cases where $i = 1, 2, \ldots, n$ we use the additional fact that since $\mathcal{B}^i_n(L_i) = L_i$ it follows that $\mathcal{B}_n^i(X)L_i = \mathcal{B}_n^i(XL_i)$ for all symmetric $|\mathcal{V}_n| \times |\mathcal{V}_n|$ matrices X. Then

$$\begin{aligned}\mathcal{B}_n^i(Y)L_i^\star &= \sum_{j=1}^{i}(\lambda_{j-1} - \lambda_j)\mathcal{B}_n^i(\Phi_n^j\ M_j\ \Phi_j^n)L_i^\star + \sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)\mathcal{B}_n^i(\Phi_n^j\ M_j\ \Phi_j^n\ L_i^\star)\\ &\overset{(a)}{=} \sum_{j=i+1}^{n}(\lambda_{j-1} - \lambda_j)\mathcal{B}_n^i(\Phi_n^j\ M_j\ \Phi_j^n\ \Phi_n^i\ Q_i\ \Phi_i^n)\\ &= \lambda_i L_i^\star\end{aligned}$$

where the equality marked (a) follows from Claim 1, and the rest of the argument is exactly the same as that used in the case where $i = 0$. □

4.8.4 Proof of Theorem 18

In this section we prove the following result from which Theorem 18 directly follows by taking $n = |\mathcal{C}(v)|$, $U = \mathcal{R}(A_v)$, $M_v = \pi_{U^\perp}^T Y \pi_{U^\perp}$, and the partition of $|\mathcal{C}(v)|$ to be that induced by the children of v.

Proposition 10. *Suppose $\mathcal{P}$ is a partition of $\{1, 2, \ldots, n\}$. Given a subspace U of $\mathbb{R}^n$ of dimension $n-k$, there is a $k \times k$ positive semidefinite matrix Y such that*

$$[\pi_{U^\perp}^T Y \pi_{U^\perp}]_{\mathcal{I}} = I \quad \textit{for all}\ \mathcal{I} \in \mathcal{P}$$

as long as $[P_U]_{\mathcal{I}} \prec 1/3I$ for all $\mathcal{I} \in \mathcal{P}$.

Our proof of Proposition 10 follows closely the proof of Theorem 5 in Chapter 3. Before proceeding, we introduce some convenient notation. Let us fix a partition $\mathcal{P}$ of $\{1, 2, \ldots, n\}$ throughout this section and label its elements $\mathcal{I}_1, \ldots, \mathcal{I}_m$. For the purposes of this proof we let $\text{diag} : \mathcal{S}^n \to \mathcal{S}^{|\mathcal{I}_1|} \times \cdots \times \mathcal{S}^{|\mathcal{I}_m|}$ be defined by

$$\text{diag}(X) = (X_{\mathcal{I}_1}, \ldots, X_{\mathcal{I}_m}).$$

Let diag^* denote its adjoint, so that diag^*diag is the orthogonal projector onto the set of block diagonal matrices with support corresponding to the partition $\mathcal{P}$.

Define the cone $\mathcal{K} = \mathcal{S}_+^{|\mathcal{I}_1|} \times \cdots \times \mathcal{S}_+^{|\mathcal{I}_m|}$ and write $\prec_{\mathcal{K}}$ to indicate the order induced by the (interior of the) cone $\mathcal{K}$. We also use the norm $\|(X_1, \ldots, X_m)\|_{\mathcal{K}} = \max_{1\leq i\leq m} \|X_i\|$ throughout this section. Note that it follows from the Russo-Dye theorem [8] that if $\mathcal{B}$ preserves $\mathcal{K}$ then $\|\mathcal{B}\|_{\mathcal{K}\to\mathcal{K}} = \|\mathcal{B}(I, \ldots, I)\|_{\mathcal{K}}$.

Proof of Proposition 10. Suppose U has dimension $n-k$. Define a map $\mathcal{A} : \mathcal{S}^k \to \mathcal{S}^{|\mathcal{I}_1|} \times \cdots \times \mathcal{S}^{|\mathcal{I}_m|}$

by

$$\mathcal{A}(X) = \mathrm{diag}(\pi_{U^\perp}^T X \pi_{U^\perp}).$$

Then to show that there is a matrix Y such that $\mathrm{diag}(\pi_{U^\perp}^T Y \pi_{U^\perp}) = I$ and $Y \succeq 0$, it suffices to show that $\mathcal{A}^\dagger(I, \dots, I) \succeq 0$ and take $Y = \mathcal{A}^\dagger(I, \dots, I)$. Furthermore, since $\mathcal{A}^*$ maps $\mathcal{K}$ into $\mathcal{S}_+^n$, it suffices to show that $(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I) \in \mathcal{K}$.

