Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Entropic methods in learning stochastic systems with latent variables and homogeneous Gaussian random fields

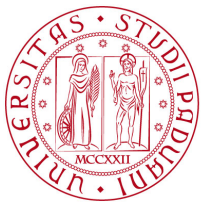
Ph.D. candidate
Valentina Ciccone

Advisor
Prof. Augusto Ferrante

Director & Coordinator
Prof. Andrea Neviani

Ph.D. School in
Information Engineering

Department of
Information Engineering
University of Padova
2019



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Head Office: Università degli Studi di Padova

Department of Information Engineering
Ph.D. course in: Information Engineering
Curriculum: Information and Communication Technologies
Series XXXII

Entropic methods in learning stochastic systems with latent variables and homogeneous Gaussian random fields

Coordinator: Prof. Andrea Neviani

Advisor: Prof. Augusto Ferrante

Ph.D. student: Valentina Ciccone

Abstract

This dissertation is divided into two main parts, the common thread being the prominent role of entropy-based methods in the identification and estimation of stochastic models and systems.
The first part of the dissertation deals with the problem of robustness in the identification of stochastic models with latent variables, namely variables that, although influencing the behaviour of some other manifest variables, are not directly observable. These models boast a long tradition and find natural application in many disciplines within engineering and applied science including psychology, econometrics, system engineering, machine learning and statistics, to name but a few. In this part of the dissertation, relying on certain invariance properties of the relative entropy and inspired by the previous contributions on robust estimation, we propose, for the case of zero-mean Gaussian random variables and processes, a novel approach for constructing a confidence region for the underlying model from a given finite sample estimate. This region depends only on the number of data and, by construction, contains the true model with a user-chosen probability. This paradigm is applied to the identification of two classes of latent variable models, namely factor models and graphical models with latent variables, for which we search the most parsimonious model in the confidence region by solving a convex optimization problem.
The second part of this dissertation focuses on homogeneous Gaussian random fields, namely stationary Gaussian processes defined over a multidimensional lattice, which find application, for instance, in multidimensional signal processing, spatial statistics and image analysis. In this part of the dissertation, relying on the properties of multi-level circulant and multi-level Toeplitz matrices, we derive an explicit formula for the computation of the relative entropy rate between two homogeneous random fields in terms of their spectral densities. Moreover, we establish a correspondence between the relative entropy rate for homogeneous Gaussian random fields and the relative entropy rate for their spectral domain representation. Both the cases of general and periodic homogeneous random fields are considered.

Sommario

La tesi è divisa in due parti il cui filo conduttore è il ruolo dei metodi entropici nell'identificazione e nella stima di modelli e sistemi stocastici.
La prima parte della tesi affronta il problema dell'identificazione robusta di modelli stocastici con variabili latenti, ossia variabili che, pur influenzando altre variabili manifeste, non sono direttamente osservabili. Questi modelli vantano una lunga tradizione e trovano applicazione in molte discipline dell'ingegneria e delle scienze applicate fra cui psicologia, econometria, ingegneria dei sistemi, machine learning e statistica, per citarne solo alcune. In questa parte della tesi, grazie a certe proprietà di invarianza dell'entropia relativa e ispirati da precedenti lavori nel campo della stima robusta, proponiamo, nel caso di variabili e processi stocastici Gaussiani a media nulla, un nuovo approccio per costruire una regione di confidenza per il modello sottostante data una sua stima campionaria. Questa regione di confidenza dipende solo dalla numerosità campionaria e, per costruzione, contiene il vero modello con una probabilità prescelta. Questo paradigma è applicato all'identificazione di due classi di modelli a fattori latenti, ossia i modelli fattoriali e i modelli grafici con variabili latenti, per i quali cerchiamo il modello più parsimonioso, rispetto alla data classe, nella regione di confidenza risolvendo un problema di ottimizzazione convessa.
La seconda parte della tesi si focalizza sui campi aleatori Gaussiani omogenei, ossia processi Gaussiani stazionari definiti su un reticolo multidimensionale, che trovano applicazione, per esempio, nell'analisi dei segnali multidimensionali, nella statistica spaziale e nell'analisi di immagini. In questa parte della tesi, grazie alle proprietà delle matrici multilivello circolanti e multilivello di Toeplitz, deriviamo una formula esplicita per il tasso di entropia relativa tra campi aleatori Gaussiani omogenei in termini delle loro densità spettrali. Inoltre stabiliamo una corrispondenza tra tasso di entropia relativa per campi aleatori Gaussiani omogenei e tasso di entropia relativa per i processi spettrali loro associati. Sono considerati sia il caso generale sia il caso di campi aleatori periodici.

Contents

List of symbols

Symbol	Description
$\mathbb{N}$	Set of natural numbers
$\mathbb{Z}$	Set of integer numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{C}$	Set of complex numbers
a^*	Complex conjugate of $a \in \mathbb{C}$
$\mathfrak{Re}[a]$	Real part of $a \in \mathbb{C}$
$\mathfrak{Im}[a]$	Imaginary part of $a \in \mathbb{C}$
$\lvert a \rvert$	Modulus of $a \in \mathbb{C}$
$\mathbb{R}^{n\times m}$	Set of $n \times m$ matrices with real entries
$\mathbb{C}^{n\times m}$	Set of $n \times m$ matrices with complex entries
$\ker(\cdot)$	Kernel of a matrix or of a linear operator
$\mathrm{range}(\cdot)$	Range of a matrix or of a linear operator
$\mathrm{rank}(\cdot)$	Rank of a matrix
$(\cdot)^\top$	Transpose of a matrix or vector
$(\cdot)^*$	Conjugate transpose of a matrix or vector, adjoint operator
$(\cdot)^{-1}$	Inverse of a square matrix
$\mathrm{tr}(\cdot)$	Trace of a square matrix
$\lvert\cdot\rvert$	Determinant of a matrix
$\sigma(\cdot)$	Spectrum of a matrix
$A_{(i,j)}$	Element of the matrix A in the i-th row and $j-$th column
I_n	Identity matrix of dimension $n \times n$ (the subscript is omitted if clear from the context)
$\langle\cdot,\cdot\rangle$	Frobenius inner product, $\langle A,B\rangle := \mathrm{tr}(A^\top B)$ for $A,B \in \mathbb{R}^{n\times n}$
$\lVert\cdot\rVert_F$	Frobenius norm induced by the Frobenius inner product
$\lVert\cdot\rVert_2$	Matrix spectral norm

Symbol	Description
$\mathbf{Q}_n$	Vector space of real symmetric matrices of size $n \times n$
$\mathbf{D}_n$	Vector space of real diagonal matrices of size $n \times n$
$\mathbf{D}_n^{\perp}$	Orthogonal complement of $\mathbf{D}_n$ in $\mathbf{Q}_n$ with respect to the inner product $\langle \cdot, \cdot \rangle$
$\mathbf{M}_{m,n}$	Vector spaces of matrices of the form $[Y_0\, Y_1\, ...\, Y_n]$, $Y_0 \in \mathbf{Q}_m$, $Y_1, ..., Y_n \in \mathbb{R}^{m \times m}$
$\mathbf{Q}_{m(n+1)}$	Vector space of block-matrices with $(n+1) \times (n+1)$ square blocks of dimension $m \times m$
$\mathscr{C}^{\mathbf{N}}$	Set of (real) multi-level circulant matrices of index $\mathbf{N} \in \mathbb{N}^d$
$\mathscr{T}^{\mathbf{N}}$	Set of (real) multi-level Toeplitz matrices of index $\mathbf{N} \in \mathbb{N}^d$
$\succeq$, $\succ$	Partial ordering induced by the cone of symmetric positive semi-definite matrices
$\mathrm{diag}(A)$	Operator mapping a matrix $A \in \mathbb{R}^{n \times n}$ into a $n-$dimensional vector containing the diagonal elements of A
$\mathrm{diag}(\mathbf{a})$	Operator mapping an $n-$dimensional vector $\mathbf{a} \in \mathbb{R}^n$ into a diagonal matrix $A \in \mathbf{D}_n$ such that $\mathrm{diag}(A) = \mathbf{a}$
$\mathrm{diag}^2(A)$	Operator mapping a square matrix A into a diagonal matrix of the same size having the same main diagonal
$\mathrm{ofd}(A)$	Operator orthogonally projecting a matrix $A \in \mathbf{Q}_n$ onto $\mathbf{D}_n^{\perp}$
$\otimes$	Kronecker product between two matrices
$\mathrm{vec}(\cdot)$	Vectorization of a matrix
$\mathscr{S}_m^+$	Space of $m \times m$ matrix-valued coercive and bounded spectral densities defined on $\{e^{i\theta};\, \theta \in [-\pi, \pi]\}$
$\mathscr{Q}_{m,n}$	Set of $m \times m$ Hermitian pseudo-polynomial matrices of order n
$\mathbb{E}\{\cdot\}$	Expected value of a random variable or vector
$\mathscr{N}(\mu, \Sigma)$	Normal distribution with mean μ and covariance Σ
$\lvert\mathcal{A}\rvert$	Cardinality of a set $\mathcal{A}$

1 Introduction

Stochastic modelling is nowadays of prime importance in describing many physical, social, economic, and technological phenomena. Stochastic models, in their more general definition, arise naturally in various disciplines within engineering and applied science, including control, telecommunications and networks, signal processing, biology and finance, to name but a few.
Entropic functionals and divergence criteria have a long-standing tradition in the identification and estimation of stochastic models and systems, a very popular choice being the *relative entropy* or *Kullback-Leibler divergence* which is nowadays an ubiquitous tool in system identifications, spectral estimation, statistics and machine learning.
The first part of this dissertation is concerned with a particular family of stochastic models in which some variables are not directly observable: they are the so-called *latent variable models*. The second part of this dissertation is oriented toward the problem of modelling stochastic systems that exhibit both time and spatial dynamics. This is achieved by considering *homogeneous random fields* and the related problem of their spectral estimation.
The *leitmotiv* across all the analysis in this dissertation is the prominent role of the relative entropy, and in particular of certain *invariance properties* of the relative entropy.

1.1 Latent variable modelling

A main objective when modelling big data is to provide a concise, parsimonious representation of the statistical dependencies among a collection of observable random variables. Latent variable models are a well-established and widely used tool to achieve this goal.

Latent variable models encompass a wide family of statistical models in which some variables, the so-called latent variables, are not directly observable but are instead inferred from a set of observed, or manifest, variables by means of a mathematical model.
Latent variable models boast a long tradition with their origin dating back at the beginning of the last century in the psychology community. Nowadays they find natural application in many disciplines including psychology, economics, demography, social science, medicine, system engineering, bioinformatics, machine learning, artificial intelligence, natural language processing and many others.
A main advantage of latent variable models is that they form a powerful tool for data dimensionality reduction. A key assumption in this family of models is that by accounting for the contribution of a few latent variables the statistical description of the observed variables significantly simplifies. More formally, in latent variable modelling we assume the existence of a small set of latent variables such that the conditional distribution of the observed variables given this set of latent variables has a much simpler structure. This is of paramount importance when dealing with data complexity as the simplified data structure provides a more easily interpretable relation between random variables.
We stress that latent variables are, in principle, a pure mathematical construction derived from observed phenomena and they may not have direct interpretability. The problem of attributing an interpretable meaning to latent variables is by itself an active vein of research with evident impact in many scientific and technological applications.
Latent variable models may be classified according to the continuous or categorical nature of the latent and of the observed variables, we refer to Skrondal and Rabe-Hesketh (2007) for a detailed survey. Depending on the type of model several inference methods are available, both frequentist and Bayesian. At a general level, frequentist approaches may include the classical maximum likelihood and Expectation Maximization (EM) see Dempster, Laird, and Rubin (1977) and the more recent composite likelihood and variational approximation Varin and Vidoni (2005); Blei, Kucukelbir, and McAuliffe (2017) while Bayesian approaches may include Monte Carlo Markov Chains (MCMC) and sequential Monte Carlo methods Doucet, De Freitas, and Gordon (2001).
The latent variable models that we shall consider in this dissertation are *factor analysis models*, or *factor models*, and *graphical models with latent variables*.

Factor models

Factor models aim to describe the variability among a large number of observed, correlated variables in terms of a smaller number of common relevant factors. Factor models have their origin in the seminal work Spearman (1904) in psychology. Since then they have been an active area of research with countless contributions in many disciplines including psychometric Burt (1909); Spearman and Holzinger (1924); Thurstone (1935); Bekker and de Leeuw (1987), econometrics Geweke (1977); Chamberlain and Rothschild (1983); Picci (1989); Forni and Lippi (2001); H.Stock and W.Watson (2010), statistics Anderson and Rubin (1956), system engineering Kalman (1983); Van Schuppen (1986) ;

Picci and Pinzoni (1986); Bottegal and Picci (2015), just to name a few. It can be seen that the problem of learning a factor model, in its original linear and static formulation, can be stated in mathematical terms as follows. A covariance matrix must be additively decomposed as the sum of two positive semi-definite matrices: a diagonal one, accounting for the idiosyncratic noise affecting the data, and a low rank one, whose rank must be as small as possible in order to describe the data in terms of the smallest possible number of latent factors.

Graphical models with latent variables

Gaussian Graphical models are statistical models describing a collection of jointly Gaussian random variables by means of an undirected graph in which each node corresponds to one of these variables and edges represent conditional dependence relations: absence of an edge between two nodes means that the corresponding random variables are conditionally independent given all the other, Lauritzen (1996). It can be shown that in these models the sparsity of the graph structure corresponds to the sparsity pattern of the concentration matrix, i.e. the inverse of the covariance matrix. When looking for a sparse graph representation a major challenge arises from the fact that some relevant variables, although directly influencing the manifest variables, may be unobservable causing the resulting graph to be dense or even complete. Latent variable graphical models aim to remedy this problem by considering the presence of a small number of latent variables, Chandrasekaran, Parrilo, and Willsky (2010); Chandrasekaran, Sanghavi, Parrilo, and Willsky (2011). Hence the problem of learning a latent variable Gaussian graphical model is the problem of learning a structured model in which the presence of a small number of latent variables is allowed in order to enforce the sparsity of the sub-graph corresponding to the observed variables. In mathematical terms, this amount to find a decomposition of the concentration matrix as the difference between a positive definite sparse matrix and a low-rank positive semi-definite matrix such that (a trade-off between) the number of non-zero elements of the former and the rank of the latter are minimized.

1.2 Preliminaries on random fields

The study of random fields is, in its most common understanding, the study of stochastic processes defined over some N-dimensional space.
Classical textbook examples of random fields include the modelling of the ocean surfaces, whose heights can be mathematically described by a three dimensional random field (where one dimension represents time while the other two are spatial dimensions), or the roughness of a metallic surface which can be described by a two dimensional random field, Adler (2010).

The theory of random fields is extensive, ranging from their finest probabilistic description, to the more modelling-concerned aspects, to the huge amount of applications across different disciplines for which they are naturally suited.

Among the many applications, besides the aforementioned ocean and metallic surface modelling, Longuet-Higgins (1952, 1957); Greenwood and Williamson (1966); Whitehouse and Archard (1970), we mention applications in forestry with the seminal work Matérn (2013) [1], geomorphology Mandelbrot (1975b), turbulence studying Mandelbrot (1975a), meteorology Handcock and Wallis (1994), agriculture production Besag (1974); Whittle (1954), computer vision where the use of Markov random fields and its variants abound Li (1994). More in general, random fields appear to be a natural tool for modelling any physical, social or natural phenomenon which exhibits random variations in time and space, see Vanmarcke (2010) and references therein.

Concerning the systems and signals engineering community, the interest in these processes is motivated by their application for instance in imaging processing Ringh, Karlsson, and Lindquist (2016); Ringh, Karlsson, and Lindquist (2015); Ekstrom (2012), texture generations Ringh, Karlsson, and Lindquist (2017); Kashyap and Lapsa (1984), parameter estimation in automotive radar applications and sensor arrays Lang and McClellan (1983); Zhu, Ferrante, Karlsson, and Zorzi (2019).

In this dissertation we focus on homogeneous discrete random fields, namely stationary stochastic processes defined over a multidimensional lattice. As it is the case for their one-dimensional counterpart, Cramér and Leadbetter (1967); Yaglom (2004), homogeneous random fields admit a spectral representation, Yaglom (1957); Yaglom et al. (1961); Miller (1975). For an overview of the spectral theory of homogeneous random fields we refer, for example, to the monograph (Adler, 2010, Chapter 2) and references therein.

It is worth noticing that, whereas in the last decades the problem of spectral estimation has received a great deal of attention in the one-dimensional setting, Byrnes, Georgiou, and Lindquist (2000); Georgiou and Lindquist (2003); Ferrante, Pavon, and Ramponi (2008); Georgiou, Karlsson, and Takyar (2009); Ramponi, Ferrante, and Pavon (2009); Jiang, Ning, and Georgiou (2012); Ferrante, Masiero, and Pavon (2012) ; Zorzi (2014, 2015), these topics have received less attention in the multidimensional setting. Among the notable exceptions we mention Lang and McClellan (1982a,b); Lev-Ari, Parker, and Kailath (1989); Georgiou (2006); Ringh et al. (2015); Ringh et al. (2016); Ringh et al. (2017).

Motivated by this, in the second part of this dissertation we have considered the problem of deriving a suitable entropic pseudo-distance between multidimensional spectral densities.

[1] Originally published in 1960.

1.3 Outline of the manuscript

This dissertation is divided into two parts, the *fil rouge* being the prominent role of the entropy-based methods in the identification and estimation of stochastic models and systems. Some basic properties of the relative entropy are recalled in Appendix A.

Part I: Learning latent variable models with relative entropy constraints

The first part of this dissertation proposes a robust estimation paradigm that hinges on a scale invariance property of the pseudo-distance induced by the relative entropy. This scale invariance property allows us to construct, for the case of zero-mean Gaussian random variables and processes, a *confidence region* centered in a given estimate of the second order statistics of the underlying data generating process. This estimate is based on a finite sample that may be understood as the observed data. The confidence region thus constructed contains the true model with a user chosen probability. Moreover, its "radius" depends *only* on the number of available data. This paradigm is applied to the identification of stochastic systems with latent variables for which we search the most parsimonious model – according to some ranking criteria – in the confidence region by solving a convex optimization problem. Two classes of latent variables models are considered, namely factor models and graphical models with latent variables.

Chapter 2 deals with the factor analysis problem in the realistic situation in which only a finite sample estimate of the covariance matrix is actually available. This has a strong motivation on the fact that even if the underlying data generating process is genuinely *low rank*, the minimum rank solution of the classical factor analysis problem rapidly degrades when a certain degree of uncertainty affects the covariance matrix. In this chapter we propose a strategy to cope with this problem by introducing an adequate confidence region for the covariance matrix. This leads to a novel optimization problem whose dual problem can be efficiently solved numerically by an ADMM-type algorithm. The results presented in this chapter are published in:

I. **Ciccone V., Ferrante A., Zorzi M.** Factor analysis with finite data, *In 56th IEEE Conference on Decision and Control (CDC)*, pages 4046–4051, 2017.

II. **Ciccone V., Ferrante A., Zorzi M.** Factor Models with Real Data: a Robust Estimation of the Number of Factors. *IEEE Transactions on Automatic Control* 64(6), 2412 - 2425, 2019.

In **Chapter 3** we consider the problem of robustly identifying a latent variable graphical model for a Gaussian random vector given a finite number of "observations" (i.e. independent realizations) of this vector. As in the case of factor analysis in fact, it appears

that the accuracy of the estimation may severely affect the goodness of the result, in terms of minimum rank and maximum sparsity, of the related optimization problem. By relying on the robust estimation paradigm proposed in Chapter 2 we propose a novel optimization problem for which we derive the dual problem which can be efficiently solved numerically. The results presented in this chapter are published in :

III. **Ciccone V., Ferrante A., Zorzi M.** Robust Identification of "Sparse Plus Low-rank" Graphical Models: An Optimization Approach. *In 57th IEEE Conference on Decision and Control (CDC)*, pages 2241–2246, 2018.

Chapter 4 generalizes the paradigm proposed in Chapter 3 to the dynamic case by addressing the problem of robustly identifying latent variable *dynamic* graphical models from a given a finite sample estimate of the spectral density of the underlying data generating process. In this chapter we deal with this problem by introducing an adequate confidence region for the spectral density. As a by-product, the resulting optimization problem involves just one regularization parameter balancing the trade-off between the number of latent variables and the sparsity of the learned graph. Consequently, the cross-validation procedure, required to solve the problem numerically, is performed on a one-dimensional grid significantly reducing the computational burden with respect to existing methods which typically require a two-dimensional grid. The problem considered in this chapter is stated in terms of spectral densities. To make the problem tractable we first derive a convenient matrix formulation. Then, the dual problem is derived and an ADMM-type algorithm is presented to solve the dual problem numerically. The results of this chapter appear in:

IV. **Ciccone V., Ferrante A., Zorzi M.** Learning Latent Variable Dynamic Graphical Models by Confidence Sets Selection. Submitted to *IEEE Transactions on Automatic Control*, 2018.

Part II: Entropic methods in learning homogeneous random fields

The second part of this dissertation focuses on homogeneous Gaussian random fields. A classical result in information theory shows that the relative entropy rate between two zero-mean stationary Gaussian processes can be computed explicitly in terms of their spectral densities, hence inducing a pseudo-distance in the cone of positive definite spectra. In **Chapter 5** we rely on the properties of *multi-level circulant* and *multi-level Toeplitz* matrices in order to establish a similar theory for homogeneous Gaussian random fields. Both the general case and the case of periodic Gaussian random fields are considered. Consequently, a natural entropic pseudo-distance on the cone of positive definite multidimensional spectral densities is obtained. Moreover, we define the *spectral*

entropy rate as an entropic divergence index between the spectral random fields associated to two homogeneous Gaussian random fields and we show that the relative entropy rate and the spectral relative entropy rate are in fact equal. The results in this chapter are the subject of:

V. **Ciccone V., Ferrante A.** Space and Spectral Domain Relative Entropy for Homogeneous Random Fields. Submitted to *Automatica*, 2019.