Using the fact that $\mathrm{diag}(A\mathrm{diag}^*(B_1, \dots, B_m)) = ([A]_{\mathcal{I}_1} B_1, \dots, [A]_{\mathcal{I}_m} B_m)$, we can write

$$\begin{aligned}(\mathcal{A}\mathcal{A}^*)(X_1, \dots, X_m) &= \mathrm{diag}(P_{U^\perp}\mathrm{diag}^*(X_1, \dots, X_m)P_{U^\perp}) \\ &= \mathrm{diag}((I - P_U)\mathrm{diag}^*(X_1, \dots, X_m)(I - P_U)) \\ &= (X_1 - [P_U]_{\mathcal{I}_1} X_1 - X_1[P_U]_{\mathcal{I}_1}, \dots, X_m - [P_U]_{\mathcal{I}_m} X_m - X_m[P_U]_{\mathcal{I}_m}) \quad + \\ &\quad\ \mathrm{diag}(P_U \mathrm{diag}^*(X_1, \dots, X_m)P_U) \\ &= (L_1(X_1), \dots, L_m(X_m)) + \mathrm{diag}(P_U \mathrm{diag}^*(X_1, \dots, X_m)P_U)\end{aligned}$$

where $L_i(X_i) = ((1/2)I - [P_U]_{\mathcal{I}_i})X_i + X_i((1/2)I - [P_U]_{\mathcal{I}_i})$. Define the maps $\mathcal{L}(X_1, \dots, X_m) = (L_1(X_1), \dots, L_m(X_m))$ and $\mathcal{B}(X_1, \dots, X_m) = \mathrm{diag}(P_U\mathrm{diag}^*(X_1, \dots, X_m)P_U)$ so that $\mathcal{A}\mathcal{A}^* = \mathcal{L} + \mathcal{B}$. Since $[P_U]_{\mathcal{I}_i} \prec (1/3)I$ for all $i = 1, 2, \dots, m$, the matrices $(1/2)I - [P_U]_{\mathcal{I}_i}$ are positive definite, and so the inverses of the Lyapunov operators L_i^{-1} are positive maps, in the sense that they map positive semidefinite matrices to positive semidefinite matrices [8]. Furthermore, we have that $L_i(I) = I - 2[P_U]_{\mathcal{I}_i} \succ (1/3)I$ so that $L_i^{-1}(I) \prec L_i^{-1}(3L_i(I)) = 3I$.

Recall that our aim is to show that $(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I) \in \mathcal{K}$. Since each L_i^{-1} is a positive map it follows that $\mathcal{L}^{-1}(X_1, \dots, X_m) = (L_1^{-1}(X_1), \dots, L_m^{-1}(X_m))$ preserves the cone $\mathcal{K}$. It is easily checked that $\mathcal{B}$ also preserves $\mathcal{K}$. Furthermore, $\mathcal{B}(I, \dots, I) = ([P_U]_{\mathcal{I}_1}, \dots, [P_U]_{\mathcal{I}_m}) \prec_{\mathcal{K}} (1/3)(I, \dots, I)$ and $\mathcal{L}^{-1}(I, \dots, I) = (L_1^{-1}(I), \dots, L_m^{-1}(I)) \prec_{\mathcal{K}} 3(I, \dots, I)$.

We now expand $(\mathcal{A}\mathcal{A}^*)^{-1} = (\mathcal{L} + \mathcal{B})^{-1}$ as a Neumann series. This is valid as

$$\mathcal{L}^{-1}\mathcal{B}(I, \dots, I) = \mathcal{L}^{-1}([P_U]_{\mathcal{I}_1}, \dots, [P_U]_{\mathcal{I}_m}) \prec_{\mathcal{K}} (1/3)\mathcal{L}^{-1}(I, \dots, I) \prec_{\mathcal{K}} (I, \dots, I)$$

and so $\|\mathcal{L}^{-1}\mathcal{B}\|_{\mathcal{K}\to\mathcal{K}} = \|\mathcal{L}^{-1}\mathcal{B}(I, \dots, I)\|_{\mathcal{K}} < 1$. The expansion yields

$$(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I) = \mathcal{L}^{-1}[(I, \dots, I) - \mathcal{B}\mathcal{L}^{-1}(I, \dots, I)] + \mathcal{L}^{-1}\mathcal{B}\mathcal{L}^{-1}\mathcal{B}[(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I)]$$

so that

$$(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I) = \sum_{i=0}^{\infty}(\mathcal{L}^{-1}\mathcal{B}\mathcal{L}^{-1}\mathcal{B})^i\{\mathcal{L}^{-1}[(I, \dots, I) - \mathcal{B}\mathcal{L}^{-1}(I, \dots, I)]\}.$$