Part I

Learning latent variable models with relative entropy constraints

2

Minimum trace factor analysis with confidence constraints

2.1 Introduction

Describing a large amount of data by means of a small number of factors carrying most of the information is a main objective in modern data analysis with applications ranging in all fields of science. One of the classical methods for this purpose is to resort to factor models. Factor models were first developed at the beginning of the last century by Spearman, see Spearman (1904), in the framework of the so-called *mental tests* as an attempt at "the procedure of eliciting verifiable facts" in determining psychical tendencies from the tests results. From this first seed a rich stream of literature was developed at the interface between psychology and mathematics with the main focus on the case of a single common factor underlying the available data: necessary and sufficient conditions for the data to be compatible with a single common factor were derived in Burt (1909); Spearman and Holzinger (1924), see also Bekker and de Leeuw (1987) and references therein for a detailed historical reconstruction of the derivation of these conditions. The interest for this kind of model has grown rapidly also outside the psychology community and the analysis of factor models, or *factor analysis* has become an important tool in statistics, econometrics, systems theory and many other engineering fields Kalman (1983); Van Schuppen (1986); Bekker and de Leeuw (1987); Picci (1989); Geweke (1977); Picci and Pinzoni (1986); Pena and Box (1987); Hu and Chou (2004); Deistler and Zinner (2007); Heij, Scherrer, and Deistler (1997); Anderson and Deistler (1993); Sargent and Sims (1977); Forni and Lippi (2001); Engle and Watson

(1981); Watson and Engle (1983), McLachlan and Krishnan (1997); see also the more recent papers Bottegal and Picci (2015); Zorzi and Sepulchre (2015); Deistler, Scherer, and Anderson (2015); Fan, Liao, and Mincheva (2013), Bertsimas, Copenhaver, and Mazumder (2017); Delgado, Agüero, and Goodwin (2014) where many other references are listed. In particular we mention that a detailed geometric description of this problem can be found in Scherrer and Deistler (1998) whereas a maximum likelihood approach in a statistical testing framework is proposed in the seminal paper Anderson and Rubin (1956).

In the original formulation the construction of a factor model is equivalent to the mathematical problem of additively decomposing a given positive definite matrix Σ – modeling the covariance of the data – as

$$\Sigma = L + D \tag{2.1}$$

where both L and D are positive semi-definite, and D – modeling the covariance of the idiosyncratic noise – is diagonal. The rank of L is the number of (latent or hidden) common factors that explain the available data. One of the key aspects of factor analysis is to determine the minimum number of latent factors or, equivalently, a decomposition (2.1) where the rank of L is minimal. This is therefore a particular case of a matrix additive decomposition problem that arises naturally in numerous frameworks and have therefore received a great deal of attention, see Chandrasekaran et al. (2011); Agarwal, Negahban, and Wainwright (2012); Saunderson, Chandrasekaran, Parrilo, and Willsky (2012); Chandrasekaran et al. (2010) and references therein. We hasten to remark that the problem of minimizing the rank of L in the decomposition (2.1) is, in general, NP hard so that, the convex relaxation is usually considered where, in place of the rank, the nuclear norm (i.e. the trace) of L is minimized.

2.1.1 Motivating considerations

In Thurstone (1935), Kelley (1928), Ledermann (1937) an upper bound $r(n)$ – known as *Ledermann bound* – was proposed for the minimal rank $r_m(\Sigma)$ of L in terms if the dimension n of the matrix Σ:

$$r_m(\Sigma) \leq r(n) := \left\lfloor \frac{2n+1-\sqrt{8n+1}}{2} \right\rfloor.$$

This bound, however, is based on heuristics that have never been proven rigorously; a *pétale de rose* is the prize for a positive demonstration of this fact Hakim, Lochard, Olivier, and Terouanne (1976).[1] Interestingly, almost half a century later in Shapiro (1982) a related result was established: the set of symmetric $n \times n$ matrices Σ for which $r_m(\Sigma) < r(n)$ has zero Lebesgue measure. As a consequence of this result we have the

[1] Indeed, not only is a rigorous proof missing but a precise statement is also needed. In fact, some further assumptions must be added for the validity of this bound as counterexamples can, otherwise, be easily produced Guttman (1958).

following observation that may be regarded as the basic premise of our effort. When n is large the Ledermann bound $r(n)$ is not much smaller than n. Therefore, even if our data do come from a factor model with a small number r of latent factors, only a set of zero measure of $\hat{\Sigma}$ in a neighbourhood of Σ can be decomposed in such a way that the corresponding L matrix in its decomposition (2.1) has rank r. Thus, unless we know Σ with absolute precision, we cannot rely only on the decomposition (2.1) to recover such r. An example of this phenomenon is illustrated in Figure 2.1.

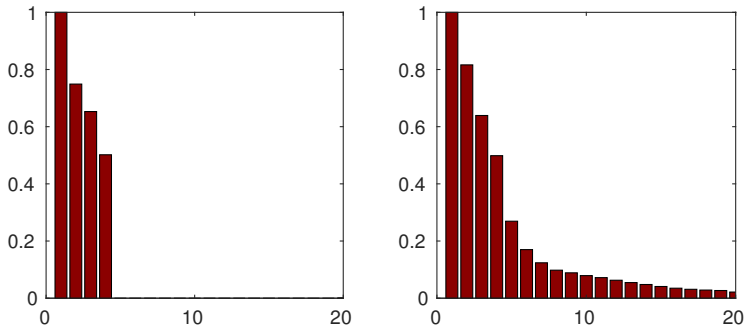


Figure 2.1: First 20 singular values of the matrices L (on the left) and $\hat{L}$ (on the right) obtained, respectively, by applying the minimum trace factor analysis decomposition algorithm to a "true" covariance matrix $\Sigma \in \mathbb{R}^{40\times 40}$ of a model with $r = 4$ latent factors and to an estimate $\hat{\Sigma}$ of Σ obtained by generating $N = 1000$ independent samples from a normal distribution $\mathcal{N}(0,\Sigma)$ and computing the corresponding sample covariance.

The problem of estimating r from an estimate $\hat{\Sigma}$ of Σ is therefore of crucial importance and has been addressed in Bai and Ng (2002) and Lam and Yao (2012) by means of statistical methods. A similar issue has been addressed also in Ning, Georgiou, Tannenbaum, and Boyd (2015) in the framework of the robustness of Frisch's scheme. Here, we propose an alternative optimization-based approach which is based only on the estimate $\hat{\Sigma}$ and takes into account the uncertainty of this estimate. Hence, even if we can start from N n-dimensional vectors (observations) the data of our problem are just the sample covariance $\hat{\Sigma}$ of these vectors and their number N. These two quantities summarize all the relevant information for our method in which we *compute* the matrix Σ in such a way that the trace of L in its additive decomposition (2.1) is minimized under a constraint limiting the Kullback-Leibler divergence between Σ and $\hat{\Sigma}$ to a prescribed tolerance that depends on the precision of our estimate $\hat{\Sigma}$ and hence may be reliably chosen on the basis of the data numerosity N.

The proposed problem is analyzed by resorting to duality theory. Indeed, the dual analysis yields a problem whose solution can be efficiently computed by employing an alternating direction method of multipliers (ADMM) algorithm.

Outline of the chapter

The chapter is outlined as follows. In the Section 2.2 we recall the classical approach to factor analysis and, from it, we derive the formulation of our factor analysis problem. In Section 2.3 we describe how to establish, for a desired tolerance, an upper bound on the aforementioned Kullback-Leibler divergence. In Section 2.4 we derive a dual formulation of our problem and we show existence and uniqueness of the solution for the dual problem. Then, in Section 2.5 we show how to recover the solution of the primal problem. In Section 2.6 we present the numerical algorithm for solving the dual problem, while in Section 2.7 the results of numerical simulations are presented. Finally, some conclusions and future research directions are provided.

2.2 Problem formulation

We consider a standard factor model in its static linear formulation

$$\mathbf{y} = A\mathbf{x} + \mathbf{z} \tag{2.2}$$

where $A \in \mathbb{R}^{n\times r}$, with $r \ll n$, is the factor loading matrix, $\mathbf{x}$ represents the (independent) latent factors and $\mathbf{z}$ is the idiosyncratic component. $\mathbf{x}$ and $\mathbf{z}$ are independent Gaussian random vectors with zero mean and covariance matrix equal to the identity matrix of dimension r and $D \in \mathbf{D}_n$, respectively. Note that $A\mathbf{x}$ represents the latent variable. Consequently, $\mathbf{y}$ is a Gaussian random vector with zero mean; we denote by Σ its covariance matrix. Since $\mathbf{x}$ and $\mathbf{z}$ are independent we get that Σ may be additively decomposed as in (2.1), where $L := AA^\top$ and D are the covariance matrices of $A\mathbf{x}$ and $\mathbf{z}$, respectively. Thus, L has rank equal to r, and D is diagonal.

The objective of factor analysis consists in finding the most parsimonious "low-rank plus diagonal" decomposition of Σ, that is a decomposition (2.1) for which the rank of L is minimal. This amounts to solve the minimum rank problem

$$\begin{aligned} \underset{L,D\in\mathbf{Q}_n}{\arg\min} \quad & \operatorname{rank}(L) \\ \text{subject to} \quad & L, D \succeq 0, \\ & D \in \mathbf{D}_n, \\ & \Sigma = L + D \end{aligned} \tag{2.3}$$

which is, in general, an NP hard problem, Fazel (2002). A well-known and widely used

heuristic is the convex relaxation of (2.3), i.e. the trace minimization problem

$$\begin{aligned} \underset{L,D\in\mathbf{Q}_n}{\arg\min} \quad & \operatorname{tr}(L) \\ \text{subject to} \quad & L, D \succeq 0, \\ & D \in \mathbf{D}_n, \\ & \Sigma = L + D. \end{aligned} \tag{2.4}$$

The substitution of the rank with the trace is justified by the fact that $\operatorname{tr}(L)$, i.e. the nuclear norm of L, is the convex hull of $\operatorname{rank}(L)$ over the set $\mathscr{S} := \{L \in \mathbf{Q}_n \text{ s.t. } \|L\|_2 \leq 1\}$, Fazel (2002). The relation between problem (2.3) and problem (2.4) has been first studied in Della Riccia and Shapiro (1982) and while these two problems are, in general, not equivalent, very often they have the same solution.

In practice, however, the matrix Σ is not known and needs to be estimated from a N-length realization (i.e. a data record) $\mathbf{y}_1 \dots \mathbf{y}_N$ of $\mathbf{y}$. The typical choice is to take the sample covariance estimate

$$\hat{\Sigma} := \frac{1}{N} \sum_{k=1}^{N} \mathbf{y}_k \mathbf{y}_k^\top .$$

As discussed in the introduction to this chapter, by replacing Σ with $\hat{\Sigma}$ the solution, in terms of minimum rank, will rapidly degrade. Indeed a delicate problem in factor analysis is the one of estimating the number of factors. Such a problem has been addressed by several important contributions, see the seminal works Bai and Ng (2002) and Lam and Yao (2012) and the references therein. Our objective is to address the same problem from a different perspective. In fact, we propose an optimization problem whose solution provides an estimate of the minimum number of factors by introducing an appropriate model for the error in the estimation of Σ. This model is based on an auxiliary Gaussian random vector $\hat{\mathbf{y}}$ with zero mean and covariance matrix $\hat{\Sigma}$ that is regarded as a "model approximation" for $\mathbf{y}$. To account for the estimation uncertainty, we assume that the distribution of $\mathbf{y}$ (that is completely specified by its covariance matrix and hence is referred to by Σ) belongs to a "ball" centered in $\hat{\mathbf{y}}$

$$\mathscr{B} := \{\Sigma \in \mathbf{Q}_n \text{ s.t. } \Sigma \succ 0,\ \mathbb{D}(\Sigma \| \hat{\Sigma}) \leq \delta/2\}$$

which is formed by placing a bound (i.e. tolerance) on the *relative entropy*, or *Kullback-Leibler* divergence, between $\mathbf{y}$ and $\hat{\mathbf{y}}$:

$$\mathbb{D}(\Sigma \| \hat{\Sigma}) := \frac{1}{2} \left(-\log|\Sigma| + \log|\hat{\Sigma}| + \operatorname{tr}(\Sigma \hat{\Sigma}^{-1}) - n \right).$$

This way to deal with model uncertainty has been successfully applied in econometrics for model misspecification Hansen and Sargent (2008) and in robust filtering Levy and Nikoukhah (2013, 2004); Zorzi and Levy (2015), Zorzi (2017a,b,c). Accordingly, in order

to estimate the minimum number of factors, we propose the following "robustification" of the minimum trace problem:

$$\begin{aligned} \underset{\Sigma,L,D\in\mathbf{Q}_n}{\arg\min} \quad & \operatorname{tr}(L) \\ \text{subject to} \quad & L,D \succeq 0, \\ & D \in \mathbf{D}_n, \\ & \Sigma = L+D, \\ & \Sigma \in \mathscr{B}. \end{aligned} \tag{2.5}$$

Note that in (2.5) we can eliminate variable D, obtaining the equivalent problem

$$\begin{aligned} \underset{\Sigma,L\in\mathbf{Q}_n}{\arg\min} \quad & \operatorname{tr}(L) \\ \text{subject to} \quad & L,\Sigma-L \succeq 0, \\ & \operatorname{ofd}(\Sigma-L) = 0, \\ & \Sigma \succ 0, \\ & 2\mathbb{D}(\Sigma||\hat{\Sigma}) \leq \delta. \end{aligned} \tag{2.6}$$

It is worth noting that an alternative in the same spirit of problem (2.6) is to consider $\mathbb{D}(\Sigma||\hat{\Sigma})$ as a penalty term in the objective function rather than as a constraint. Such approach, however, would require a cross validation procedure to set the regularization parameter λ, i.e. it would require to solve an optimization problem for many values of λ. In contrast, the proposed problem is solved only once provided that δ is chosen in a suitable way, as explained next.

2.3 The choice of δ

The tolerance δ may be chosen by taking into account the accuracy of the estimate $\hat{\Sigma}$ of Σ which, in turn, depends on the numerosity of the available data. This can be done by choosing a probability $\alpha \in (0,1)$ and a neighborhood of "radius" δ_α (in the Kullback-Leibler topology) centered in $\hat{\Sigma}$ containing the "true" Σ with probability α. The Kullback-Leibler divergence in (2.6) is a function of the estimated sample covariance and as such its accuracy depends crucially on the numerosity of the available data. To assess this accuracy we propose an approach that hinges on the following scale-invariance property of the Kullback-Leibler divergence.

Lemma 2.3.1. *Let* $\mathbf{y}_i \sim \mathscr{N}(0,\Sigma)$, $i = 1,\ldots,N$ *be i.i.d.* random variables *taking values in*

$\mathbb{R}^n$ and define the sample covariance estimator

$$\hat{\hat{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^\top .$$

Then, the Kullback-Leibler divergence between Σ and $\hat{\hat{\Sigma}}$ is a random variable whose distribution depends only on the sample size N and on the dimension n of the random variables.

Proof. We have

$$\hat{\hat{\Sigma}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^\top = \frac{1}{N}\Sigma^{1/2}\sum_{i=1}^{N} \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^\top \Sigma^{1/2} = \Sigma^{1/2} Q_N \Sigma^{1/2}$$

with $\tilde{\mathbf{y}}_i = \Sigma^{-1/2}\mathbf{y}_i \sim \mathcal{N}(0, I_n)$ and $Q_N := \frac{1}{N}\sum_{i=1}^{N} \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^\top$, is a random matrix taking values in $\mathbf{Q}_n$. Notice that at this point $\tilde{\mathbf{y}}$ are normalized Gaussian random vectors and hence do not depend on the data nor on Σ. Thus, Q_N is a random matrix whose distribution only depends on N and n (see Section 2.3.1 for more details). Hence, the Kullback-Leibler divergence between Σ and the sample covariance estimator is

$$\begin{aligned} d :&= \mathbb{D}(\Sigma \| \hat{\hat{\Sigma}}) = \frac{1}{2}\Big(\log(|\hat{\hat{\Sigma}}\Sigma^{-1}|) + \operatorname{tr}(\Sigma\hat{\hat{\Sigma}}^{-1}) - n\Big) \\ &= \frac{1}{2}\Big(\log(|Q_N|) + \operatorname{tr}(Q_N^{-1}) - n\Big). \end{aligned} \tag{2.7}$$

■

In view of this result we can easily approximate the distribution of the random variable $2d = 2\mathbb{D}(\Sigma\|\hat{\hat{\Sigma}})$ by a standard Monte Carlo method. In particular, we can reliably estimate with arbitrary precision the value of δ for which $Pr\big(2\mathbb{D}(\Sigma||\hat{\Sigma}) \leq \delta\big) = \alpha$. As an alternative to this empiric approach for determining δ_α, we can also resort to an analytic one as discussed below.

2.3.1 Gaussian Orthogonal Ensemble

Let us focus on the random matrix Q_N that we have defined as $Q_N := \frac{1}{N}\sum_{i=1}^{N} \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^\top$ where $\tilde{\mathbf{y}}_i \in \mathbb{R}^n$, $i = 1, ..., N$, are i.i.d. random variables distributed as $\mathcal{N}(0, I_n)$. We now introduce a new matrix $\tilde{Q}_N := \sqrt{N}(Q_N - I_n) = \sqrt{N}\frac{1}{N}\sum_{i=1}^{N} C_i$, where $C_i := \tilde{\mathbf{y}}_i \tilde{\mathbf{y}}_i^T - I$ are i.i.d. symmetric random matrices with zero mean. It is immediate to check that, for each i, any two distinct elements $[C_i]_{(h,j)}$ and $[C_i]_{(k,l)}$ of C_i are uncorrelated as long as they do not occupy symmetric positions, i.e. whenever $(h, j) \neq (l, k)$, and

$\text{Var}\left[[C_i]_{(h,j)}\right] = \begin{cases} 1, & \text{if } h \neq j \\ 2, & \text{if } h = j \end{cases}$. By the multivariate Central Limit Theorem, we have that $\tilde{Q}_N$ converges in distribution to the random matrix

$$X = \begin{pmatrix} \sqrt{2}\xi_{1,1} & \cdots & \cdots & \xi_{1,n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \xi_{1,n} & \cdots & \cdots & \sqrt{2}\xi_{n,n} \end{pmatrix} \in \mathbf{Q}_n,$$

where $\{\xi_{i,j}\}$ are i.i.d. Gaussian random variables with mean 0 and variance 1. The set of these matrices is known as the *Gaussian Orthogonal Ensemble*, see Anderson, Guionnet, and Zeitouni (2010). It is well known that the joint distribution of the eigenvalues $\lambda_1(X) \leq ... \leq \lambda_n(X)$ of such matrices takes the following form:

$$p(\lambda_1, ..., \lambda_n) = \bar{C}_n |\Delta(\lambda)| \prod_{i=1}^{n} \mathrm{e}^{-\lambda_i^2/4}$$

where $\lambda := (\lambda_1, ..., \lambda_n)$, $|\Delta(\lambda)|$ is the Vandermonde determinant associated with λ, which is given by:

$$|\Delta(\lambda)| = \prod_{i<j} (\lambda_j - \lambda_i)$$

and $\bar{C}_n$ is defined as:

$$\bar{C}_n = \left(\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} |\Delta(\lambda)| \prod_{i=1}^{n} \mathrm{e}^{-\lambda_i^2/4} d\lambda_i \right)^{-1}.$$

It is not difficult to see that (2.7) can be rewritten as:

$$d = d(\lambda_1, ..., \lambda_n) = \sum_{i=1}^{n} \frac{1}{2} \left(\log \left(\frac{\lambda_i}{\sqrt{N}} + 1 \right) - \frac{\lambda_i}{\lambda_i + \sqrt{N}} \right)$$

where $\lambda_i \in \sigma(\tilde{Q}_N)$. Then, for a desired α, we are interested in finding δ_α such that

$$Pr(2d \leq \delta_\alpha) = \alpha.$$

Such a value for δ_α is given by the cumulative distribution function $F(\cdot)$:

$$F(\delta_\alpha) = Pr(2d \leq \delta_\alpha) = \int_{I(\delta_\alpha)} 2d(\lambda_1, ..., \lambda_n) p(\lambda_1, ..., \lambda_n) d\lambda \tag{2.8}$$

where $p(\lambda_1,...,\lambda_n)$ denotes the joint probability density function of the eigenvalues $\lambda_1,...,\lambda_n$ and $I(\delta_\alpha) := \{(\lambda_1,...,\lambda_n) : d(\lambda_1,...,\lambda_n) \leq \delta_\alpha/2\}$. Given α the integral in (2.8) can be solved numerically for δ_α.

2.3.2 An upper bound for δ_α

If the chosen level α is too large with respect to the sample size N, the computed δ_α become excessively large so that there are diagonal matrices Σ_D such that $2\mathbb{D}(\Sigma_D||\hat{\Sigma}) \leq \delta_\alpha$. In this case problem (2.6) admits the trivial solution $L = 0$ and $D = \Sigma_D$. In order to rule out this trivial situation we need to require that the maximum value for δ in (2.6) is strictly less than a certain δ_{max} that can be determined as follows: since the trivial solution $L = 0$ would imply a diagonal Σ, that is $\Sigma = \Sigma_D := \text{diag}(d_1,...,d_n) > 0$, δ_{max} can be determined by solving the following minimization problem

$$\delta_{max} := \min_{\Sigma_D \in \mathbf{D}_n} 2\mathbb{D}(\Sigma_D||\hat{\Sigma}). \tag{2.9}$$

The following proposition shows how to solve this problem.

Proposition 2.3.2. *Let γ_i denote the i-th element in the main diagonal of the inverse of the sample covariance $\hat{\Sigma}^{-1}$. Then, the optimal Σ_D which solves the minimization problem in (2.9) is given by*

$$\Sigma_D = diag(\gamma_1^{-1},...,\gamma_n^{-1}).$$

Moreover, δ_{max} can be determined as

$$\delta_{max} = 2\mathbb{D}(\Sigma_D||\hat{\Sigma}) = \log|\,\text{diag}^2(\hat{\Sigma}^{-1})\hat{\Sigma}|. \tag{2.10}$$

Proof. For $\Sigma_D \in \mathbf{D}_n$, the Kullback-Leibler divergence can be rewritten as

$$\begin{aligned} 2\mathbb{D}(\Sigma_D||\hat{\Sigma}) &= \sum_{j=1}^{n} -\log d_j + \text{tr}(\text{diag}(d_1,...,d_n)\hat{\Sigma}^{-1}) + \log|\hat{\Sigma}| - n \\ &= \left(\sum_{j=1}^{n} -\log d_j + d_j\gamma_j\right) + \log|\hat{\Sigma}| - n. \end{aligned}$$

Thus, the minimization problem in (2.9) is equivalent to

$$\left(\sum_{j=1}^{n} \min_{d_j} -\log d_j + d_j\gamma_j\right) + \log|\hat{\Sigma}| - n.$$

Since $-\log d_j + d_j\gamma_j$ is convex with respect to d_j, by setting to zero the first derivative

with respect to $d_j, j = 1,...,n$, it easily follows that

$$\Sigma_D^{OPT} = \mathrm{diag}(\gamma_1^{-1},...,\gamma_n^{-1}).$$

Then, δ_{max} can be determined as

$$\begin{aligned}\delta_{max} &= 2\mathbb{D}(\Sigma_D^{OPT}\|\hat{\Sigma}) \\ &= \left(\sum_{j=1}^{n} -\log(\gamma_j^{-1}) + 1\right) + \log|\hat{\Sigma}| - n \\ &= -\log|\mathrm{diag}(\gamma_1^{-1},...,\gamma_n^{-1})| + \log|\hat{\Sigma}| \\ &= \log|\,\mathrm{diag}^2(\hat{\Sigma}^{-1})\hat{\Sigma})|.\end{aligned}$$

■

In what follows, we always assume that δ in (2.6) strictly less than δ_{max}, so that the trivial solution $L = 0$ is ruled out.

2.4 The dual problem

To to derive the dual of problem (2.6) we introduce the Lagrangian function

$$\begin{aligned}\mathscr{L}(L,\Sigma,\lambda,\Lambda,\Gamma,\Theta) =& \mathrm{tr}(L) + \lambda(-\log|\Sigma| + \log|\hat{\Sigma}| - n + \mathrm{tr}(\hat{\Sigma}^{-1}\Sigma) - \delta) \\ &- \mathrm{tr}(\Lambda L) - \mathrm{tr}(\Gamma(\Sigma - L)) + \mathrm{tr}(\Theta\,\mathrm{ofd}(\Sigma - L)) \\ =& \mathrm{tr}(L) + \lambda(-\log|\Sigma| + \log|\hat{\Sigma}| - n + \mathrm{tr}(\hat{\Sigma}^{-1}\Sigma) - \delta) \\ &- \mathrm{tr}(\Lambda L) - \mathrm{tr}(\Gamma(\Sigma - L)) + \mathrm{tr}(\mathrm{ofd}^*(\Theta)(\Sigma - L)) \\ =& \mathrm{tr}(L) + \lambda(-\log|\Sigma| + \log|\hat{\Sigma}| - n + \mathrm{tr}(\hat{\Sigma}^{-1}\Sigma) - \delta) \\ &- \mathrm{tr}(\Lambda L) - \mathrm{tr}(\Gamma(\Sigma - L)) + \mathrm{tr}(\mathrm{ofd}(\Theta)(\Sigma - L))\end{aligned} \tag{2.11}$$

where $\lambda \in \mathbb{R}, \lambda \geq 0$, and $\Lambda, \Gamma, \Theta \in \mathbf{Q}_n$ with $\Lambda, \Gamma \succeq 0$. In the last equality we exploited the fact that the operator $\mathrm{ofd}(\cdot)$ is self-adjoint. Note that the Lagrangian (2.11) does not include the constraint $\Sigma \succ 0$: as we will see this condition is automatically met by the solution of the dual problem.

The dual function is defined as the infimum of $\mathscr{L}(L,\Sigma,\lambda,\Lambda,\Gamma,\Theta)$ over L and Σ.
Thanks to the convexity of the Lagrangian, we rely on standard variational methods to characterize the minimum.
The first variation of the Lagrangian (2.11) at Σ in direction $\delta\Sigma \in \mathbf{Q}_n$ is

$$\delta\mathscr{L}(\Sigma;\delta\Sigma) = \mathrm{tr}(-\lambda\Sigma^{-1}\delta\Sigma + \lambda\hat{\Sigma}^{-1}\delta\Sigma - \Gamma\delta\Sigma + \mathrm{ofd}(\Theta)\delta\Sigma).$$

We impose the optimality condition

$$\delta\mathscr{L}(\Sigma;\delta\Sigma)=0, \qquad \forall\delta\Sigma\in\mathbf{Q}_n,$$

which is equivalent to require $\operatorname{tr}(-\lambda\Sigma^{-1}\delta\Sigma+\lambda\hat{\Sigma}^{-1}\delta\Sigma-\Gamma\delta\Sigma+\operatorname{ofd}(\Theta)\delta\Sigma)=0$ for all $\delta\Sigma\in\mathbf{Q}_n$, obtaining

$$\Sigma=\lambda(\lambda\hat{\Sigma}^{-1}-\Gamma+\operatorname{ofd}(\Theta))^{-1} \tag{2.12}$$

provided that $\lambda\hat{\Sigma}^{-1}-\Gamma+\operatorname{ofd}(\Theta)\succ 0$ and $\lambda>0$, which is clearly equivalent to require that the optimal Σ that minimizes the Lagrangian satisfies the constraint $\Sigma\succ 0$.

The first variation of the Lagrangian (2.11) at L in direction $\delta L\in\mathbf{Q}_n$ is

$$\delta\mathscr{L}(L;\delta L)=\operatorname{tr}(\delta L-\Lambda\delta L+\Gamma\delta L-\operatorname{ofd}(\Theta)\delta L).$$

Again, we impose the optimality condition

$$\delta\mathscr{L}(L;\delta L)=0, \qquad \forall\delta L\in\mathbf{Q}_n,$$

which is equivalent to require $\operatorname{tr}(\delta L-\Lambda\delta L+\Gamma\delta L-\operatorname{ofd}(\Theta)\delta L)=0$ for all $\delta L\in\mathbf{Q}_n$ and we get that

$$I-\Lambda+\Gamma-\operatorname{ofd}(\Theta)=0. \tag{2.13}$$

The following result provides a precise formulation of the dual problem.

Proposition 2.4.1. *The dual of problem (2.6) is*

$$\max_{(\lambda,\Gamma,\Theta)\in\mathscr{C}}\ \lambda(\log|(\hat{\Sigma}^{-1}+\lambda^{-1}(\operatorname{ofd}(\Theta)-\Gamma))|+\log|\hat{\Sigma}|-\delta) \tag{2.14}$$

where $\mathscr{C}$ is defined as

$$\mathscr{C}:=\{(\lambda,\Gamma,\Theta):\ \lambda>0,\ I+\Gamma-\operatorname{ofd}(\Theta)\succeq 0,\ \Gamma\succeq 0,\hat{\Sigma}^{-1}+\lambda^{-1}(\operatorname{ofd}(\Theta)-\Gamma)\succ 0\}.$$

Proof. The result immediately follows by substituting the optimal conditions (2.13) and (2.12) into (2.11) and carrying out the following straightforward computation where, for notation convenience, we set $\Delta:=\lambda(\lambda\hat{\Sigma}^{-1}+\operatorname{ofd}(\Theta)-\Gamma)^{-1}$:

$$\begin{aligned}
&\operatorname{tr}(L)+\lambda(-\log|\Delta|+\operatorname{tr}(\hat{\Sigma}^{-1}\Delta)+\log|\hat{\Sigma}|-n-\delta)\\
&\qquad-\operatorname{tr}((I+\Gamma-\operatorname{ofd}(\Theta))L)-\operatorname{tr}(\Gamma\Delta)+\operatorname{tr}(\operatorname{ofd}(\Theta)\Delta)+\operatorname{tr}(\Gamma L)-\operatorname{tr}(\operatorname{ofd}(\Theta)L)\\
&=\lambda(-\log|\Delta|+\log|\hat{\Sigma}|-n-\delta+\operatorname{tr}(\hat{\Sigma}^{-1}\Delta))-\operatorname{tr}(\Gamma\Delta)+\operatorname{tr}(\operatorname{ofd}(\Theta)\Delta)\\
&=\lambda(-\log|\Delta|+\log|\hat{\Sigma}|-n-\delta)+\operatorname{tr}(\lambda\hat{\Sigma}^{-1}\Delta)-\operatorname{tr}(\Gamma\Delta)+\operatorname{tr}(\operatorname{ofd}(\Theta)\Delta)\\
&=\lambda(-\log|\Delta|+\log|\hat{\Sigma}|-\delta).
\end{aligned}$$

Since the last term does not depend on Λ, we can get rid of it and, in view of (2.13), condition $\Lambda \succeq 0$ is replaced by the constraint $I+\Gamma-\text{ofd}(\Theta) \succeq 0$. ■

For convenience, we reformulate the maximization problem in (2.14) as a minimization problem:

$$\min_{(\lambda,\Gamma,\Theta)\in\mathscr{C}} J(\lambda,\Gamma,\Theta) \tag{2.15}$$

where $J(\lambda,\Gamma,\Theta) := \lambda(-\log|\hat{\Sigma}^{-1}+\lambda^{-1}(\text{ofd}(\Theta)-\Gamma)|-\log|\hat{\Sigma}|+\delta)$.

2.4.1 Existence of solutions

The aim of this subsection is to show that the dual problem (2.15) admits solution. This is done by showing that we can restrict the set $\mathscr{C}$ to a smaller compact set $\mathscr{C}_C$ over which the minimization problem is equivalent to the one in (2.15). Then, since the objective function is continuous over $\mathscr{C}$, and hence over $\mathscr{C}_C$, by Weierstrass's theorem we will conclude that J admits a minimum.
First, we recall that the operator $\text{ofd}(\cdot)$ is self-adjoint. Moreover, we notice that $\text{ofd}(\cdot)$ is not injective on Θ, thus we can restrict the domain of $\text{ofd}(\cdot)$ to those Θ such that $\text{ofd}(\cdot)$ is injective. Since ofd is self-adjoint we have that:

$$\ker(\text{ofd}) = [\text{range }(\text{ofd})]^{\perp}.$$

Thus, by restricting Θ to range(ofd)$= [\ker(\text{ofd})]^{\perp} = \mathbf{D}_n^{\perp}$, the map becomes injective. Therefore, without loss of generality, from now on we can safely assume that $\Theta \in \mathbf{D}_n^{\perp}$ so that $\text{ofd}(\Theta)=\Theta$ and we restrict our set $\mathscr{C}$ to $\mathscr{C}_1$:

$$\begin{aligned}\mathscr{C}_1 :=& \{(\lambda,\Gamma,\Theta)\in\mathscr{C} : \Theta\in\mathbf{D}_n^{\perp}\}\\ =& \{(\lambda,\Gamma,\Theta) : \lambda>0, I+\Gamma-\Theta\succeq 0, \Gamma\succeq 0, \Theta\in\mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1}+\lambda^{-1}(\Theta-\Gamma)\succ 0\}.\end{aligned}$$

Moreover, since Θ and Γ enter into the problem always through their difference they cannot be determined univocally. However, their difference can. This allows us to restrict Γ to the space of the diagonal positive semi-definite matrices. Indeed, for any sequence $(\lambda_k,\Gamma_k,\Theta_k)_{k\in\mathbb{N}} \in \mathscr{C}_1$ such that $\inf J(\lambda,\Gamma,\Theta) = \lim_{k\to\infty} J(\lambda_k,\Gamma_k,\Theta_k)$ we can always consider a different sequence $(\lambda_k,\tilde{\Gamma}_k,\tilde{\Theta}_k)_{k\in\mathbb{N}}$ with $\tilde{\Gamma}_k := \text{diag}^2(\Gamma_k)$ and $\tilde{\Theta}_k := \Theta_k - \text{ofd}(\Gamma_k)$. It is now immediate to check that the new sequence still belongs to $\mathscr{C}_1$ and that we still have $\inf J(\lambda,\Gamma,\Theta) = \lim_{k\to\infty} J(\tilde{\lambda}_k,\tilde{\Gamma}_k,\tilde{\Theta}_k)$. For this reason, we can further restrict our set $\mathscr{C}_1$ to $\mathscr{C}_2$:

$$\begin{aligned}\mathscr{C}_2 := \{(\lambda,\Gamma,\Theta) : \lambda>0, I+\Gamma-\Theta\succeq 0, \Gamma\succeq 0, \Gamma\in\mathbf{D}_n,\\ \Theta\in\mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1}+\lambda^{-1}(\Theta-\Gamma)\succ 0\}.\end{aligned}$$

Lemma 2.4.2. *Let $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_2$ such that*

$$\lim_{k\to\infty} \lambda_k = 0.$$

Then $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ is not an infimizing sequence for J .

Proof. We consider two cases separately. Let us first analyse the case of sequences $(\lambda_k, \Gamma_k, \Theta_k)$ in which, besides $\lambda_k \to 0$, we also have $\|\lambda_k^{-1}(\Theta_k - \Gamma_k)\| \to \infty$ as $k \to \infty$. This implies that the largest singular value of $\lambda_k^{-1}(\Theta_k - \Gamma_k)$ tends to infinity and this, by symmetry, implies in turn that

$$\lim_{k\to\infty} \max_{\alpha_k \in \sigma(\lambda_k^{-1}(\Theta_k - \Gamma_k))} |\alpha_k| = +\infty. \tag{2.16}$$

We now show that this implies

$$\lim_{k\to\infty} \min_{\alpha_k \in \sigma(\lambda_k^{-1}(\Theta_k - \Gamma_k))} \alpha_k = -\infty. \tag{2.17}$$

To this end, we observe that from (2.16) it follows that at least one of the following statements is true: (2.17) holds (and in this case we are done) or

$$\lim_{k\to\infty} \max_{\alpha_k \in \sigma(\lambda_k^{-1}(\Theta_k - \Gamma_k))} \alpha_k = +\infty. \tag{2.18}$$

In the latter case, we use the fact that $\Gamma_k \succeq 0$ and $\lambda_k > 0$, so that

$$\max_{\alpha_k \in \sigma(\lambda_k^{-1}\Theta_k)} \alpha_k \geq \max_{\alpha_k \in \sigma(\lambda_k^{-1}(\Theta_k - \Gamma_k))} \alpha_k$$

which, together with (2.18) gives

$$\lim_{k\to\infty} \max_{\alpha_k \in \sigma(\lambda_k^{-1}\Theta_k)} \alpha_k = +\infty. \tag{2.19}$$

Since $\mathrm{tr}(\lambda_k^{-1}\Theta_k) = 0$, (2.19) implies that

$$\lim_{k\to\infty} \min_{\alpha_k \in \sigma(\lambda_k^{-1}\Theta_k)} \alpha_k = -\infty. \tag{2.20}$$

Now we use again the fact that $\Gamma_k \succeq 0$ and $\lambda_k > 0$, so that

$$\min_{\alpha_k \in \sigma(\lambda_k^{-1}(\Theta_k - \Gamma_k))} \alpha_k \leq \min_{\alpha_k \in \sigma(\lambda_k^{-1}\Theta_k)} \alpha_k$$

which, together with (2.20), implies (2.17). In conclusion, for sequences $(\lambda_k, \Gamma_k, \Theta_k)$ of

this type and for a sufficiently large k, $\hat{\Sigma}^{-1}+\lambda_k^{-1}(\Theta_k-\Gamma_k)$ is no longer positive definite and therefore these sequences does not belong to $\mathscr{C}_2$.

Second, we consider the case of sequences $(\lambda_k,\Gamma_k,\Theta_k)$ in which, besides $\lambda_k\to 0$, we also have $\|\lambda_k^{-1}(\Theta_k-\Gamma_k)\|\to c$ as $k\to\infty$, where $c<+\infty$ is a non-negative value.

In this case, it is not difficult to see that $\forall\varepsilon>0$, $\exists\ \bar{k}$ such that the dual functional satisfies $J(\lambda_k,\Gamma_k,\Theta_k)>-\varepsilon$, $\forall k\geq\bar{k}$. In fact, since $\|\lambda_k^{-1}(\Theta_k-\Gamma_k)\|$ is bounded, there exists $l_0>0$ such that $\lambda_k^{-1}(\Theta_k-\Gamma_k)\leq l_0 I$ for all k. Therefore, there exists $l_1>0$ such that for all k, $\hat{\Sigma}^{-1}+\lambda_k^{-1}(\Theta_k-\Gamma_k)\leq l_1 I$ and hence there exists $l_2>0$ such that for all k, $|\hat{\Sigma}^{-1}+\lambda_k^{-1}(\Theta_k-\Gamma_k)|\leq l_2$. In turn, there exists $l_3\in\mathbb{R}$ such that for all k, $\log|\hat{\Sigma}^{-1}+\lambda_k^{-1}(\Theta_k-\Gamma_k)|\leq l_3$ and $-\log|\hat{\Sigma}^{-1}+\lambda_k^{-1}(\Theta_k-\Gamma_k)|\geq -l_3$. Finally, there exists a real constant $l_4:=-l_3-\log|\hat{\Sigma}|+\delta$ such that, for all k, $J(\lambda_k,\Gamma_k,\Theta_k)\geq\lambda_k l_4$. Since l_4 is constant, the right-hand side of this inequality converges to zero so that, by definition, $\forall\varepsilon>0$, $\exists\,\bar{k}$ such that $\lambda_k l_4>-\varepsilon\ \forall k\geq\bar{k}$. As a consequence, $J(\lambda_k,\Gamma_k,\Theta_k)>-\varepsilon$, $\forall k\geq\bar{k}$. It is therefore sufficient to exhibit a triple $(\bar{\lambda},\bar{\Gamma},\bar{\Theta})\in\mathscr{C}_2$ for which the dual functional is negative to conclude that sequences $(\lambda_k,\Gamma_k,\Theta_k)$ of this kind cannot be minimizing sequences. Let us consider $(\bar{\lambda},\bar{\Gamma},\bar{\Theta})$ such that $\bar{\lambda}>0$, $\bar{\Gamma}=0$ and

$$\bar{\Theta}=-\bar{\lambda}\,\mathrm{ofd}(\hat{\Sigma}^{-1}).$$

For $\bar{\lambda}$ sufficiently small, but strictly greater than zero, it is immediate to check that this triple is in $\mathscr{C}_2$. For this choice of the multipliers and taking into account (2.10) we have that

$$\begin{aligned}
J(\bar{\lambda},\bar{\Gamma},\bar{\Theta}) &= -\bar{\lambda}\log|\hat{\Sigma}^{-1}+\bar{\lambda}^{-1}(\bar{\Theta}-\bar{\Gamma})|-\bar{\lambda}\log|\hat{\Sigma}|+\bar{\lambda}\delta\\
&= -\bar{\lambda}\log|(\hat{\Sigma}^{-1}+\bar{\lambda}^{-1}\bar{\Theta})\hat{\Sigma}|+\bar{\lambda}\delta\\
&= -\bar{\lambda}\log|\,\mathrm{diag}^2(\hat{\Sigma}^{-1})\hat{\Sigma}|+\bar{\lambda}\delta\\
&= -\bar{\lambda}(\delta_{max}-\delta)<0.
\end{aligned}$$

This is sufficient to conclude the proof. In fact, the only other possible case is the one in which $\lim_{k\to\infty}\|\lambda_k^{-1}(\Theta_k-\Gamma_k)\|$ does not exist. In this case however, we can consider a sub-sequence $(\lambda_{k_j},\Gamma_{k_j},\Theta_{k_j})$ for which the corresponding limit does exist (finite or infinite) and we are thus reduced to one of the previous two cases. ■

As a consequence of the previous result we have that the minimization of the dual functional over the set $\mathscr{C}_2$ is equivalent to minimization over the set:

$$\begin{aligned}
\mathscr{C}_3:=\{(\lambda,\Gamma,\Theta):\lambda\geq\varepsilon, I+\Gamma-\Theta\succeq 0,\Gamma\succeq 0,\Gamma\in\mathbf{D}_n,\\
\Theta\in\mathbf{D}_n^{\perp},\hat{\Sigma}^{-1}+\lambda^{-1}(\Theta-\Gamma)\succ 0\}
\end{aligned}$$

for a certain $\varepsilon > 0$.

The next result provides an upper bound for λ.