As such, to show that $(\mathcal{A}\mathcal{A}^*)^{-1}(I, \dots, I) \in \mathcal{K}$ it suffices to show that $\mathcal{L}^{-1}\mathcal{B}\mathcal{L}^{-1}\mathcal{B}$ preserves $\mathcal{K}$ and

$\mathcal{L}^{-1}[(I,\ldots,I) - \mathcal{B}\mathcal{L}^{-1}(I,\ldots,I)] \in \mathcal{K}$. Since $\mathcal{B}$ and $\mathcal{L}^{-1}$ preserve $\mathcal{K}$ it follows that $\mathcal{L}^{-1}\mathcal{B}\mathcal{L}^{-1}\mathcal{B}$ also has this property. Furthermore, $\mathcal{L}^{-1}[(I,\ldots,I) - \mathcal{B}\mathcal{L}^{-1}(I,\ldots,I)] \in \mathcal{K}$ because

$$\mathcal{B}\mathcal{L}^{-1}(I,\ldots,I) \prec_{\mathcal{K}} 3\mathcal{B}(I,\ldots,I) = 3([P_U]_{\mathcal{I}_1},\ldots,[P_U]_{\mathcal{I}_m}) \prec_{\mathcal{K}} (I,\ldots,I)$$

and $\mathcal{L}^{-1}$ preserves $\mathcal{K}$. This completes the proof. □

Chapter 5

Conclusion

5.1 Contributions

In this thesis we examined two subspace identification problems—factor analysis and its refinement, learning Gaussian latent tree models given the tree and the covariance among the leaves. Both of these problems are intractable to solve exactly in high dimensions. We provided a new analysis of a semidefinite programming-based heuristic, minimum trace factor analysis, for the factor analysis problem, and extended this convex optimization based method and its analysis to the Gaussian latent tree setting.

In our analysis of minimum trace factor analysis, we show that under simple incoherence conditions on the subspace we are trying to identify, minimum trace factor analysis can successfully identify that subspace. These conditions are sufficiently simple that they can easily be translated into problem-specific conditions when the subspaces that arise in a given problem have particular structure. As an example of this, we show how to convert our incoherence conditions into more problem-specific conditions when we use a factor analysis model in a subspace-based method for direction of arrival estimation. We also consider when minimum trace factor analysis succeeds on random problem instances, when the subspace to be identified is chosen uniformly from the set of r dimensional subspaces of $\mathbb{R}^n$. We show that for large n, with high probability minimum trace factor analysis succeeds for such problem instances as long as $r \leq n - cn^{5/6}$ for some constant c. This gives a precise sense in which minimum trace factor analysis is a good heuristic for factor analysis.

We then extend minimum trace factor analysis to the problem of learning the parameters and state dimensions of a Gaussian latent tree model given the index tree and the covariance among the leaves. We show that if the underlying tree model has all scalar states and the parameters satisfy certain 'balance' conditions then our semidefinite programming-based heuristic correctly identifies the model. In the case where the underlying tree model has non-scalar states, we develop incoherence-based conditions on the model parameters (closely related to our conditions for the success of minimum trace factor analysis) that ensure our semidefinite program successfully identifies the model parameters and state dimensions. We propose a modification of our method to deal with the case where the covariance matrix we are given does not arise as the covariance of some Gaussian latent tree model, and demonstrate by simulation that this method can identify the state dimensions of a true underlying model given only sample covariance information.

5.2 Further Work

In this final section we discuss some of the many research directions that arise from the work presented in this thesis.

5.2.1 Diagonal and Low-Rank Decompositions

The work in this thesis does not give a complete understanding of the largest rank of a low-rank matrix with random row/column space that can be recovered by minimum trace factor analysis. Simulations suggest that our result that matrices of rank $n - O(n^{5/6})$ can be recovered is clearly not optimal. Indeed numerical evidence suggests that the correct bound is of the form $n - O(\sqrt{n})$. This has the interesting interpretation in terms of ellipsoid fitting that we can fit an ellipsoid to $\sim k^2$ i.i.d. Gaussian points in $\mathbb{R}^k$.