Lemma 2.4.3. *Let $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_3$ such that*

$$\lim_{k\to\infty} \lambda_k = \infty. \tag{2.21}$$

Then $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ is not an infimizing sequence for J.

Proof. Let us consider a sequence $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ such that (2.21) holds. It follows from the condition $\Theta_k - \Gamma_k \preceq I$ that

$$\lambda_k^{-1}(\Theta_k - \Gamma_k) \preceq \lambda_k^{-1} I$$

which implies that

$$\begin{aligned} J(\lambda_k, \Gamma_k, \Theta_k) &= \lambda_k(\log|(\hat{\Sigma}^{-1} + \lambda_k^{-1}(\Theta_k - \Gamma_k))^{-1}\hat{\Sigma}^{-1}| + \delta) \\ &\geq \lambda_k(\log|((\hat{\Sigma}^{-1} + \lambda_k^{-1} I)^{-1}\hat{\Sigma}^{-1})| + \delta) \\ &\longrightarrow +\infty \end{aligned}$$

so that $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ cannot be an infimizing sequence. ∎

As a consequence, the set $\mathscr{C}_3$ can be further restricted to the set:

$$\begin{aligned} \mathscr{C}_4 := \{(\lambda, \Gamma, \Theta) : \varepsilon \leq \lambda \leq M, I + \Gamma - \Theta \succeq 0, \Gamma \succeq 0, \\ \Gamma \in \mathbf{D}_n, \Theta \in \mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1} + \lambda^{-1}(\Theta - \Gamma) \succ 0\} \end{aligned}$$

for a certain $M < \infty$.

The next result provides an upper bound for $\Theta - \Gamma$.

Lemma 2.4.4. *Let $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_4$ such that*

$$\lim_{k\to\infty} \|\Theta_k - \Gamma_k\| = +\infty. \tag{2.22}$$

Then $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ is not an infimizing sequence for J.

Proof. From (2.22) if follows that the largest singular value of $(\Theta_k - \Gamma_k)$ tends to $+\infty$ as $k \to \infty$. This in turn implies that, as $k \to \infty$, at least one of the eigenvalues of $(\Theta_k - \Gamma_k)$ diverges, because $(\Theta_k - \Gamma_k)$ is symmetric so that its singular values are the absolute values of its eigenvalues. As before, since $(\Theta_k - \Gamma_k) \preceq I$ holds, the diverging eigenvalues have to tend to $-\infty$. This implies that also $\hat{\Sigma}^{-1} + \lambda_k^{-1}(\Theta_k - \Gamma_k)$ has an eigenvalue which

tends to $-\infty$ as $k \to \infty$. But, this cannot be the case, because we have the positive definiteness constraint on $\hat{\Sigma}^{-1} + \lambda_k^{-1}(\Theta_k - \Gamma_k)$. ■

It follows from the previous result that there exists ρ such that $|\rho| < \infty$ and

$$\Theta - \Gamma \succeq \rho I.$$

Therefore, the set $\mathscr{C}_4$ can be further restricted to the set:

$$\begin{aligned}\mathscr{C}_5 := \{(\lambda, \Gamma, \Theta) : \varepsilon \leq \lambda \leq M, \rho I \preceq \Theta - \Gamma \preceq I, \Gamma \succeq 0,\\ \Gamma \in \mathbf{D}_n, \Theta \in \mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1} + \lambda^{-1}(\Theta - \Gamma) \succ 0\}.\end{aligned}$$

Now observe that, in $\mathscr{C}_5$, Θ and Γ are orthogonal so that if $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ is a sequence of elements in $\mathscr{C}_5$ such that

$$\lim_{k\to\infty} \|\Gamma_k\| = +\infty \tag{2.23}$$

or

$$\lim_{k\to\infty} \|\Theta_k\| = +\infty \tag{2.24}$$

then (2.22) holds. Then we have the following corollary.

Corollary 2.4.5. *Let $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_5$ such that (2.23) or (2.24) holds. Then $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}}$ is not an infimizing sequence for J.*

Thus, minimizing over the set $\mathscr{C}_5$ is equivalent to minimize over:

$$\begin{aligned}\mathscr{C}_6 := \{(\lambda, \Gamma, \Theta) : \varepsilon \leq \lambda \leq M, \rho I \preceq \Theta - \Gamma \preceq I, 0 \preceq \Gamma \preceq \alpha I,\\ \Gamma \in \mathbf{D}_n, \Theta \in \mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1} + \lambda^{-1}(\Theta - \Gamma) \succ 0\}\end{aligned}$$

for a certain α such that $0 < \alpha < +\infty$.

Finally, let us consider a sequence $(\lambda_k, \Gamma_k, \Theta_k)_{k\in\mathbb{N}} \in \mathscr{C}_6$ such that, as $k \to \infty$, the minimum eigenvalue of $\hat{\Sigma} + \lambda_k^{-1}(\Theta_k - \Gamma_k)$ tends to zero. This implies that $\left|\hat{\Sigma}^{-1} + \lambda_k^{-1}(\Theta_k - \Gamma_k)\right| \to 0$ and hence $J \to +\infty$. Thus, such sequence does not infimize the dual functional. The final set $\mathscr{C}_C$ is

$$\begin{aligned}\mathscr{C}_C := \{(\lambda, \Gamma, \Theta) : \varepsilon \leq \lambda \leq M, \rho I \preceq \Theta - \Gamma \preceq I, 0 \preceq \Gamma \preceq \alpha I,\\ \Gamma \in \mathbf{D}_n, \Theta \in \mathbf{D}_n^{\perp}, \hat{\Sigma}^{-1} + \lambda^{-1}(\Theta - \Gamma) \succeq \beta I\}\end{aligned}$$

for a suitable $\beta > 0$. Summing up we have the following result.

Theorem 2.4.6. *Problem (2.15) is equivalent to*

$$\min_{(\lambda,\Gamma,\Theta)\in\mathscr{C}_C} J(\lambda, \Gamma, \Theta).$$

Both these problems admit solution.

Proof. The equivalence of the two problems has already been proven by the previous argument. Since $\mathscr{C}_C$ is closed and bounded, and hence compact, and J is continuous over $\mathscr{C}_C$, by Weierstrass's Theorem the minimum exists. ∎

Remark 2.4.7. Clearly, the dual objective function J is bounded from below as the optimal value of the dual (maximization) problem is upper-bounded by the optimal value of the primal problem which is finite (as can be seen for instance by choosing $\Sigma = \hat{\Sigma}$ which, by assumption, is positive definite and bounded element wise).

Before discussing the uniqueness of the solution to (2.15), it is convenient to further simplify the dual optimization problem: consider the function

$$F(\lambda, X) := -\lambda[\log(|\hat{\Sigma}^{-1} + \lambda^{-1}X|) + \log|\hat{\Sigma}| - \delta]$$

where $\lambda > 0$ and $X \in \mathbf{Q}_n$. Note that

$$F(\lambda, \Theta - \Gamma) = J(\lambda, \Gamma, \Theta).$$

Moreover, Θ and Γ are orthogonal over $\mathscr{C}_C$ so that minimizing J over $\mathscr{C}_C$ is equivalent to minimize F over the corresponding set

$$\mathscr{C}_F := \{(\lambda, X) : \ \lambda > 0,\ X \in \mathbf{Q}_n, X \preceq I,\ -\operatorname{diag}^2(X) \succeq 0,\ \hat{\Sigma}^{-1} + \lambda^{-1}X \succ 0\}.$$

Therefore, from now on we can consider the following problem

$$\min_{(\lambda, X) \in \mathscr{C}_F} F(\lambda, X). \tag{2.25}$$

Once obtained the optimal solution (λ°, X°) we can recover the optimal values of the original multipliers simply by setting $\Theta^\circ = \operatorname{ofd}(X^\circ)$ and $\Gamma^\circ = -\operatorname{diag}^2(X^\circ)$.

2.4.2 Uniqueness of the solution

The aim of this subsection is to show that problem (2.25) (and hence problem (2.15)) admits a unique solution. It is easy to check that F is a convex function over the set $\mathscr{C}_F$. However, as we will see, F is not strictly convex. Accordingly, establishing the uniqueness of the minimum is not a trivial task.
The following proposition characterizes the second variation of F in direction $(\delta\lambda, \delta X)$, i.e. $\delta^2 F(\lambda, X; \delta\lambda, \delta X)$.

Proposition 2.4.8. *Let* $\mathbf{x} := \mathrm{vec}(X)$, $\delta\mathbf{x} := \mathrm{vec}(\delta X)$, *and* $K := (\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1} \otimes (\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}$. *Let also*

$$H := \begin{bmatrix} \lambda^{-3}\mathbf{x}^\top K\mathbf{x} & -\lambda^{-2}\mathbf{x}^\top K \\ -\lambda^{-2}K\mathbf{x} & \lambda^{-1}K \end{bmatrix} \in \mathbb{R}^{(1+n^2)\times(1+n^2)}.$$

Then, we have

$$\delta^2 F(\lambda, X; \delta\lambda, \delta X) = [\delta\lambda \;\; \delta\mathbf{x}^\top] H \begin{bmatrix} \delta\lambda \\ \delta\mathbf{x} \end{bmatrix}.$$

Proof. Consider the function

$$\tilde{F}(\lambda, X) = -\lambda \log|\hat{\Sigma} + \lambda^{-1}X|.$$

Since $\tilde{F}(\lambda, X)$ differs from $F(\lambda, X)$ only by terms which are linear in (λ, X) the second variations of the two functions are equal. Thus, in what follows we will focus on $\tilde{F}(\lambda, X)$. The first variation of $\tilde{F}(\lambda, X)$ in direction $(\delta\lambda, \delta X)$ is

$$\begin{aligned} \delta\tilde{F}(\lambda, X; \delta\lambda, \delta X) = &- \log|\hat{\Sigma} + \lambda^{-1}X|\delta\lambda \\ &+ \lambda^{-1}\mathrm{tr}((\hat{\Sigma} + \lambda^{-1}X)^{-1}X)\delta\lambda - \mathrm{tr}((\hat{\Sigma} + \lambda^{-1}X)^{-1}\delta X). \end{aligned}$$

The second variation of $\tilde{F}(\lambda, X)$ in direction $(\delta\lambda, \delta X)$ is

$$\begin{aligned} \delta^2\tilde{F}(\lambda, X; \delta\lambda, \delta X) = &\lambda^{-1}\mathrm{tr}\big((\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}\delta X(\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}\delta X\big) \\ &- 2\left[\lambda^{-2}\mathrm{tr}\big((\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}\delta X(\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}X\big)\delta\lambda\right] \\ &+ \lambda^{-3}\mathrm{tr}\big((\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}X(\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}X\big)\delta\lambda^2. \end{aligned}$$

Now, by using the Kronecker product and the vec operator and defining $\mathbf{x} := \mathrm{vec}(X)$, $\delta\mathbf{x} := \mathrm{vec}(\delta X)$, and $K := (\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1} \otimes (\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}$ the Hessian in the statement immediately follows. ∎

Since in $\mathscr{C}_F$ we have that $K \in \mathbf{Q}_n$ is positive definite and $\lambda > 0$, the matrix H, which has clearly the meaning of the Hessian of F, has at least rank equal to n^2. Moreover, $Hw = 0$ with $w = [\lambda \; \mathbf{x}^\top]^\top$. We conclude that H has rank equal to n^2.
This means that F is convex and there is exactly one direction along which F is not strictly convex. We now analyse this direction in the neighbourhood of the optimal solution.

Lemma 2.4.9. *Any optimal solution* (λ°, X°) *minimizing* F *over* $\mathscr{C}_F$ *lies on the boundary of* $\mathscr{C}_F$ *and, specifically, is such that* $I - X^\circ$ *is singular.*

Proof. Let (λ°, X°) be an optimal solution and assume, by contradiction, that (λ°, X°) does not belong to the boundary of the feasible set $\mathscr{C}_F$, so that, in particular, $X^\circ \prec I$.

Thus there exists $\varepsilon > 0$ such that $(1+\varepsilon)X^\circ \prec I$ so that

$$((1+\varepsilon)\lambda^\circ, (1+\varepsilon)X^\circ) \in \mathscr{C}_F.$$

Now a direct computation yields

$$F((1+\varepsilon)\lambda^\circ, (1+\varepsilon)X^\circ) = (1+\varepsilon)F(\lambda^\circ, X^\circ) < F(\lambda^\circ, X^\circ)$$

where the last inequality is a consequence of the fact that, as we have already seen in the proof of Lemma 2.4.2, the optimal value of J (and, hence, of F) is negative. This a contradiction as $F(\lambda^\circ, X^\circ)$ is assumed to be a minimum. ∎

Remark 2.4.10. Notice that for any $(\lambda_0, X_0) \in \mathscr{C}_F$, the direction $(\varepsilon\lambda_0, \varepsilon X_0)$ (which, by the way, is the direction considered in Lemma 2.4.9 for the specific case of the optimal solution (λ°, X°)) is exactly the unique direction along which F is not strictly convex. In fact, along this direction F is clearly a linear function of λ. Notice also that F is constant along this direction if and only if $F(\lambda_0, X_0) = 0$. Since at any optimal solution (λ°, X°) F is necessarily negative, F is not constant along the direction $(\varepsilon\lambda^\circ, \varepsilon X^\circ)$ (which is the only direction along which F is not strictly convex).

As a consequence of this observation, we have the following result.

Corollary 2.4.11. *Let (λ_0, X_0) be a given point in $\mathscr{C}_F$. If $w := (\delta\lambda, \delta X)$ is any direction along which $F(\lambda_0, X_0)$ is constant, i.e. $F(\lambda_0, X_0) = F(\lambda_0 + \alpha\delta\lambda, X_0 + \alpha\delta X)$ for any α such that $|\alpha| > 0$ is sufficiently small, then $F(\lambda_0, X_0) = 0$.*

We are now ready to state our main result.

Theorem 2.4.12. *The dual problem admits a unique solution.*

Proof. Assume, by contradiction, that there are two optimal solutions $(\lambda_1^\circ, X_1^\circ)$ and $(\lambda_2^\circ, X_2^\circ)$. By the convexity of $\mathscr{C}_F$, the whole segment $\mathcal{S}$ connecting $(\lambda_1^\circ, X_1^\circ)$ to $(\lambda_2^\circ, X_2^\circ)$ belongs also to $\mathscr{C}_F$. It follows by the convexity of $F(\cdot,\cdot)$ that all the points in $\mathcal{S}$ are optimal solutions. Notice, *en passant*, that in view of Lemma 2.4.9, this implies that $\mathcal{S}$ belongs to the boundary of $\mathscr{C}_F$. Now, F is clearly negative and constant along $\mathcal{S}$ and this is a contradiction in view of Corollary 2.4.11. ∎

2.5 Recovering the solution of the primal problem

The optimal Σ° can be easily recovered by substituting the optimal solution of the dual problem $(\lambda^\circ, \Theta^\circ, \Gamma^\circ)$ into (2.12). Recovering the optimal L° is slightly more involved. To this end, we observe that the primal problem is strictly feasible (which can be seen by taking, for instance, $\Sigma = \hat{\Sigma}$) and hence, by virtue of the properties of the primal and dual

problems, the following extremality relations hold, see e.g. (Ekeland and Temam, 1999, Theorem 5.1):

$$\operatorname{tr}(\Lambda^{\circ} L^{\circ}) = 0 \tag{2.26}$$

$$\operatorname{tr}(\Gamma^{\circ}(\Sigma^{\circ} - L^{\circ})) = 0 \tag{2.27}$$

$$\operatorname{tr}(\Theta^{\circ}(\Sigma^{\circ} - L^{\circ})) = 0. \tag{2.28}$$

We begin by considering (2.26). It follows from (2.13) that

$$\Lambda^{\circ} = I + \Gamma^{\circ} - \Theta^{\circ}$$

where we now know that Λ° has deficient rank. Thus, we consider the following reduced singular value decomposition

$$\Lambda^{\circ} = USU^{\top} \tag{2.29}$$

with $S \in \mathbf{Q}_{n-r}$ positive definite, i.e. $n-r$ is the rank of Λ°, and $U \in \mathbb{R}^{n \times n-r}$ such that $U^{\top}U = I_{n-r}$. We plug (2.29) in (2.26) and get

$$0 = \operatorname{tr}[\Lambda^{\circ} L^{\circ}] = \operatorname{tr}[USU^{\top} L^{\circ}] \Rightarrow U^{\top} L^{\circ} U = 0.$$

Then, by selecting a matrix $\tilde{U} \in \mathbb{R}^{n \times r}$ whose columns form an orthonormal basis of $[\operatorname{im}(U)]^{\perp}$, we can express L° as:

$$L^{\circ} = \tilde{U} Q \tilde{U}^{\top} \tag{2.30}$$

with $Q \in \mathbf{Q}_r$. Note that, in view of the fact that the columns of $\tilde{U}$ form the orthogonal complement of the image of U, the relationship $U^{\top}\tilde{U} = 0$ holds.

By (2.28), we know that $\Sigma^{\circ} - L^{\circ}$ is diagonal. Thus, we plug (2.30) into (2.28) and obtain a linear system of equations: $\operatorname{ofd}(\Sigma^{\circ} - \tilde{U} Q \tilde{U}^{\top}) = 0$, or equivalently,

$$\operatorname{ofd}(\tilde{U} Q \tilde{U}^{\top}) = \operatorname{ofd}(\Sigma^{\circ}).$$

In an analogous fashion, using (2.27) we obtain an additional system of linear equations. By virtue of the fact that both the dual and the primal problem admit solution the resulting system of equations always admits solution in Q. Moreover, the solution of this system of equations is unique if and only if the solution of the primal problem is unique.

2.6 Numerical implementation

In this section we propose an algorithm for computing the numerical solution of problem (2.25). First, recall that the optimal solution lies on the boundary and it is characterized

by the constraints $-\operatorname{diag}^2(X) \succeq 0$ and $X \preceq I$. Finding a descending direction (λ, X) for $F(\lambda, X)$ satisfying simultaneously these two constraints may be a difficult task. We resort to the Alternating Direction Method of Multipliers (ADMM) algorithm, Boyd, Parikh, Chu, Peleato, and Eckstein (2011), for decoupling such constraints. The corresponding ADMM updates can be performed by using a projection gradient algorithm. To this end, we rewrite (2.25) by introducing the new variable $Y \in \mathbf{Q}_n$ defined as $Y := I - X$:

$$\begin{aligned} \min_{(\lambda,X)\in\mathscr{C}_{\lambda,X}, Y\in\mathscr{C}_Y} \quad & F(\lambda, X) \\ \text{subject to} \quad & Y = I - X \end{aligned}$$

where $\mathscr{C}_{\lambda,X}$, and $\mathscr{C}_Y$ are defined, respectively, as

$$\begin{aligned} \mathscr{C}_{\lambda,X} :=& \{(\lambda, X) : \lambda > 0, X \in \mathbf{Q}_n, \hat{\Sigma}^{-1} + \lambda^{-1} X \succ 0, -\operatorname{diag}^2(X) \succeq 0\} \\ \mathscr{C}_Y :=& \{Y : Y \in \mathbf{Q}_n, Y \succeq 0\}. \end{aligned}$$

The *augmented Lagrangian* (see Boyd et al. (2011)) for the problem is

$$\mathscr{L}_\rho(\lambda, X, Y, M) = F(\lambda, X) + \langle M, Y - I + X\rangle + \frac{\rho}{2}\|Y - I + X\|_F^2$$

where $M \in \mathbf{Q}_n$. Accordingly, given the initial values λ^0, X^0, Y^0 and M^0, the ADMM updates are:

$$(\lambda^{(k+1)}, X^{(k+1)}) := \operatorname*{arg\,min}_{(\lambda,X)\in\mathscr{C}_{\lambda,X}} \mathscr{L}_\rho(\lambda, X, Y^{(k)}, M^{(k)}) \tag{2.31}$$

$$Y^{(k+1)} := \operatorname*{arg\,min}_{Y\in\mathscr{C}_Y} \mathscr{L}_\rho(\lambda^{(k+1)}, X^{(k+1)}, Y, M^{(k)}) \tag{2.32}$$

$$M^{(k+1)} := M^{(k)} + \rho(Y^{(k+1)} - I + X^{(k+1)}) \tag{2.33}$$

where $\rho > 0$ is the penalty parameter. Here, we choose $\rho = 0.5$.

Problem (2.31) has not a closed form solution. Thus, the solution is approximated by a projective gradient whose updating steps are:

$$\begin{aligned} \lambda^{(k+1)} &:= \lambda^{(k)} - t_k \nabla_\lambda \mathscr{L}_\rho(\lambda^{(k)}, X^{(k)}, Y^{(k)}, M^{(k)}) \\ X^{(k+1)} &:= \Pi_{\mathscr{C}_X^*}(X^{(k)} - t_k \nabla_X \mathscr{L}_\rho(\lambda^{(k)}, X^{(k)}, Y^{(k)}, M^{(k)})) \end{aligned}$$

where $\Pi_{\mathscr{C}_X}$ denotes the projector operator from $\mathbf{Q}_n$ onto

$$\mathscr{C}_X := \{X : X \in \mathbf{Q}_n, -\operatorname{diag}^2(X) \succeq 0\},$$

and $\nabla_\lambda \mathscr{L}_\rho$, $\nabla_X \mathscr{L}_\rho$ denotes the gradient with respect to λ and X, respectively:

$$\nabla_\lambda \mathscr{L}_\rho(\lambda, X, Y, M) := -\log|\hat{\Sigma}^{-1} + \lambda^{-1}X| - \log|\hat{\Sigma}| + \delta + \lambda^{-1}\operatorname{tr}\left((\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1}X\right)$$
$$\nabla_X \mathscr{L}_\rho(\lambda, X, Y, M) := -(\hat{\Sigma}^{-1} + \lambda^{-1}X)^{-1} + M + \rho(Y - I + X).$$

It is not difficult to see that

$$[\Pi_{\mathscr{C}_X}(A)]_{(i,j)} = \begin{cases} 0, & \text{if } i = j \text{ and } A_{(i,j)} > 0 \\ A_{(i,j)}, & \text{otherwise} \end{cases}$$

where $A_{(i,j)}$ denotes the entry in position (i, j) of matrix $A \in \mathbf{Q}_n$. The step size t_k is determined at each step k in a iterative fashion: we start by setting $t_k = 1$ and we decrease it progressively until the two conditions $\lambda^{(k+1)} > 0$ and $\hat{\Sigma}^{-1} + \lambda^{-1}X \succ 0$ are met and the so-called Armijo's conditions, Boyd and Vandenberghe (2004), are satisfied.
Problem (2.32) can be rewritten as

$$Y^{(k+1)} = \underset{Y \in \mathscr{C}_Y}{\arg\min} \|I - X^{(k+1)} - \frac{1}{\rho}M^{(k)} - Y\|_F.$$

We introduce the projection operator $\Pi_{\mathscr{C}_Y} : \mathbf{Q}_n \to \mathscr{C}_Y$ which is defined by

$$\Pi_{\mathscr{C}_Y}(W) := \underset{Z \in \mathscr{C}_Y}{\arg\min} \|W - Z\|_F^2.$$

It is not difficult to see that, if $A = UDU^\top$ is the eigenvalue decomposition of the matrix $A \in \mathbf{Q}_n$, then

$$\Pi_{\mathscr{C}_Y}(A) = U \operatorname{diag}(f(d_1), ..., f(d_n))U^\top$$

where

$$f(d_i) := \begin{cases} d_i, & \text{if } d_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then the solution of (2.32) becomes

$$Y^{(k+1)} = \Pi_{\mathscr{C}_Y}(I - X^{(k+1)} - \frac{1}{\rho}M^{(k)}).$$

In order to set the stopping criteria for the algorithm we define the primal and dual residual matrices:

$$R^{(k+1)} := Y^{(k+1)} - I + X^{(k+1)},$$
$$S^{(k+1)} := \rho_k(Y^{(k+1)} - Y^{(k)}).$$

The algorithm reaches an acceptable solution when the following conditions are met,

Boyd et al. (2011):

$$\|R^{(k+1)}\|_F \leq n\varepsilon^{abs} + \varepsilon^{rel} \max\{\sqrt{n}, \|X^k\|_F, \|Y^k\|_F\},$$
$$\|S^{(k+1)}\|_F \leq n\varepsilon^{abs} + \varepsilon^{rel} \|M^{(k)}\|_F$$

where $\varepsilon^{rel} = 10^{-4}$ and $\varepsilon^{abs} = 10^{-4}$ are the relative and the absolute tolerances, respectively.

2.7 Numerical examples

In this section we consider Monte Carlo studies composed by 200 experiments whose structure is as follows. For each experiment:

- we consider a factor model having the structure of (2.2) with cross sectional dimension $n = 40$; L and D are randomly generated in such a way that L has rank equal to r (*a priori* fixed), D is diagonal and the signal-to-noise ratio (defined as $\|L\|/\|D\|$) between the latent and the idiosyncratic components is equal to one;

- a data sequence of length N for $\mathbf{y}$ is generated;

- we compute the sample covariance matrix $\hat{\Sigma}$ from this data;

- we compute the estimate δ_α of δ using the empirical procedure of Section 2.3 with $\alpha = 0.5$;

- we compute the solution (L_{OPT}, Σ_{OPT}) of problem (2.6) where we replace δ with δ_α. Let λ_i, $i = 1 \dots n$, denote the singular values of L_{OPT} and define i_{max} as the first i such that $\lambda_{i+1}/\lambda_1 < 0.05$. Then, we define the "numerical rank" of L_{OPT} as:

$$r_{OPT} := \max_{i \leq i_{max}} \lambda_i/\lambda_{i+1}$$

- we compute the solution of the standard problem (2.4) (exact decomposition with trace heuristic) and, with the same procedure of the previous point, we compute the numerical rank, r_{ED}, of the corresponding low rank matrix;

- we compute the minimum number of factors from the data sequence for $\mathbf{y}$ by applying the three methods proposed by Bai and Ng (see Bai and Ng (2002)), namely: ICP1, ICP2 and ICP3. We denote the corresponding estimates by r_{ICP1}, r_{ICP2} and r_{ICP3}, respectively;

	Proposed method	Exact Decomposition	Bai & Ng			Lam & Yao
			ICP1	ICP2	ICP3	
N=200	0.500	3.752	3.589	2.546	7.506	5.529
N=500	0.000	0.9618	2.271	2.273	4.236	5.347
N=1000	0.000	0.6557	3.587	3.213	3.927	5.421

Table 2.1: Average root mean squared error between the estimated numerical rank and the true rank $r = 4$.

- we compute the minimum number of factors from the data sequence for $\mathbf{y}$ by applying the method proposed by Lam and Yao (see Lam and Yao (2012)).[2] We denote by r_{LY} the resulting estimate of the rank.

Finally, we compute the root mean squared error:

$$e = \sqrt{\frac{1}{200}\sum_{i=1}^{200}(r_i^\sharp - r)^2} \tag{2.34}$$

for $r^\sharp = \{r_{OPT}, r_{ED}, r_{ICP1}, r_{ICP2}, r_{ICP3}, r_{LY}\}$ and where r is the true rank of the data generating process.
Table 2.1 shows the error (2.34) for three Monte Carlo studies where $r = 4$ and the sample size is $N = 200$, $N = 500$ and $N = 1000$, respectively.
Usually, the problem becomes more challenging when the rank r of the data generating process increases (yet remaining below the Ledermann bound). For this reason we repeat the above Monte Carlo studies for the case $r = 10$ (considering again the three sample sizes $N = 200$, $N = 500$, $N = 1000$). The corresponding root mean squared errors (2.34) are reported in Table 2.2.
As one can see, in all these six Monte Carlo studies the proposed method outperforms the others.

We now analyze how well the proposed method recovers the subspace of L by considering the following measure of discrepancy. Let $L = AA^\top$ be the low rank matrix of the data generating process and consider the singular value decomposition of L_{OPT}, that is $L_{OPT} = USV^\top$. Let $\tilde{U} := U_{[1:n,1:r_{OPT}]}$, $\tilde{U} \in \mathbb{R}^{n\times r_{OPT}}$ be the matrix formed by the first r_{OPT} columns of U and $\tilde{S} := S_{[1:r_{OPT},1:r_{OPT}]}$, $\tilde{S} \in \mathbb{R}^{r_{OPT}\times r_{OPT}}$ be the top left $r_{OPT} \times r_{OPT}$ sub-matrix of S. We define the orthogonal projector onto the subspace generated by the

[2] The estimation procedure for this method requires to set a parameter k_0 for the selection which only general considerations are provided: we decided to select k_0 using an "oracle" procedure i.e. for each Monte Carlo run we choose the value of k_0 which yields the most favourable result.

	Proposed method	Exact Decomposition	Bai & Ng			Lam & Yao
			ICP1	ICP2	ICP3	
N=200	2.170	7.218	5.888	5.629	8.254	6.943
N=500	0.174	5.221	5.214	5.258	5.812	6.536
N=1000	0	2.961	5.302	5.196	5.490	6.669

Table 2.2: Average root mean squared error between the estimated numerical rank and the true rank $r = 10$.

columns of $A_{OPT} := \tilde{U}\tilde{S}$ as

$$P := A_{OPT}(A_{OPT}^{\top}A_{OPT})^{-1}A_{OPT}^{\top}.$$

Then, a measure of discrepancy between the subspace of A and the subspace of A_{OPT} is given by:

$$s(A_{OPT}) := \mathrm{tr}(A^{\top}PA)/\mathrm{tr}(A^{\top}A) \tag{2.35}$$

where $s(A_{OPT})$ takes value between 0 and 1. Note that, if $s(A_{OPT}) = 1$ then A_{OPT} recovers exactly the image of A. Figure 2.2 shows the box-plots for the quantity (2.35) in the six Monte Carlo studies.

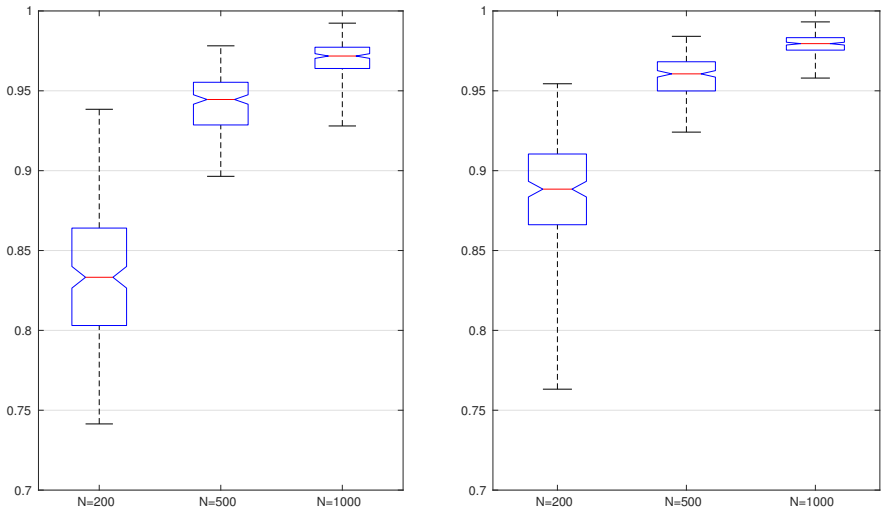


Figure 2.2: Performances of the subspace recovery: box-plots of the quantity (2.35) for the Monte Carlo studies with sample sizes $N = 200, 500, 1000$ and for the case $r = 4$ (left hand side panel) and $r = 10$ (right hand side panel).

Finally, we consider the example illustrated in Figure 2.1 of the Introduction. By applying our method, we obtain the situation illustrated in Figure 2.3 showing that our approach provides a numerical rank equal to the true value of r.

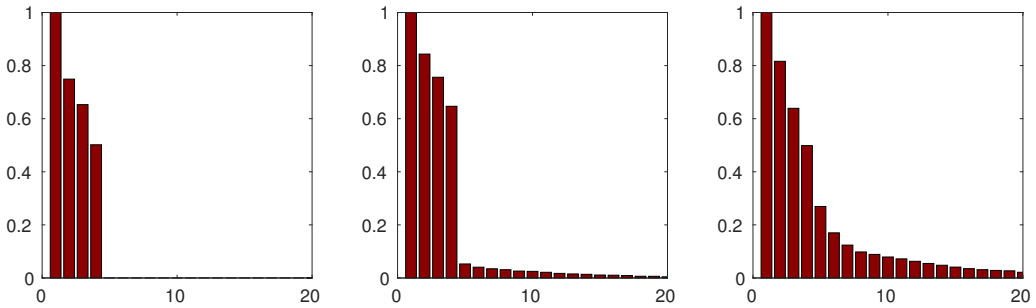


Figure 2.3: A sample of numerosity 1000 has been generated from a factor model with $n = 40$ and $r = 4$. The Figure displays the first twenty singular values of the true matrix L (on the left) and of the matrices L_{OPT} (in the middle) and L_{ED} (on the right) estimated, respectively, with the proposed method and with the trace heuristic with exact decomposition.

2.8 Concluding remarks and future directions

In this chapter we have proposed a new method to estimate the number of factors for the realistic situation in which the covariance matrix of the data is estimated with an error that is not negligible.

A question which arises naturally concerns the statistical properties of the proposed estimator, and, in particular, its asymptotic properties as the sample size approaches infinity. This is a complex issue and we present only some ideas and a heuristic road-map that may be followed. A primary issue is the rank consistency of the minimum trace estimator, note that a similar matter has been studied e.g. in Bach (2008) for the Lasso problem. Restricting to the cases in which the minimizers of the minimum rank problem and of the minimum trace problem coincide, a possible argument may be the following. By the construction presented in Section 2.3, it holds that

$$Pr[2\mathbb{D}(\Sigma\|\hat{\Sigma}) < \delta] = \alpha.$$

We can let the desired precision α be a function of the sample size N, and choose $\alpha(N)$ such that, as $N \to \infty$, it holds that $\alpha(N) \to 1$. Moreover, we let $\alpha(N) \to 1$ sufficiently slowly so that it is reasonable to expect that $\delta(\alpha(N)) \to 0$ because $\hat{\Sigma}$ converges to Σ as $N \to \infty$. Consequently, as $N \to \infty$, the neighbourhood of $\hat{\Sigma}$ in which we seek for the solution becomes smaller and smaller and it contains the "true" Σ with probability tending to 1. Moreover, the minimum rank problem (2.3) is a lower semi-continuous

function of Σ (see Ning et al. (2015), Proposition 1) and, being integer valued, it does not decrease in a sufficiently small neighborhood of Σ. Therefore, it seems reasonable to conclude that $\forall\, \varepsilon > 0\ \exists\, \bar{N}$ such that $\forall N > \bar{N}$ it holds that

$$Pr[r_{OPT} \neq r_{true}] < \varepsilon.$$

A rigorous study of this heuristic argument will be the subject of future investigation.

Another related question, that may arise naturally when considering, for example, the existing methods for estimating the number of factors as done in Section 2.7 is the following: for a given covariance matrix Σ and a given estimate r° of the number of factors, find a decomposition $\Sigma = L + D + E$ where D is a positive semi-definite diagonal matrix, L is a positive semi-definite matrix of rank r° and the error E is minimal in some norm. This is a non-convex problem that has been addressed in Ciccone, Ferrante, and Zorzi (2019a) where a projection type algorithm has been proposed together with a local convergence analysis.

Finally, another natural direction of research is the extension of the paradigm proposed in this chapter to the dynamic case in the spirit of the analysis that we will present in Chapter 4 for latent variable dynamic graphical models.

3

Robust identification of latent variable graphical models

3.1 Introduction

Graphical Models are used to provide a graph representation of the relations between random variables. In particular, Gaussian graphical models can be used to describe the conditional independence relations between the m components of a zero-mean Gaussian random vector $\mathbf{x}$ by means of an *interaction graph* $\mathscr{G}(V_m, E_m)$. This is an undirected graph where the set V_m contains m nodes and E_m is the subset of the pairs of nodes that are directly connected by an edge. In this representation, the i-th node represents the i-th component x_i of vector $\mathbf{x}$; each edge represents a dependence relation: no edge between the nodes i and j indicates that x_i and x_j are conditionally independent, given all the others x_k; in more formal terms, for any pair of distinct nodes i and j, $(i, j) \notin E_m$ if and only if $x_i \perp x_j | \{x_k\}_{k \neq i,j}$. This graphical structure is very powerful to represent interdependence of the various components in a complex system with many variables, and for this reason it has been used and analyzed in a huge amount of papers in Statistics, Engineering, and Signal processing, to mention only the main applications, Dempster (1972); Chandrasekaran et al. (2010); Chandrasekaran et al. (2011); Tao and Yuan (2011); Zhou and Tao (2011).

The case of graphs with a small number of edges is very interesting in applications because such a pattern highlights a simple structure where the values of most pairs of variables does not have a *direct* influence on each other. Moreover, this structure clarifies the paths of interdependence along which the value of each variable may affect the value

of the other ones. In most situations, however, the graph is complete (or almost complete) as there is no pairs of variables that are conditionally independent given the others. Of course, this *may* reveal that we are considering a genuinely complex system where each variable has direct influence on each other, but it may also point to a different and much more interesting situation. The essence of this situation is well described by the following very simple example. Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} y_1 + y_{m+1} \\ y_2 + y_{m+1} \\ \vdots \\ y_m + y_{m+1} \end{bmatrix},$$

where y_i, $i = 1, \ldots, m+1$ are independent Gaussian random variables. It is easy to see that the graph associated with $\mathbf{x}$ is complete (i.e. each pair of nodes is connected by an edge). The interdependence pattern between the variables of the vector $\mathbf{x}$, however, has a very interesting structure providing a powerful interpretation. This structure can be highlighted by considering the *augmented* vector $\tilde{\mathbf{x}} := [\mathbf{x}^\top \ (y_{m+1})^\top]^\top$; indeed, it is easy to see that the graph associated with $\tilde{\mathbf{x}}$ has only the m edges connecting y_{m+1} to each of the x_i which provides the following interpretation: the interdependence between all the *manifest* (observed) variables x_i is completely explained by the dependence of each x_i with a common *latent* (hidden) variable y_{m+1}. It is now clear why a great effort has been dedicated to uncover this hidden structure by only observing the manifest variables in $\mathbf{x}$.

In this chapter and in the next one, we will focus on Gaussian graphical models with a two-layer structure: our aim is to integrate the set of manifest variables - that we think to be arranged in a bottom layer - with a *small* number of latent variables - arranged in a top layer - in such a way that a drastic reduction in the number of edges between the manifest variables is achieved. An example of this structure is depicted in Figure 3.1 where it is easy to see that if we considered only the manifest variables $x_1, \ldots, x_6$ the corresponding graph would be complete. However, almost all of the interdependence between these variables is explained by the two latent variables x_7, x_8 so that an insightful structure emerges when we integrate these two variables with the observed ones.

As an application of the theory developed in Dempster's seminal paper ,Dempster (1972), an identification procedure for such Gaussian graphical models has been developed which is based on the *"Sparse plus Low-rank"* decomposition of the manifest concentration matrix Σ_m^{-1} of the manifest vector $\mathbf{x}$ (Σ_m being the covariance matrix of $\mathbf{x}$), Chandrasekaran et al. (2010, 2011):

$$\Sigma_m^{-1} = S - L. \tag{3.1}$$

Indeed, such a decomposition, where S is symmetric and positive definite and L is symmetric and positive semi-definite, provides a two-layers graphical model with $\text{rank}(L)$

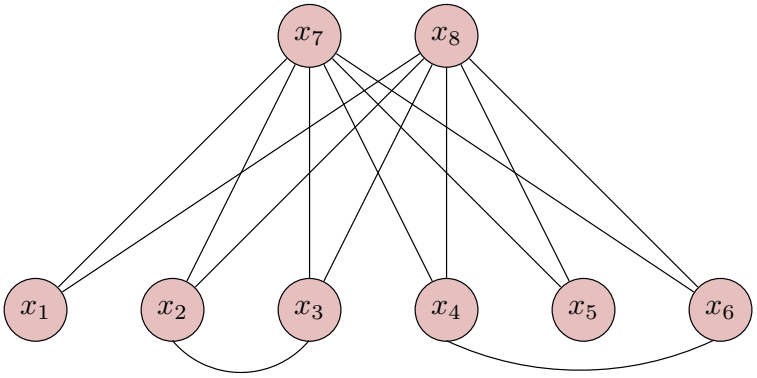


Figure 3.1: Example of a latent variable graphical model: the nodes x_7, x_8 represent the latent variables while the nodes $x_1, x_2, ..., x_6$ represent the manifest variables.

latent variables and a number of edges between the observed variables that is equal to the number of non-zero upper-off-diagonal (or, equivalently, lower-off-diagonal) entries of the matrix S. Therefore, for the solution of our problem we seek for a decomposition of the form (3.1) where the rank of the matrix L and the number of non-zero entries of the matrix S are minimized.