All of our results only apply to the exact decomposition problem. In particular we always assume that the 'input' to our methods, Σ, admits an exact decomposition into a diagonal and a low-rank matrix. In practice we would not expect this assumption to hold. As such it is important to analyze a modification of minimum trace factor analysis (along the lines of the approximate covariance decomposition SDP in Chapter 4). Ideally we would seek structural stability results from such an analysis. By this we mean that if the input Σ is close (in some sense) to admitting a decomposition into a diagonal and a rank r matrix then the convex program decomposes Σ as the sum of a diagonal matrix, a rank r matrix, and a small error term. Such results have been derived for related problems [15] and we expect them also to hold in this case under appropriate tightenings of the conditions on the model required for exact diagonal and low-rank decomposition.

Recall that we can think of the diagonal and low-rank decomposition problem as an instance of low-rank matrix completion with a fixed pattern of unknown entries in the matrix. It would be interesting to extend our deterministic conditions for decomposition to other deterministic patterns (beyond diagonal and block diagonal) of unknown entries, such as banded matrices, or trees. We expect that such an analysis would also involve some of the combinatorial properties of the graph corresponding to the unknown entries.

5.2.2 Gaussian Latent Tree Models

As for diagonal and low-rank decompositions, our analysis only applies to the case where we are given a covariance matrix that arises as the covariance among the leaves of a Gaussian latent tree model. Extending our analysis to the case where we are only given a covariance matrix that approximates the covariance matrix among the leaves of the tree is a natural next step. A natural extension of this would be to assume we are only given certain projections of the covariance matrix among the leaves, or projections of the sample covariance among the leaves. Data of this type arises in problems in oceanography, for example, where the data available for some parts of the area being observed are at a much lower resolution than in other parts of the observed area.

In Chapter 4 we always assume that we are given the index tree with respect to which the Gaussian latent tree model is defined. While there are a number of methods in the phylogenetics and machine learning literature ([17,22,45] for example) for learning such a tree from leaf covariance information, it would be interesting if we could extend the convex optimization-based framework introduced in this thesis to also learn the tree structure.

It would also interesting to develop alternative convex optimization-based approaches to the problem of learning parameters and state dimensions of Gaussian latent tree models that are less global in nature but computationally very efficient. Methods that operate on a single parent-children cluster of nodes at a time would be much more computationally efficient and may offer similar performance in cases where the given tree matches the conditional independence structure in the data very well. Such approaches may, however, be quite sensitive to the choice of tree structure.

Bibliography

[1] H. Akaike, "Markovian representation of stochastic processes by canonical variables," *SIAM Journal on Control*, vol. 13, p. 162, 1975.

[2] A. Albert, "The matrices of factor analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 30, no. 4, p. 90, 1944.

[3] T. Anderson and H. Rubin, "Statistical inference in factor analysis," in *Proceedings of the third Berkeley Symposium on mathematical statistics and probability: contributions to the theory of statistics.* University of California Press, 1956, pp. 111–150.

[4] W. Barrett and S. Pierce, "Null spaces of correlation matrices," *Linear Algebra and its Applications*, vol. 368, pp. 129–157, 2003.

[5] M. Basseville, A. Benveniste, and A. Willsky, "Multiscale autoregressive processes. I. Schur-Levinson parametrizations," *IEEE Transactions on Signal Processing*, vol. 40, no. 8, pp. 1915–1934, 1992.

[6] S. Becker, E. Candes, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Arxiv preprint arXiv:1009.2065*, 2010.

[7] P. Bekker and J. ten Berge, "Generic global indentification in factor analysis," *Linear Algebra and its Applications*, vol. 264, pp. 255–263, 1997.

[8] R. Bhatia, *Positive definite matrices.* Princeton Univ Pr, 2007.

[9] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 31, no. 5, pp. 1235–1248, 1983.

[10] C. Borell, "The Brunn-Minkowski inequality in Gauss space," *Inventiones Mathematicae*, vol. 30, no. 2, pp. 207–216, 1975.

[11] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge University Press, 2004.

[12] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Arxiv preprint arXiv:0912.3599*, 2009.

[13] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[14] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[15] V. Chandrasekaran, P. Parrilo, and A. Willsky, "Latent Variable Graphical Model Selection via Convex Optimization," *Arxiv preprint arXiv:1008.1290*, 2010.

[16] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition," *Arxiv preprint arXiv:0906.2220*, 2009.

[17] M. Choi, V. Tan, A. Anandkumar, and A. Willsky, "Learning latent tree graphical models," *Arxiv preprint arXiv:1009.2722*, 2010.

[18] G. Della Riccia and A. Shapiro, "Minimum rank and minimum trace of covariance matrices," *Psychometrika*, vol. 47, no. 4, pp. 443–448, 1982.

[19] C. Delorme and S. Poljak, "Combinatorial properties and the complexity of a max-cut approximation," *European Journal of Combinatorics*, vol. 14, no. 4, pp. 313–333, 1993.