3.1.1 Motivating considerations

In real applications, however, Σ_m is normally estimated from the data so that only a noisy version $\hat{\Sigma}_m$ of Σ_m is usually available and the accuracy of this estimation may severely affect the goodness of the result – in terms of minimum rank and maximum sparsity – of the aforementioned optimization problem. More precisely, even in the case where the data are indeed produced by a mechanism in which a few non-observable variables explain most of the interdependence between the observed variables, relatively small variations of the covariance matrix $\hat{\Sigma}_m$ from the true value Σ_m may produce significant changes in the numerical rank of L and the numerical sparsity of S. To see this, we considered a sparse matrix S_0 of dimension 10 and a positive semi-definite matrix L_0 of dimension 10 and rank 3 such that $S_0 - L_0 \succ 0$; then we considered a matrix $\Sigma_m := (S_0 - L_0)^{-1}$. We generated a sample of $N = 1000$ independent realizations of a Gaussian random vector with zero mean and covariance matrix Σ_m and from them we estimated the sample covariance $\hat{\Sigma}_m$. Then we computed the "Sparse plus Low-rank" decompositions $\Sigma_m^{-1} = \tilde{L}_0 - \tilde{S}_0$ and $\hat{\Sigma}_m^{-1} = \hat{L}_0 - \hat{S}_0$, using an "exact" decomposition approach, Chandrasekaran et al. (2011). In Figure 3.2 the sparsity pattern of S_0, $\tilde{S}_0$ and $\hat{S}_0$ and the eigenvalues of $L_0, \tilde{L}_0$ and $\hat{L}_0$ are depicted providing evidence of the degradation of the solution when $\hat{\Sigma}_m$ is substituted to Σ_m. In fact, it is apparent that when the true covariance matrix is employed the algorithm recovers a solution with the correct numbers of latent variables and of non-zero elements

of the matrix S, while when the covariance is estimated (even from as many as $N = 1000$ data) the "sparse plus low-rank" structure is completely lost. Therefore, we are dealing

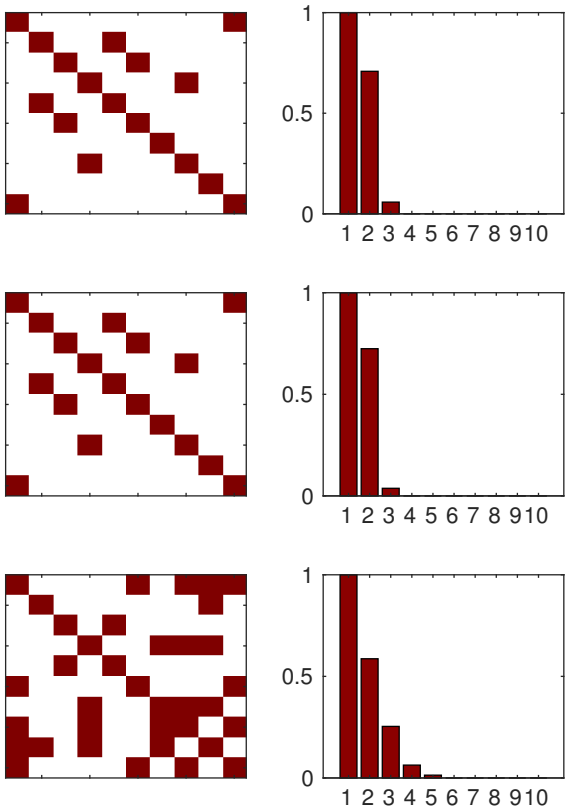


Figure 3.2: First row: sparsity pattern of the true sparse matrices S_0 and eigenvalues of the true low rank matrix L_0. Second and third row: sparsity pattern of the matrices $\tilde{S}_0$ and $\hat{S}_0$ and eigenvalues of the matrices $\tilde{L}_0$ and $\hat{L}_0$ obtained by applying the "exact" decomposition algorithm to the true covariance matrix Σ_m and to the estimated covariance matrix $\hat{\Sigma}_m$, respectively.

with a delicate problem where the solution is highly sensible to the observed data. In this chapter we deal with this issue by taking into account the uncertainty in the estimation of Σ_m. This is done by adopting the same strategy proposed in Chapter 2 for the factor analysis problem and leads to the following optimization problem: for a given $\hat{\Sigma}_m$, we propose to *compute* the matrix Σ_m in such a way that in the decomposition $\Sigma_m^{-1} = S - L$ the rank of L is minimized while the sparsity of S is maximized under a constraint which limits the Kullback-Leibler divergence between Σ_m and the sample covariance $\hat{\Sigma}_m$ to a prescribed tolerance depending on the precision of $\hat{\Sigma}_m$.

Outline of the chapter

The chapter is organized as follows. In Section 3.2 Gaussian graphical models and the *Sparse plus Low-rank* structure for the inverse covariance matrix are introduced. Then, the main optimization problem is presented together with some connections and comparison with the negative log-likelihood approach. In Section 3.3 we define the dual problem and we establish existence and uniqueness of its solution. Then, in Section 3.4 we discuss how to recover the solution of the primal problem. Finally, in Section 3.5 we discuss the numerical implementation and we propose a numeric example.

3.2 Problem formulation

3.2.1 Latent variable Gaussian graphical models

Let $\tilde{\mathbf{x}}$ be a zero-mean Gaussian random vector of dimension $m+l$, that is $\tilde{\mathbf{x}} := [\mathbf{x}^\top \mathbf{y}^\top]^\top$, where $\mathbf{x} := [x_1 \dots x_m]^\top$ plays the role of the manifest vector and $\mathbf{y} := [x_{m+1} \dots x_{m+l}]^\top$ plays the role of the latent vector. We partition the covariance $\Sigma \in \mathbf{Q}_{m+l}$ of $\tilde{\mathbf{x}}$ conformably with the partition of $\tilde{\mathbf{x}}$ as:

$$\Sigma = \begin{bmatrix} \Sigma_m & \Sigma_{lm}^\top \\ \Sigma_{lm} & \Sigma_l \end{bmatrix}. \tag{3.2}$$

We are interested in the Gaussian graphical model of $\tilde{\mathbf{x}}$ and hence in the conditional independence among the component of $\tilde{\mathbf{x}}$. Thus we recall a fundamental result stating that two distinct elements of a Gaussian random vector are conditionally independent given all the others if and only if the corresponding element in the concentration matrix (the inverse of the covariance) is zero, see e.g Dempster (1972). In our case, this reads:

$$\forall i \neq j, \quad x_i \perp x_j | \{x_k\}_{k \neq i,j} \Leftrightarrow [\Sigma^{-1}]_{ij} = 0.$$

Let us denote by K the inverse of covariance matrix Σ. Then, K can be also partitioned in the same way as (3.2):

$$K = \Sigma^{-1} = \begin{bmatrix} K_m & K_{lm}^\top \\ K_{lm} & K_l \end{bmatrix}$$

and, by using the *Schur complement*, we can obtain the relationship

$$\Sigma_m^{-1} = K_m - K_{lm}^\top K_l^{-1} K_{lm}$$

where the sparsity pattern of K_m provides the relations of conditional independence between the manifest variables in $\mathbf{x}$ and the rank of $K_{lm}^\top K_l^{-1} K_{lm}$ provides (an upper bound for) the number l of latent variables.
Now we recall that the only data that we can access are $\mathbf{x}$ and its covariance matrix Σ_m

while the rest of Σ is a purely artificial construction. The previous argument, however, provides a procedure for identifying a model Σ from Σ_m. In fact, if we decompose Σ_m^{-1} as

$$\Sigma_m^{-1} = S - L, \qquad S \succ 0,\ L \succeq 0,$$

where the rank l of L is as small as possible and S is as sparse as possible (of course there is a trade off between these two conditions) we may identify S with K_m and L with $K_{lm}^\top K_l^{-1} K_{lm}$.
This argument naturally leads to an optimization problem where the optimal S and L must minimize a combination of two penalty functions, ϕ_1 and ϕ_* , inducing sparsity and low-rankness on S and L, respectively, Chandrasekaran et al. (2011):

$$\begin{aligned} \underset{S,L\in\mathbf{Q_m}}{\arg\min} \quad & \phi_*(L) + \gamma\phi_1(S) \\ \text{subject to} \quad & \Sigma_m^{-1} = S - L, \\ & L \succeq 0 \end{aligned} \tag{3.3}$$

where the parameter γ balances the trade-off between the two penalties. We make the following natural choice for the penalty functions, Chandrasekaran et al. (2010, 2011) :

- the penalty function inducing sparsity is given by an l_1-like function: $\phi_1(Y) = h_1(Y) = \sum_{k>h} |Y_{(h,k)}|$, for any matrix $Y \in \mathbf{Q}_m$;
- the penalty function inducing low-rankness is the nuclear norm, which for positive semi-definite matrices is the trace, so that we let $\phi_*(Y) = \mathrm{tr}(Y)$.

3.2.2 Robust sparse plus low rank identification

In practical applications, however, the matrix Σ_m is unknown and needs to be estimated from the observed data that we assume to be N independent realizations $\mathbf{x}_i$ of $\mathbf{x}$. The typical choice is the sample covariance matrix $\hat{\Sigma}_m = N^{-1}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i^\top$. As discussed in the introduction to this chapter, however, when replacing Σ_m with $\hat{\Sigma}_m$ in (3.3) the corresponding solution may rapidly degrade. To deal with this problem we adopt a similar strategy to the one proposed in Chapter 2.
Let $\hat{\Sigma}_m$ be given. We assume that the "actual" Σ_m belongs to a ball centred in $\hat{\Sigma}_m$:

$$\mathscr{B} := \{\Sigma_m \in \mathbf{Q}_m \ s.t.\ \Sigma_m \succ 0, \mathbb{D}(\hat{\Sigma}_m \| \Sigma_m) \leq \delta/2\}$$

where $\delta/2$ is the prescribed tolerance and $\mathbb{D}(\hat{\Sigma}_m \| \Sigma_m)$ is the Kullback-Leibler divergence

$$\mathbb{D}(\hat{\Sigma}_m || \Sigma_m) := \frac{1}{2}(\log|\Sigma_m| - \log|\hat{\Sigma}_m| + \mathrm{tr}(\Sigma_m^{-1}\hat{\Sigma}_m) - m). \tag{3.4}$$

Therefore, we consider the following problem:

$$\begin{aligned} \underset{S,L,\Sigma_m \in \mathbf{Q_m}}{\arg\min} \quad & \operatorname{tr}(L) + \gamma h_1(S) \\ \text{subject to} \quad & L \succeq 0,\ S - L \succ 0, \\ & \Sigma_m^{-1} = S - L, \\ & 2\mathbb{D}(\hat{\Sigma}_m || \Sigma_m) \leq \delta. \end{aligned} \tag{3.5}$$

To streamline the notation let us denote by X the inverse of the matrix Σ_m, that is $X := \Sigma_m^{-1} = S - L$. Then, observing that in (3.5) we can eliminate, for example, the variable L, we get the equivalent minimization problem:

$$\begin{aligned} \underset{S,X \in \mathbf{Q_m}}{\arg\min} \quad & \operatorname{tr}(S - X) + \gamma h_1(S) \\ \text{subject to} \quad & S - X \succeq 0, \\ & X \succ 0, \\ & 2\mathbb{D}(\hat{\Sigma}_m || X^{-1}) \leq \delta \end{aligned} \tag{3.6}$$

where $2\mathbb{D}(\hat{\Sigma}_m || X^{-1})$ is given by $-\log|X| - \log|\hat{\Sigma}_m| + \operatorname{tr}(X\hat{\Sigma}_m) - m$.

Remark 3.2.1. In the majority of the applications in statistics, probability and information theory it is usual to consider $\mathbb{D}(\Sigma_m || \hat{\Sigma}_m)$ instead of $\mathbb{D}(\hat{\Sigma}_m || \Sigma_m)$, as has been done in Chapter 2 for the factor analysis problem. The choice of considering $\mathbb{D}(\hat{\Sigma}_m || \Sigma_m)$ in problem (3.6) is motivated by the fact that $\mathbb{D}(\hat{\Sigma}_m || \Sigma_m)$ is convex in $X = \Sigma_m^{-1}$.

3.2.3 Negative log-likelihood approach

Before entering into the study of problem (3.6), we discuss the connections of our problem with some approaches proposed in the literature.
First, we observe that a very natural alternative to problem (3.6) would be to consider a regularized minimization problem of the following type:

$$\begin{aligned} \underset{S,L \in \mathbf{Q}_m}{\arg\min} \quad & d(\hat{\Sigma}_m, S - L) + \lambda(\operatorname{tr}(L) + \gamma h_1(S)) \\ \text{subject to} \quad & S - L \succ 0, \\ & L \succeq 0 \end{aligned} \tag{3.7}$$

where $\lambda > 0$ is a regularization parameter and $d(\cdot,\cdot)$ is a distance or divergence or fitting function. In particular, in Chandrasekaran et al. (2010) a regularized negative log-likelihood has been considered. In formulation 3.7, however, two regularization parameters, λ and γ , are involved. Their optimal values have to be estimated by cross-validation. This requires a search over a two dimensional grid. On the contrary,

in our approach only one regularization parameter is involved and, therefore, only one dimensional grid is required for the cross-validation procedure with a consequent significant reduction in the number of candidate models that need to be computed.
Secondly, it is interesting to observe that there is a different route leading essentially to the same problem which may provide an interesting alternative interpretation of our proposed approach. To this end, let $\mathbf{X}_N := (\mathbf{x}_1, ..., \mathbf{x}_N)$ be an i.i.d. sample drawn from $\mathscr{P}(\Sigma_m) := \mathscr{N}(0, \Sigma_m)$. The log-likelihood function is

$$\log p(\mathbf{X}_N|\Sigma_m) = -\frac{Nm}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma_m| - \frac{1}{2}\sum_{k=1}^{N}\mathbf{x}_k^\top \Sigma_m^{-1}\mathbf{x}_k.$$

By using the so-called "trace-trick" we can rewrite it as: $\log p(\mathbf{X}_N|\Sigma_m) = \frac{N}{2}[m\log(2\pi) + \log|\Sigma_m| + \mathrm{tr}(\hat{\Sigma}_m\Sigma_m^{-1})]$. Thus, the negative log-likelihood is:

$$l(\mathbf{X}_N|\Sigma_m) = \frac{N}{2}[\log|\Sigma_m| + \mathrm{tr}(\hat{\Sigma}_m\Sigma_m^{-1}) + m\log(2\pi)]. \tag{3.8}$$

At this point observe that the Kullback-Leibler divergence (3.4) and the negative log-likelihood (3.8) differ only by a term not depending on Σ_m. Accordingly, imposing an upper bound δ on the Kullback-Leibler divergence is equivalent to impose an upper bound $\bar{l}$ on the desired negative log-likelihood. This leads to the equivalent problem:

$$\begin{aligned} \underset{S,X\in\mathbf{Q}_m}{\arg\min} \quad & \mathrm{tr}(S-X) + \gamma h_1(S) \\ \text{subject to} \quad & S - X \succeq 0, \\ & X \succ 0, \\ & l(\mathbf{X}_N|\Sigma_m) \leq \bar{l} \end{aligned}$$

where $\bar{l} := \frac{N}{2}(\delta + \log|\hat{\Sigma}| + m + m\log(2\pi))$.

3.2.4 An upper bound for δ

The allowed tolerance δ can be computed in a fashion similar to that of Section 2.3 by choosing a probability $\alpha \in (0,1)$ and a neighborhood of "radius" δ_α centered in $\hat{\Sigma}_m$ which contains the "true" Σ_m with probability α. In analogy with the factor analysis problem, if the level of α that has been chosen is too large with respect to the numerosity of the available data N, the computed δ_α may turn out to be excessively large so that there exist diagonal matrices Σ_m such that $2\mathbb{D}(\hat{\Sigma}_m||\Sigma_m) \leq \delta_\alpha$. In this case problem (3.6) admits the trivial solution $L = 0$ and S diagonal. From now on we assume that δ in (3.6) is strictly smaller than a certain upper bound δ_{max} that can be computed by solving the

minimization problem:

$$\delta_{max} := \min_{D \in \mathbf{D}_m,\, D \succ 0} 2\mathbb{D}(\hat{\Sigma}_m \| D). \tag{3.9}$$

The next result provides the solution to this problem.

Proposition 3.2.2. *The optimal D solving problem (3.9) is given by* $D^{opt} = diag^2(\hat{\Sigma}_m)$, *so that*

$$\delta_{max} = 2\mathbb{D}(\hat{\Sigma}_m \| D^{opt}) = \log|\hat{\Sigma}_m^{-1}\,\mathrm{diag}^2(\hat{\Sigma}_m)|.$$

The proof follows by mimicking the computation in the proof of Proposition 2.3.2 .

3.3 The dual problem

To derive the dual of problem (3.6) we introduce the Lagrangian:

$$\begin{aligned}\mathscr{L}(X,S,U,\lambda) =& \mathrm{tr}(S) - \mathrm{tr}(X) + \gamma h_1(S) - \mathrm{tr}(U(S-X)) \\ &+ \lambda\left(-\log|\hat{\Sigma}_m| - \log|X| + \mathrm{tr}(X\hat{\Sigma}_m) - m - \delta\right) \\ =& \langle S, I-U\rangle + \gamma h_1(S) + \langle X, U - I + \lambda\hat{\Sigma}_m\rangle \\ &+ \lambda\left(-\log|\hat{\Sigma}_m| - \log|X| - m - \delta\right)\end{aligned}$$

where $\lambda \in \mathbb{R}$, $\lambda \geq 0$, and $U \in \mathbf{Q}_m$, $U \succeq 0$ are the Lagrange multipliers. Then, the dual objective function is the infimum of $\mathscr{L}$ over X and S.

1.) Partial minimization over S: $\mathscr{L}$ depends on S solely through the terms

$$\gamma h_1(S) - \langle S, U - I\rangle. \tag{3.10}$$

The non-linear term does not depend on the elements in the main diagonal of S. Therefore, the minimization over the diagonal elements is unbounded from below unless $\mathrm{diag}^2(U - I) = 0$, i.e.

$$U_{(i,i)} = 1,\ i = 1, \ldots, m. \tag{3.11}$$

The minimization over the off-diagonal entries of S translates into an independent minimization of

$$-\left(U_{(i,j)}S_{(i,j)} + S_{(j,i)}U_{(j,i)}\right) + \gamma|S_{(i,j)}|$$

for each element i, j with $j > i$, which is unbounded from below unless:

$$2|U_{(i,j)}| \leq \gamma, \quad i \neq j. \tag{3.12}$$

If (3.12), (3.11) hold the infimum of (3.10) is equal to zero. Therefore, the partial

minimization of the Lagrangian over S is

$$\inf_S \mathscr{L} = \begin{cases} \langle X, \mathrm{ofd}(U) + \lambda \hat{\Sigma}_m \rangle - \lambda (\log|\hat{\Sigma}_m| + \log|X| + m + \delta) & \text{if (3.11), (3.12) hold;} \\ -\infty & \text{otherwise} \end{cases}$$

2.) Partial minimization over X: if (3.11) and (3.12) hold, $\mathscr{L}$ depends on X only through

$$\langle X, \mathrm{ofd}(U) + \lambda \hat{\Sigma}_m \rangle - \lambda \log|X|$$

which is bounded from below if and only if

$$\mathrm{ofd}(U) + \lambda \hat{\Sigma}_m \succ 0 \tag{3.13}$$

which implies

$$\lambda > 0. \tag{3.14}$$

If (3.13), (3.14) hold, by taking convexity into account, the matrix X achieving the minimum is easily obtained by annihilating the first derivative which yields

$$X = (\lambda^{-1} \mathrm{ofd}(U) + \hat{\Sigma}_m)^{-1} \tag{3.15}$$

and the minimum is given by

$$\lambda m + \lambda \log|\lambda^{-1} \mathrm{ofd}(U) + \hat{\Sigma}_m|.$$

Therefore, the result of the minimization of the Lagrangian over S and X is:

$$\inf_{S,X} \mathscr{L} = \begin{cases} \lambda \log|\lambda^{-1} \mathrm{ofd}(U) + \hat{\Sigma}_m| + \lambda(-\log|\hat{\Sigma}_m| - \delta) & \text{if (3.11), (3.12), (3.13), (3.14) hold;} \\ -\infty & \text{otherwise.} \end{cases}$$

Let us define the convex, closed and bounded set $\mathscr{U}$ as:

$$\mathscr{U} := \{U \in \mathbf{Q}_m : U \succeq 0, \mathrm{diag}^2(U) = I, 2|U_{(h,k)}| \leq \gamma, k \neq h\}.$$

Then, the dual problem is

$$\max_{(U,\lambda)\in\mathscr{C}} \quad \lambda \log|\lambda^{-1} \mathrm{ofd}(U) + \hat{\Sigma}_m| - \lambda(\log|\hat{\Sigma}_m| + \delta) \tag{3.16}$$

where the set $\mathscr{C}$ is defined as

$$\mathscr{C} := \{(\lambda, U) : U \in \mathscr{U}, \lambda > 0, (\lambda^{-1} \mathrm{ofd}(U) + \hat{\Sigma}_m) \succ 0\}.$$

For convenience we reformulate (3.16) as a minimization problem:

$$\min_{(\lambda,U)\in\mathscr{C}} \quad J(\lambda, U) \tag{3.17}$$

with $J(\lambda, U) := -\lambda(\log|\lambda^{-1}\operatorname{ofd}(U) + \hat{\Sigma}_m| - \log|\hat{\Sigma}_m| - \delta)$.

3.3.1 Existence of solutions

In this subsection we establish existence of solutions to problem (3.17) by showing that we can restrict the set $\mathscr{C}$ to a smaller compact set $\mathscr{C}_C$ over which the minimization problem is equivalent. Then, by continuity of the objective function over $\mathscr{C}$, and hence over $\mathscr{C}_C$, we conclude by Weierstrass's Theorem that J admits a minimum.

Lemma 3.3.1. *Let $(\lambda_k, U_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}$ such that*

$$\lim_{k\to\infty} \lambda_k = 0.$$

Then $(\lambda_k, U_k)_{k\in\mathbb{N}}$ is not an infimizing sequence for J.

Proof. We consider two possible cases separately.

First we analyze the case of sequences (λ_k, U_k) in which, besides $\lambda_k \to 0$, we also have $\|\lambda_k^{-1}\operatorname{ofd}(U_k)\| \to \infty$ as $k \to \infty$. This means that the largest singular value of $\lambda_k^{-1}\operatorname{ofd}(U_k)$ tends to $+\infty$, and this, by the symmetry of the matrix $\lambda_k^{-1}\operatorname{ofd}(U_k)$, implies that

$$\lim_{k\to\infty} \max_{\alpha_k \in \sigma(\lambda_k^{-1}\operatorname{ofd}(U_k))} |\alpha_k| = +\infty.$$

Since for all k $\operatorname{tr}(\lambda_k^{-1}\operatorname{ofd}(U_k)) = 0$, this in turn implies

$$\lim_{k\to\infty} \min_{\alpha_k \in \sigma(\lambda_k^{-1}\operatorname{ofd}(U_k))} \alpha_k = -\infty.$$

Therefore, for this type of sequences, and for a k sufficiently large , $\lambda_k^{-1}\operatorname{ofd}(U_k) + \hat{\Sigma}_m$ is no longer positive definite and therefore these sequences do not belong to the set $\mathscr{C}$.

Second, we consider sequences (λ_k, U_k) in which, besides $\lambda_k \to 0$, we also have

$$\|\lambda_k^{-1}\operatorname{ofd}(U_k)\| \to c$$

as $k \to +\infty$, where $0 \le c < \infty$. In this case, we show that $\forall \varepsilon > 0$, $\exists\, \bar{k}$ such that the dual functional satisfies $J(\lambda_k, U_k) > -\varepsilon$, $\forall k > \bar{k}$. Indeed, since $\|\lambda_k^{-1}\operatorname{ofd}(U_k)\|$ is bounded, $\exists\, l > 0$, $l \in \mathbb{R}$ such that, for all k, it holds that $\lambda_k^{-1}\operatorname{ofd}(U_k) + \hat{\Sigma}_m \le lI$. Therefore for all k,

$$|\lambda_k^{-1}\operatorname{ofd}(U_k) + \hat{\Sigma}_m| \le l^m,$$

$$\log|\lambda_k^{-1}\operatorname{ofd}(U_k) + \hat{\Sigma}_m| \le m\log l,$$

$$-\log|\lambda_k^{-1}\operatorname{ofd}(U_k)+\hat{\Sigma}_m|\geq -m\log l.$$

Thus, we can define the real constant (i.e. independent of k) $l_1 := m\log l - \log|\hat{\Sigma}_m| - \delta$ and, for all k, it holds that $J(\lambda_k,U_k)\geq -\lambda_k l_1$.
Since l_1 is constant $-\lambda_k l_1$ converges to zero so that, by definition, $\forall\varepsilon>0$, $\exists\,\bar{k}$ such that $-\lambda_k l_1 > -\varepsilon\ \forall k\geq\bar{k}$.
Now, it is sufficient to exhibit a couple $(\bar{\lambda},\bar{U})\in\mathscr{C}$ such that the dual functional is strictly negative to conclude that sequences (λ_k,U_k) cannot be infimizing sequences.
To this end, let us consider

$$\bar{U} := I+\bar{\lambda}[-\hat{\Sigma}_m+\operatorname{diag}^2(\hat{\Sigma}_m)]$$

with $\bar{\lambda}$ sufficiently small but strictly greater than zero, in this way $(\bar{\lambda},\bar{U})\in\mathscr{C}$. Then we have that $\operatorname{ofd}(\bar{U})=\bar{\lambda}[-\hat{\Sigma}_m+\operatorname{diag}^2(\hat{\Sigma}_m)]$ and therefore

$$J(\bar{\lambda},\bar{U})=-\bar{\lambda}(\delta_{max}-\delta)<0.$$

This suffices to conclude the proof. Indeed, the only other possible case is that for which $\lim_{k\to\infty}\|\lambda_k^{-1}\operatorname{ofd}(U_k)\|$ does not exist. However, in this case it is always possible to consider a sub-sequence (λ_{k_j},U_{k_j}) for which the corresponding limit does exist (finite or not) and thus we can reduce this case to one of the previous two. ■

As a consequence, minimizing the dual functional over the set $\mathscr{C}$ is equivalent to minimize over the set:

$$\mathscr{C}_1 := \{(\lambda,U): U\in\mathscr{U},\lambda\geq\varepsilon,(\lambda^{-1}\operatorname{ofd}(U)+\hat{\Sigma}_m)\succ 0\}$$

for a certain $\varepsilon>0$.

The next result allows to further restrict $\mathscr{C}_1$ to a set where λ is bounded.

Lemma 3.3.2. *Let $(\lambda_k,U_k)_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_1$ such that*

$$\lim_{k\to\infty}\lambda_k=\infty.$$

Then $(\lambda_k,U_k)_{k\in\mathbb{N}}$ cannot be an infimizing sequence for J.

Proof. Let us consider a sequence such that λ_k tends to $+\infty$ when $k\to+\infty$. Then $\|\lambda_k^{-1}\operatorname{ofd}(U_k)\|\to 0$, because U is bounded element-wise. Therefore

$$\lim_{k\to+\infty}J\to+\infty.$$

Therefore, such a sequence cannot be an infimizing sequence. ■

As a consequence, we can further restrict set $\mathscr{C}_1$ to the set:

$$\mathscr{C}_2 := \{(\lambda, U) : U \in \mathscr{U}, \xi \geq \lambda \geq \varepsilon, (\lambda^{-1}\,\mathrm{ofd}(U) + \hat{\Sigma}_m) \succ 0\}$$

for a certain $\xi > 0$.
Finally, we consider a sequence $(\lambda_k, U_k)_{k\in\mathbb{N}}$ such that as $k \to \infty$ the minimum eigenvalue of $\lambda_k^{-1}\,\mathrm{ofd}(U_k) + \hat{\Sigma}_m$ tends to zero. This implies $|\lambda_k^{-1}\,\mathrm{ofd}(U_k) + \hat{\Sigma}_m| \to 0$ and hence $J \to +\infty$. Therefore, such a sequence cannot be an infimizing sequence.
In conclusion, we can restrict our search for the optimal solution to the following *compact* set

$$\mathscr{C}_C := \{(\lambda, U) : U \in \mathscr{U}, \xi \geq \lambda \geq \varepsilon, (\lambda^{-1}\,\mathrm{ofd}(U) + \hat{\Sigma}_m) \succeq \beta I\}$$

for a certain $\beta > 0$.

Theorem 3.3.3. *Problem* (3.17) *admits a solution.*

Proof. Since $\mathscr{C}_C$ is compact and J is continuous over $\mathscr{C}$ and hence over $\mathscr{C}_C$, by Weierstrass's Theorem the minimum exists. ∎

Remark 3.3.4. Clearly, the dual objective function J is bounded from below as the optimal value of the dual (maximization) problem is upper-bounded by the optimal value of the primal problem which is finite (as can be seen for instance by choosing $\Sigma = \hat{\Sigma}$ which, by assumption, is positive definite and bounded element wise).

3.3.2 Uniqueness of the solution

J is the opposite of the dual objective function hence J is convex over $\mathscr{C}_C$. However, as we will show, J is not strictly convex. Therefore establishing the uniqueness of the minimum is not a trivial task.
To streamline the notation, let us define $\tilde{U} := \mathrm{ofd}(U)$. Then, the following proposition characterizes the second variation of J in direction $(\delta\lambda, \delta\tilde{U})$, i.e. $\delta^2 J(\lambda, \tilde{U}; \delta\lambda, \delta\tilde{U})$.

Proposition 3.3.5. *Let* $\tilde{\mathbf{u}} = \mathrm{vec}(\tilde{U})$, $\delta\tilde{\mathbf{u}} := \mathrm{vec}(\delta\tilde{U})$, *and* $K := (\lambda^{-1}\tilde{U} + \hat{\Sigma}_m)^{-1} \otimes (\lambda^{-1}\tilde{U} + \hat{\Sigma}_m)^{-1}$. *Let also*

$$H := \begin{bmatrix} \lambda^{-3}\tilde{\mathbf{u}}^\top K\tilde{\mathbf{u}} & -\lambda^{-2}\tilde{\mathbf{u}}^\top K \\ -\lambda^{-2}K\tilde{\mathbf{u}} & \lambda^{-1}K \end{bmatrix} \in \mathbb{R}^{(1+n^2)\times(1+n^2)}.$$

Then, we have

$$\delta^2 J(\lambda, \tilde{U}; \delta\lambda, \delta\tilde{U}) = [\delta\lambda \;\; \delta\tilde{\mathbf{u}}^\top] H \begin{bmatrix} \delta\lambda \\ \delta\tilde{\mathbf{u}} \end{bmatrix}.$$

Proof. Consider the function

$$F(\lambda,\tilde{U}) = -\lambda \log|\lambda^{-1}\tilde{U} + \hat{\Sigma}_m|.$$

$F(\lambda,\tilde{U})$ differs from $J(\lambda,\tilde{U})$ only by terms which are linear in $(\lambda,\tilde{U})$ therefore the second variations of the two functions are equal. In the rest of the proof we will focus on $F(\lambda,\tilde{U})$. The first variation of $F(\lambda,\tilde{U})$ in direction $(\delta\lambda,\delta\tilde{U})$ is

$$\begin{aligned}\delta F(\lambda,\tilde{U};\delta\lambda,\delta\tilde{U}) = &-\log|\lambda^{-1}\tilde{U}+\hat{\Sigma}_m|\delta\lambda \\ &+\lambda^{-1}\mathrm{tr}((\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\tilde{U})\delta\lambda - \mathrm{tr}((\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\delta\tilde{U}).\end{aligned}$$

The second variation of $F(\lambda,\tilde{U})$ in direction $(\delta\lambda,\delta\tilde{U})$ is

$$\begin{aligned}&\delta^2 F(\lambda,\tilde{U};\delta\lambda,\delta\tilde{U}) = \\ &\lambda^{-1}\mathrm{tr}\big((\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\delta\tilde{U}(\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\delta\tilde{U}\big) \\ &-2\left[\lambda^{-2}\mathrm{tr}\big((\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\delta\tilde{U}(\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\tilde{U}\big)\delta\lambda\right] \\ &+\lambda^{-3}\mathrm{tr}\big((\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\tilde{U}(+\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\tilde{U}\big)\delta\lambda^2.\end{aligned}$$

Now, by using the Kronecker product and the vec operator and defining $\tilde{\mathbf{u}} := \mathrm{vec}(\tilde{U})$, $\delta\tilde{\mathbf{u}} := \mathrm{vec}(\delta\tilde{U})$, and $K := (\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}\otimes(\lambda^{-1}\tilde{U}+\hat{\Sigma}_m)^{-1}$ the Hessian in Proposition 3.3.5 immediately follows. ∎

Corollary 3.3.6. *The functional J is convex and for any point $(\lambda_0,\tilde{U}_0)$ there is exactly one direction along which it is not strictly convex. This direction is*

$$(\delta\lambda_0,\delta\tilde{U}_0) = (h\lambda_0,h\tilde{U}_0), \quad \textit{with } h\neq 0. \tag{3.18}$$

Proof. Since in $\mathscr{C}$ we have that $K\in\mathbf{Q}_{m^2}$ is positive definite and $\lambda>0$, the Hessian matrix H has at least rank equal to n^2. Hence there is at most one direction along which J is not strictly convex. It is easy to check that the second variation along the direction (3.18) is zero. ∎

The next result establishes that if at a certain point the functional J is constant along an arbitrary direction then J vanishes at this point.

Lemma 3.3.7. *Let $(\lambda_0,\tilde{U}_0)$ be a given point in the feasible set $\mathscr{C}$. If $\mathbf{w} := (\delta\lambda,\delta\tilde{U}) \neq (0,0)$ is any direction along which $J(\lambda_0,\tilde{U}_0)$ is constant, that is if there exists $\varepsilon>0$ such that $f(\alpha) := J(\lambda_0+\alpha\delta\lambda,\tilde{U}_0+\alpha\delta\tilde{U})$ is constant for any α such that $|\alpha|<\varepsilon$, then $J(\lambda_0,\tilde{U}_0)=0$.*

Proof. By assumption we have that $f(\alpha)$ is constant in a neighbourhood of zero. Hence the first derivative $f'(0)$, and thus the second derivative $f''(0)$, must vanish and hence the second variation of $f(0)$ in direction 1 is zero. On the other hand this second variation

is, by definition, the second variation $\delta^2 J(\lambda_0, \tilde{U}_0, \delta\lambda, \delta\tilde{U})$. Hence this second variation vanishes and, by Corollary 3.3.6, this implies that $(\delta\lambda, \delta\tilde{U}) = (h\lambda_0, h\tilde{U}_0)$, for a certain real constant h. Hence, for $|\alpha|$ sufficiently small, we have

$$J(\lambda_0, \tilde{U}_0) = f(0) = f(\alpha) = J((1+\alpha h)\lambda_0, (1+\alpha h)\tilde{U}_0). \tag{3.19}$$

By direct computation we get $J((1+\alpha h)\lambda_0, (1+\alpha h)\tilde{U}_0) = (1+\alpha h)J(\lambda_0, \tilde{U}_0)$ which together with (3.19) yields the conclusion. ∎

We are now ready to state our main result.

Theorem 3.3.8. *The dual problem* (3.17) *admits a unique solution.*

Proof. By contradiction, assume that there exist two optimal solutions $(\lambda_1^\circ, \tilde{U}_1^\circ)$ and $(\lambda_2^\circ, \tilde{U}_2^\circ)$. By the convexity of the set $\mathscr{C}$, the whole segment $\mathcal{S}$ connecting $(\lambda_1^\circ, \tilde{U}_1^\circ)$ to $(\lambda_2^\circ, \tilde{U}_2^\circ)$ must belong to $\mathscr{C}$. Then, by the convexity of $J(\cdot,\cdot)$ all the points in $\mathcal{S}$ are optimal solutions so that $J(\cdot,\cdot)$ is constant in $\mathcal{S}$. In view of Lemma 3.3.7 this implies that $J(\cdot,\cdot)$ is zero in $\mathcal{S}$ and this is a contradiction since as seen in the proof of Lemma 3.3.1, the optimal value of J is negative. ∎

Corollary 3.3.9. *Any optimal solution* $(\lambda^\circ, \tilde{U}^\circ)$ *minimizing J over* $\mathscr{C}$ *lies on the boundary of* $\mathscr{C}$.

Proof. Let $(\lambda^\circ, \tilde{U}^\circ)$ be an optimal solution and , by contradiction, assume that $(\lambda^\circ, \tilde{U}^\circ)$ does not belong to the boundary of $\mathscr{C}$. Then there exists $\varepsilon > 0$ such that

$$((1+\varepsilon)\lambda^\circ, (1+\varepsilon)\tilde{U}^\circ) \in \mathscr{C}.$$

Now by direct computation

$$J((1+\varepsilon)\lambda^\circ, (1+\varepsilon)\tilde{U}^\circ) = (1+\varepsilon)J(\lambda^\circ, \tilde{U}^\circ) < J(\lambda^\circ, \tilde{U}^\circ) \tag{3.20}$$

where the last inequality follows from the fact that, as seen in the proof of Lemma 3.3.1, the optimal value of J is negative. This a contradiction since $J(\lambda^\circ, \tilde{U}^\circ)$ is assumed to be a minimum. ∎

3.4 Recovering the solution of the primal problem

The optimal X° can be easily recovered by substituting the optimal solution of the dual problem (λ°, U°) into (3.15).
To identify the sparsity pattern of S we observe that the minimization of (3.10) under the constraints (3.11) and (3.12) is equivalent to an independent minimization of the non-negative functions

$$|S_{(i,j)}|(\gamma - 2|U_{(i,j)}|), \qquad i > j. \tag{3.21}$$

Since the optimal value of (3.10) is equal to zero, then the optimal value of (3.21) is equal to zero for each $i > j$. Thus $(\gamma - 2|U_{(i,j)}|) > 0$ implies that the corresponding entries $S_{(i,j)} = S_{(j,i)}$ are equal to zero and the set of zero entries of S can be defined as:

$$\mathcal{I} := \{(k,h) : k \neq h,\ 2|U_{(k,h)}| < \gamma\}.$$

Recovering the non-zero entries of S°, and therefore L°, is slightly more involved. At this point we observe that the primal problem is strictly feasible (which can be seen by taking, for instance, $\Sigma = \hat{\Sigma}$) and hence, by virtue of the properties of the primal and dual problems, the following extremality condition holds, see e.g. (Ekeland and Temam, 1999, Theorem 5.1):

$$\mathrm{tr}(U^\circ(S^\circ - X^\circ)) = \mathrm{tr}(U^\circ L^\circ) = 0.$$

Note that, if U° is full rank, which can be always the case for a value of γ sufficiently small, then the unique solution is $L^\circ = S^\circ - X^\circ = 0$.
If instead U° is not full rank, we can consider the following reduced singular value decomposition

$$U^\circ = VDV^\top,$$

where $D \in \mathbf{Q}_{m-r}$, $V \in \mathbb{R}^{m\times(m-r)}$ is such that $V^\top V = I_{m-r}$ and r is the estimated optimal rank of L°. Then

$$\mathrm{tr}(VDV^\top L^\circ) = 0 \Rightarrow V^\top L^\circ V = 0.$$

Therefore, we select a matrix $\tilde{V}$ such that its columns form an orthogonal basis of $[\mathrm{Im}(V)]^\perp$ and we express L° as

$$L^\circ = \tilde{V}Q\tilde{V}^\top$$

where $Q \in \mathbf{Q}_{m-r}$ is the unknown. Finally, Q can be obtained by solving the following system of linear equations

$$(\tilde{V}Q\tilde{V}^\top)_{(k,h)} = -X^\circ_{(k,h)}, \qquad \forall (k,h) \in \mathcal{I}.$$

Note that this is a system of $|\mathcal{I}|/2$ equations with $r(r+1)/2$ unknowns, that is the number of independent parameters in Q. If the system admits a solution and this solution is unique then the system yields the solution of the primal problem which is also unique.

3.5 Numerical Implementation

To solve numerically the dual problem we rely on an Alternating Direction Method of Multipliers (ADMM) algorithm similar to the one proposed in the previous chapter.
In particular, the constraint $U \succeq 0$ is decoupled from the constraints $\mathrm{diag}^2(U) = I_m$ and

$2|U_{(h,k)}| < \gamma$, $h \neq k$, by introducing a new variable $Y := U$. Hence, the problem becomes

$$\begin{aligned} \min_{(\lambda,U)\in\mathscr{C}_{\lambda,U}, Y\in\mathscr{C}_Y} \quad & J(\lambda,U) \\ \text{subject to} \quad & Y = U \end{aligned}$$

where

$$\begin{aligned} \mathscr{C}_{\lambda,U} := \{(\lambda,U) : \lambda > 0, U \in \mathbf{Q}_m, \operatorname{diag}^2(U) = I, 2|U_{(h,k)}| \leq \gamma, h \neq k, \\ (\lambda^{-1}\operatorname{ofd}(U) + \hat{\Sigma}_m) \succ 0\} \\ \mathscr{C}_Y := \{Y \in \mathbf{Q}_m : Y \succeq 0\} \end{aligned}$$

The *augmented Lagrangian* for this problem is

$$\mathscr{L}_\rho(\lambda,U,Y,M) = \lambda(-\log|\lambda^{-1}\operatorname{ofd}(U) + \hat{\Sigma}_m| + \log|\hat{\Sigma}_m| + \delta) + \langle M, Y-U\rangle + \frac{\rho}{2}\|Y-U\|_F^2$$

and the ADMM updates are

$$(\lambda^{(k+1)}, U^{(k+1)}) = \operatorname*{arg\,min}_{(\lambda,U)\in\mathscr{C}_{\lambda,U}} \mathscr{L}_\rho(\lambda,U,Y^{(k)},M^{(k)}) \tag{3.22}$$

$$Y^{(k+1)} = \operatorname*{arg\,min}_{Y\in\mathscr{C}_Y} \mathscr{L}_\rho(\lambda^{(k+1)},U^{(k+1)},Y,M^{(k)}) \tag{3.23}$$

$$M^{(k+1)} = M^{(k)} + \rho(Y^{(k+1)} - U^{(k+1)}). \tag{3.24}$$

Problem (3.22) does not admit a closed form solution, therefore to solve it we rely on a gradient projection method whose updates are:

$$\begin{aligned} \lambda^{(k+1)} &= \lambda^{(k)} - t_k \nabla_\lambda \mathscr{L}_\rho(\lambda^{(k)},U^{(k)},Y^{(k)},M^{(k)}) \\ U^{(k+1)} &= \Pi_{\mathrm{U}}(U^{(k)} - t_k \nabla_U \mathscr{L}_\rho(\lambda^{(k)},U^{(k)},Y^{(k)},M^{(k)})) \end{aligned}$$

where the step size t_k is chosen according to Armijo's conditions, see Boyd and Vandenberghe (2004), $\nabla_\lambda \mathscr{L}_\rho$ and $\nabla_U \mathscr{L}_\rho$ are the gradients with respect to λ and U, respectively, and Π_{U} is the projector:

$$[\Pi_{\mathrm{U}}(A)]_{(i,j)} = \begin{cases} 1 & \text{if } i = j \\ A_{(i,j)} & \text{if } i \neq j \text{ and } 2|A_{(i,j)}| \leq \gamma \\ \gamma & \text{if } i \neq j \text{ and } 2|A_{(i,j)}| > \gamma. \end{cases}$$

Problem (3.23) admits the closed form solution

$$Y^{(k+1)} = \Pi_{\mathscr{C}_Y}\left(-\frac{1}{\rho}M + U\right)$$

where $\Pi_{\mathscr{C}_Y}$ is the projector onto the cone of symmetric positive semi-definite matrices of size $m \times m$. The following quantities are used to define the stopping criterion

$$R^P = Y^{(k)} - U^{(K)}$$
$$R^D = \rho(Y^{(k)} - Y^{(k-1)}).$$

The algorithm stops when the following criteria are met

$$\|R^P\|_F \leq m\varepsilon^{abs} + \varepsilon^{rel} \max\{\sqrt{m}, \|U^{(k)}\|_F, \|Y^{(k)}\|_F\}$$
$$\|R^D\|_F \leq m\varepsilon^{abs} + \varepsilon^{rel} \|M^{(k)}\|_F$$

where ε^{abs} and ε^{rel} are the desired absolute and relative tolerances.

3.5.1 Numerical Example

In this subsection we present an illustrative numerical example. To this end we consider a latent variable graphical model with 10 manifest variables and one latent variable so that the concentration matrix of the manifest variables admits a decomposition $\Sigma_m^{-1} = S - L$ with $\text{rank}(L) = 1$. The sparsity pattern of the matrix S is displayed in Figure 3.3 on the left. From this model we have generated a sample of size $N = 1000$ and we have computed the sample covariance $\hat{\Sigma}_m$. The parameter δ_α has been computed as detailed in the previous chapter for a value of $\alpha = 0.5$. To implement the ADMM algorithm we have set $\varepsilon^{rel} = \varepsilon^{abs} = 10^{-3}$. The parameter ρ has been updated according to the following scheme: $\rho^{(k+1)} = \min\{1.1\rho^{(k)}, 10^3\}$ with $\rho^{(0)} = 1$.