[20] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[21] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[22] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison, "Biological sequence analysis," 1998.

[23] N. El Karoui, "The spectrum of kernel random matrices," *The Annals of Statistics*, vol. 38, no. 1, pp. 1–50, 2010.

[24] J. Feldman, M. Wainwright, and D. Karger, "Using linear programming to decode binary linear codes," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 954–972, 2005.

[25] A. Frakt and A. Willsky, "Computationally efficient stochastic realization for internal multiscale autoregressive models," *Multidimensional Systems and Signal Processing*, vol. 12, no. 2, pp. 109–142, 2001.

[26] P. Frankl and H. Maehara, "Some geometric applications of the beta distribution," *Annals of the Institute of Statistical Mathematics*, vol. 42, no. 3, pp. 463–474, 1990.

[27] R. Frisch, *Statistical confluence analysis by means of complete regression systems.* Universitetets Økonomiske Instituut, 1934.

[28] Y. Gordon, "On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$," *Geometric Aspects of Functional Analysis*, pp. 84–106, 1988.

[29] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *Arxiv preprint arXiv:0910.1879*, 2009.

[30] R. Horn and C. Johnson, *Topics in matrix analysis.* Cambridge Univ Pr, 1994.

[31] W. Irving and A. Willsky, "A canonical correlations approach to multiscale stochastic realization," *IEEE Transactions on Automatic Control*, vol. 46, no. 10, pp. 1514–1528, 2001.

[32] R. Kalman, "Identification of noisy systems," *Russian Mathematical Surveys*, vol. 40, p. 25, 1985.

[33] A. Kannan, M. Ostendorf, W. Karl, D. Castanon, and R. Fish, "ML parameter estimation of a multiscale stochastic process using the EM algorithm," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1836–1840, 2000.

[34] R. Keshavan, A. Montanari, and S. Oh, "Matrix completion from a few entries," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.

[35] D. Koller and N. Friedman, *Probabilistic graphical models: Principles and techniques.* The MIT Press, 2009.

[36] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[37] W. Ledermann, "On the rank of the reduced correlational matrix in multiple-factor analysis," *Psychometrika*, vol. 2, no. 2, pp. 85–93, 1937.

[38] ——, "On a problem concerning matrices with variable diagonal elements," in *Proc. R. Soc. Edinb.*, vol. 60, 1940, pp. 1–17.

[39] M. Ledoux, *The concentration of measure phenomenon.* Amer Mathematical Society, 2001.

[40] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: isoperimetry and processes.* Springer, 1991.

[41] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *IEEE International Symposium on Computer Aided Control Systems Design.* IEEE, 2004, pp. 284–289.

[42] Z. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *Selected Areas in Communications, IEEE Journal on*, vol. 24, no. 8, pp. 1426–1438, 2006.

[43] M. Mesbahi and G. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 239–243, 1997.

[44] R. Oliveira, "Sums of random hermitian matrices and an inequality by rudelson," *Elect. Comm. Probab*, vol. 15, pp. 203–212, 2010.

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

[46] M. Ramana, L. Tunçel, and H. Wolkowicz, "Strong duality for semidefinite programming," *SIAM Journal on Optimization*, vol. 7, no. 3, pp. 641–662, 1997.

[47] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[48] J. Reid, *Linear system fundamentals: continuous and discrete, classic and modern.* McGraw-Hill, 1983.

[49] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[50] R. Schmidt, "A signal subspace approach to multiple emitter location spectral estimation," *Ph. D. Thesis, Stanford University*, 1981.

[51] A. Shapiro, "Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis," *Psychometrika*, vol. 47, no. 2, pp. 187–199, 1982.

[52] ——, "Identifiability of factor analysis: Some results and open problems," *Linear Algebra and its Applications*, vol. 70, pp. 1–7, 1985.

[53] C. Spearman, "'General Intelligence,' Objectively Determined and Measured," *The American Journal of Psychology*, pp. 201–292, 1904.

[54] L. Thurstone, "Multiple factor analysis." *Psychological Review*, vol. 38, no. 5, pp. 406–427, 1931.

[55] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[56] K. Toh, M. Todd, and R. Tutuncu, "SDPT 3—a MATLAB software package for semidefinite programming, version 1.3," *Optimization Methods and Software*, vol. 11, no. 1, pp. 545–581, 1999.

[57] T. Tuncer and B. Friedlander, "Classical and modern direction-of-arrival estimation," 2009.

[58] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.

[59] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *Arxiv preprint arXiv:1011.3027*, 2010.

[60] M. Wainwright and M. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.