The results are shown in Figure 3.3 for different values of γ. The optimal model has been selected via cross validation procedure and has been highlighted.

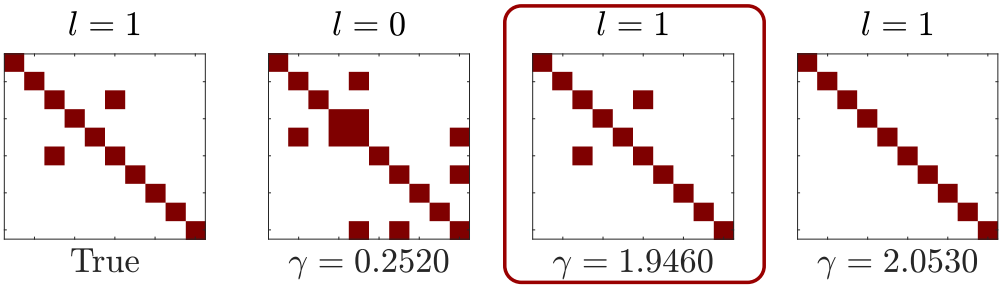


Figure 3.3: Sparsity pattern and number of latent variables l for the "true" model (left) versus sparsity pattern and number of latent variables estimated with the proposed method for $\gamma \in \{0.2520, 1.9460, 2.0530\}$ (right). The optimal model is highlighted by a red square.

3.6 Concluding remarks and future directions

In this chapter the problem of robust latent variable graphical model identification has been considered. In particular, the *Sparse plus Low Rank* decomposition problem has been reformulated for the case in which only the sample covariance is available and the difference between the sample covariance and the actual one is non-negligible.
A first natural extension of the analysis presented in this chapter is to the case of structured covariance matrices such as, for example, circulant and block-circulant covariance matrices which find applications in modelling periodic and multivariate-periodic stationary processes. For the special case in which the inverse covariance is banded block-circulant this leads to the problem of identifying latent variable *reciprocal* graphical models proposed in Alpago, Zorzi, and Ferrante (2018b) where a negative log-likelihood approach has been considered. In this regard, the approach proposed in this chapter may be an appealing alternative as only one regularization parameter is needed.
The extension of the analysis presented in this chapter to the dynamical (autoregressive) case is the subject of the next chapter.

4

Learning latent variable dynamic graphical models with confidence sets

4.1 Introduction

In modern society where everything is more and more interconnected, the importance of learning dynamical networks is booming and the open problems as well as the applications challenged by the network framework are countless. The present chapter deals with the identification of dynamic Markov networks – also known as dynamic graphical models, Lauritzen (1996) – aiming to generalize the philosophy and the estimation paradigm proposed in Chapter 3 to the dynamical framework. In analogy with the static case, one of the key issues in learning dynamic graphical models is the fact that often we receive data from a large number of possibly interconnected systems and we are interested in extracting a network model with a limited number of edges that highlights the relevant interconnections thus providing a powerful tool for data interpretation. To this end a regularized maximum likelihood approach has been proposed in Songsiri and Vandenberghe (2010); Songsiri, Dahl, and Vandenberghe (2010) for the case of autoregressive graphical models which allows for a remarkable matrix formulation. This problem has also been considered from several other perspectives: an automatic procedure for selecting the complexity of the autoregressive graphical model has been proposed in Maanan, Dumitrescu, and Giurcăneanu (2017); in Alpago, Zorzi, and Ferrante (2018a) an approach based on reciprocal processes has been considered; a Bayesian formulation has been addressed in Zorzi (2019). These approaches are strictly connected to a maximum entropy spectral estimation problem, Avventi, Lindquist, and

Wahlberg (2013), and belong to the family of moment problems extensively studied in the past years, Byrnes et al. (2000); Georgiou and Lindquist (2003); Byrnes, Enqvist, and Lindquist (2002); Ferrante et al. (2012); Zorzi (2014); Pavon and Ferrante (2013).
As for their static counterpart, sparse dynamic networks fail to provide a good interpretation in the case when the data come from dynamical systems that are genuinely highly interconnected. On the other hand, in many practical situations, most of the interconnections can be explained by a common behaviour that can be modelled by a small number of latent dynamical systems. Therefore, in these cases, at the expenses of a modest increment of the number of nodes, we can obtain a sparse network modelling our data. This situation is described by latent variable graphical models as detailed in Chapter 3. A regularized maximum likelihood method for learning such networks has been proposed, for the autoregressive case, in Zorzi and Sepulchre (2016), see also Maanan, Dumitrescu, and Giurcăneanu (2018); Alpago et al. (2018b); Liégeois, Mishra, Zorzi, and Sepulchre (2015). We recall that with this approach two regularizers are needed: one to induce a small number of latent nodes and the other to induce sparsity in the network.
In this chapter we propose a novel robust estimation paradigm for learning autoregressive latent variable graphical models. More precisely, given a finite length realization of the data generating process, we estimate its spectral density and we account for the uncertainty in the estimation by considering a "confidence neighbourhood" centered in the computed estimate which contains the true spectrum with an user-chosen probability. The "radius" of this confidence neighbourhood depends only on the number of data. Then, in this neighbourhood we search for the model that minimizes (a trade-off between) the number of latent nodes and the number of edges in the network. The resulting optimization problem involves only one regularization parameter balancing the trade-off between the number of latent nodes and the sparseness of the learned graph. This problem can be naturally formulated in terms of spectral densities. We introduce a convenient matricial re-parametrization and we derive the dual problem which can be efficiently solved numerically with an ADMM-type algorithm. A numerical simulation at the end of the chapter confirms that the our approach is particularly convenient in terms of computational burden. As a by-product of our analysis two ancillary results of independent interest are obtained: the first can be seen as a generalization of the well-known Wiener-Masani formula, and the second establishes a property of blocks matrices with diagonal blocks.

Outline of the chapter

The chapter is organized as follows. Section 4.2 we describe our model class, namely autoregressive graphical models with latent variables. We introduce a mathematical formulation of our problem in Section 4.3 and in Section 4.4 we show how to compute the "confidence neighbourhood" where the optimal solution is sought. A more convenient matrix reformulation of our problem is introduced in Section 4.5: the equivalence between

the two is shown in Section 4.7 after the variational analysis. In Section 4.6 we derive the dual problem which admits solutions as shown in Subsection 4.6.1. Uniqueness is addressed in Section 4.7. In Section 4.8 we show how to recover the solution of the primal problem. Then in Section 4.9 an ADMM algorithm is proposed to solve the dual problem. This algorithm is employed in Section 4.10 where a numerical simulation shows the effectiveness of the proposed method. Some concluding remarks are provided in the final section.

Notation and Conventions

Throughout this chapter we make use of the following notations and conventions. The symbol $\mathbf{M}_{m,n}$ denotes the vector space of matrices of the form

$$Y := [Y_0 \quad Y_1 \quad \dots \quad Y_n], \quad Y_0 \in \mathbf{Q}_m, \ Y_1, \dots, Y_n \in \mathbb{R}^{m\times m}, \tag{4.1}$$

and $\mathbf{Q}_{m(n+1)}$ denotes the space of symmetric block-matrices with $(n+1)\times(n+1)$ blocks of dimension $m\times m$; if $X \in \mathbf{Q}_{m(n+1)}$, X_{ij} is the block in position i, j with $i, j = 0, \dots, n$, so that

$$X = \begin{bmatrix} X_{00} & X_{01} & \dots & X_{0n} \\ X_{01}^\top & X_{11} & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X_{0n}^\top & X_{1n}^\top & \dots & X_{nn} \end{bmatrix}.$$

The linear mapping $T : \mathbf{M}_{m,n} \to \mathbf{Q}_{m(n+1)}$ constructs a symmetric block-Toeplitz matrix from its first block row so that if Y is given by (4.1),

$$T(Y) = \begin{bmatrix} Y_0 & Y_1 & \dots & Y_n \\ Y_1^\top & Y_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Y_1 \\ Y_n^\top & \dots & Y_1^\top & Y_0 \end{bmatrix}.$$

The adjoint of T is a mapping $D : \mathbf{Q}_{m(n+1)} \to \mathbf{M}_{m,n}$ defined by

$$D(X) = [[D(X)]_0 \quad \dots \quad [D(X)]_n]$$

with

$$[D(X)]_0 = \sum_{h=0}^{n} X_{hh}, \quad [D(X)]_j = 2\sum_{h=0}^{n-j} X_{hh+j}, \ j = 1, \dots, n.$$

We denote by $(\cdot)^*$ the complex conjugate transpose. Functions defined on the unit circle $\{e^{i\theta} : \theta \in [-\pi, \pi]\}$ are denoted by capital Greek letters, e.g. $\Phi(e^{i\theta})$ for $\theta \in [-\pi, \pi]$, and the dependence on θ is dropped if needed. Integrals are always defined from $-\pi$

to π with respect to the normalized Lebesgue measure $d\theta/2\pi$. If $\Phi(e^{i\theta})$ is positive (semi-) definite $\forall\theta \in [-\pi,\pi]$ we write $\Phi(e^{i\theta}) \succ 0$ ($\Phi(e^{i\theta}) \succeq 0$). $\mathscr{S}_m^+$ denotes the space of $m \times m$ matrix-valued spectral densities defined on the unit circle $\{e^{i\theta} : \theta \in [-\pi,\pi]\}$ which are bounded and coercive (i.e. positive definite in the unit circle). Finally, $\mathscr{Q}_{m,n} := \{\sum_{j=-n}^{n} e^{-ij\theta} R_j, \text{ s.t. } R_j = R_{-j}^\top \in \mathbb{R}^{m\times m}\}$ is the set of Hermitian pseudo-polynomial matrices of order n, and $\Delta(e^{i\theta})$ is the so-called shift operator:

$$\Delta(e^{i\theta}) := [I_m \quad e^{i\theta} I_m \quad \dots \quad e^{in\theta} I_m]. \tag{4.2}$$

4.2 Autoregressive latent variable graphical models

4.2.1 Autoregressive models

An $m-$dimensional, zero-mean, stationary, full rank, Gaussian process $\mathbf{x} = \{\mathbf{x}(t) : t \in \mathbb{Z}\}$ is an autoregressive process of order n if it satisfies the equations:

$$\mathbf{x}(t) = -\sum_{j=1}^{n} A_j \mathbf{x}(t-j) + \mathbf{w}(t)$$

where $\mathbf{w}(t) \sim \mathscr{N}(0,\Sigma_{\mathbf{w}})$ and $A_j \in \mathbb{R}^{m\times m}$. Let $R_j := \mathbb{E}\{\mathbf{x}(t+j)\mathbf{x}(t)^\top\}$ be the j-th covariance lag of $\mathbf{x}$ (recall that $R_{-j} = R_j^\top$) and define $A := [I\ A_1^\top\ \dots\ A_n^\top]$. The process $\mathbf{x}$ is characterized by its spectral density $\Phi(e^{i\vartheta}) \in \mathscr{S}_m^+$:

$$\Phi(e^{i\vartheta}) := \sum_{j=-\infty}^{\infty} e^{-ij\vartheta} R_j = (\Delta A^\top)^{-1} \Sigma_{\mathbf{w}} (A\Delta^*)^{-1}.$$

Hence, the inverse spectrum is a pseudo-polynomial matrix, $\Phi^{-1} \in \mathscr{Q}_{m,n}$:

$$\Phi(e^{i\vartheta})^{-1} = A\Delta^* \Sigma_{\mathbf{w}}^{-1} \Delta A^\top = Y_0 + \sum_{j=1}^{n} (e^{-ij\theta} Y_j + e^{ij\theta} Y_j^\top)$$

where $Y_j = \sum_{i=0}^{n-j} A_i^\top \Sigma_{\mathbf{w}}^{-1} A_{i+j}^\top$, with $A_0 = I$.

4.2.2 Latent variable dynamic graphical models

Let $\mathbf{x} := \{\mathbf{x}(t) : t \in \mathbb{Z}\}$ be a real valued, $(m+l)$-dimensional, zero-mean, stationary, full rank, Gaussian process defined on a probability space $(\Omega, \mathscr{A}, P)$. We assume $\mathbf{x}$ to be partitioned as follows:

$$\mathbf{x} = [(\mathbf{x}^m)^\top (\mathbf{x}^l)^\top]^\top$$

where $\mathbf{x}^m = [x_1, ..., x_m]^\top$ contains the manifest variables and $\mathbf{x}^l = [x_{m+1}, ..., x_{m+l}]^\top$ contains the latent variables. Accordingly, the spectral density of the process $\mathbf{x}$, $\Phi_{m+l} \in \mathscr{S}^+_{l+m}$, and its inverse can also be partitioned as:

$$\Phi_{m+l} := \begin{bmatrix} \Phi_m & \Phi^*_{lm} \\ \Phi_{lm} & \Phi_l \end{bmatrix}, \quad \Phi^{-1}_{m+l} := \begin{bmatrix} \Psi_m & \Psi^*_{lm} \\ \Psi_{lm} & \Psi_l \end{bmatrix}. \tag{4.3}$$

Let $\mathcal{I} := \{1, \ldots, m+l\}$ and denote by $\chi_\mathcal{I}$ the closure in $\mathcal{L}^2(\Omega, \mathscr{A}, P)$ of the set containing all the finite linear combinations (with real coefficients) of the variables $x_k(t)$ with $t \in \mathbb{Z}$ and $k \in \mathcal{I}$:

$$\chi_\mathcal{I} := \overline{\text{span}}\{x_k(t),\ t \in \mathbb{Z},\ k \in \mathcal{I}\}.$$

We introduce the notation $\chi_{\{i\}} \perp \chi_{\{j\}} | \chi_{\mathcal{I}\setminus\{i,j\}}$ to indicate that $\chi_{\{i\}}$ and $\chi_{\{j\}}$ are conditionally independent given $\chi_{\mathcal{I}\setminus\{i,j\}}$, Dahlhaus (2000); Avventi et al. (2013); Zorzi and Sepulchre (2016). The following relation holds, Dahlhaus (2000); Avventi et al. (2013):

$$[\Phi^{-1}_{m+l}(e^{i\theta})]_{(i,j)} = 0,\ \forall \theta \in [-\pi, \pi] \Leftrightarrow \chi_{\{i\}} \perp \chi_{\{j\}} | \chi_{\mathcal{I}\setminus\{i,j\}}.$$

Such conditional dependence relations can be represented by means of an undirected graph $\mathscr{G}(V_{m+l}, E_{m+l})$ where nodes V_{m+l} correspond to the random variables $x_1, ..., x_{m+l}$ and edges E_{m+l} represent conditionally dependence relations:

$$(i,j) \notin E_{m+l} \Leftrightarrow \chi_{\{i\}} \perp \chi_{\{j\}} | \chi_{\mathcal{I}\setminus\{i,j\}}.$$

Assume now that the number of latent variables is small compared to the number of manifest variables, i.e. $l \ll m$, and that the manifest variables are mainly correlated through the latent variables. In Zorzi and Sepulchre (2016) it has been observed that the associated graphical structure is equivalent to the fact that the inverse power spectral density $\Phi^{-1}_m \in \mathscr{S}^+_m$ of the manifest process (the only one we can actually observe) admits a "sparse plus low-rank" decomposition. Indeed, by taking the point-wise Schur complement of Ψ_l in Φ^{-1}_{m+l} in (4.3) we obtain the following relation:

$$\Phi^{-1}_m = \Psi_m - \Psi^*_{lm} \Psi^{-1}_l \Psi_{lm}$$

where the sparsity pattern of Ψ_m reflects the conditional independence relations among the manifest variables while the rank of $\Psi^*_{lm} \Psi^{-1}_l \Psi_{lm}$ gives an upper bound on the number of latent components. Then, by identifying Ψ_m with a sparse spectrum $\Sigma \in \mathscr{Q}_{m,n}$, and $\Psi^*_{lm} \Psi^{-1}_l \Psi_{lm}$ with a low rank spectrum $\Lambda \in \mathscr{Q}_{m,n}$, $\Lambda \succeq 0$ one obtains the following "sparse plus low-rank" decomposition:

$$\Phi^{-1}_m = \Sigma - \Lambda. \tag{4.4}$$

4.3 Problem formulation

Assume to collect the data $\mathbf{X}^{m,N} := \{\mathbf{x}^m(1) \dots \mathbf{x}^m(N)\}$ generated by a manifest autoregressive process $\mathbf{x}^m$. We want to estimate the spectral density Φ_m of $\mathbf{x}^m$ in such a way that the inverse of the estimated spectral density admits the sparse plus low-rank decomposition in (4.4). An estimate of the covariance lags is given by

$$\hat{R}_j = \frac{1}{N} \sum_{t=0}^{N-j} \mathbf{x}^m(t+j)\mathbf{x}^m(t)^\top. \tag{4.5}$$

Then, an estimate $\hat{\Phi}_m$ of Φ_m is obtained by solving the *maximum entropy covariance extension problem*, Burg (1975):

$$\begin{aligned} \hat{\Phi}_m = \underset{\Phi_m \in \mathscr{S}_m^+}{\arg\max} \quad & \int \log|\Phi_m| \\ \text{subject to} \quad & \int e^{ij\theta}\Phi_m = \hat{R}_j \quad j = 0, \dots, n. \end{aligned}$$

The corresponding solution $\hat{\Phi}_m$ is a spectral density of an autoregressive process of order n, i.e. $\hat{\Phi}_m^{-1} \in \mathscr{Q}_{m,n}$. On the other hand, the inverse of $\hat{\Phi}_m$ may not admit the sparse plus low-rank decomposition in (4.4). Thus, drawing inspiration from the robust estimation paradigms proposed in Levy and Nikoukhah (2004), we assume that the actual spectrum Φ_m belongs to a ball centered in the "nominal" spectrum $\hat{\Phi}_m$:

$$\mathscr{B} := \{\Phi \in \mathscr{S}_m^+ : \Phi_m^{-1} \in \mathscr{Q}_{m,n},\ \mathbb{S}(\hat{\Phi}_m||\Phi_m) \leq \delta\} \tag{4.6}$$

where $\delta > 0$ is the prescribed tolerance and $\mathbb{S}(\hat{\Phi}_m||\Phi_m)$ is the Itakura-Saito divergence between $\hat{\Phi}_m$ and Φ_m, Ferrante et al. (2012):

$$\mathbb{S}(\hat{\Phi}_m||\Phi_m) := \int \log|\Phi_m\hat{\Phi}_m^{-1}| + \mathrm{tr}[\hat{\Phi}_m\Phi_m^{-1} - I_m]. \tag{4.7}$$

Then, the proposed problem is:

$$\begin{aligned} \underset{\Phi_m^{-1}, \Sigma, \Lambda \in \mathscr{Q}_{m,n}}{\arg\min} \quad & \phi_*(\Lambda) + \gamma\phi_\infty(\Sigma) \\ \text{subject to} \quad & \Sigma - \Lambda = \Phi_m^{-1}, \\ & \Phi_m \succ 0,\ \ \Lambda \succeq 0, \\ & \mathbb{S}(\hat{\Phi}_m||\Phi_m) \leq \delta, \end{aligned} \tag{4.8}$$

where the objective function is a combination of two penalty functions, ϕ_∞ and ϕ_*, inducing sparsity on Σ and low-rankness on Λ, respectively (see Section 4.5 for details).

The regularization parameter γ balances the effects of the two penalties. It is clear that if δ (which represents a proxy for the uncertainty on the the estimate $\hat{\Phi}_m$) is known, then the solution of problem (4.8) depends only on the regularization parameter γ.

Remark 4.3.1. While in the majority of the applications in probability, statistics and information theory is usual to consider $\mathbb{S}(\Phi_m||\hat{\Phi}_m)$ instead of $\mathbb{S}(\hat{\Phi}_m||\Phi_m)$ the choice of considering $\mathbb{S}(\hat{\Phi}_m||\Phi_m)$ in (4.8) is motivated by the fact that $\mathbb{S}(\hat{\Phi}_m||\Phi_m)$ is convex in Φ_m^{-1}.

4.3.1 Connection with the literature

Before entering into the analysis of problem (4.8), we discuss some connections with related approaches proposed in the literature. In fact, a very natural alternative to problem (4.8) is the regularized negative log-likelihood problem proposed in Zorzi and Sepulchre (2016):

$$\begin{aligned} \underset{\Sigma,\Lambda\in\mathscr{Q}_{m,n}}{\arg\min} \quad & \ell(\mathbf{X}^{m,N};\Sigma,\Lambda)+\lambda\left(\phi_*(\Lambda)+\gamma\phi_\infty(\Sigma)\right) \\ \text{subject to} \quad & \Sigma-\Lambda\succ 0,\ \Lambda\succeq 0. \end{aligned} \tag{4.9}$$

where $\lambda>0$ serves as a regularization parameter while

$$\ell(\mathbf{X}^{m,N};\Sigma,\Lambda) := -\log p(\mathbf{x}^m(n+1)\dots\mathbf{x}^m(N)|\mathbf{x}^m(n)\dots\mathbf{x}^m(1);\Sigma,\Lambda) \tag{4.10}$$

is the negative conditional log-likelihood under the model whose spectrum admits the sparse plus low-rank decomposition in (4.4). In what follows we assume that N is sufficiently large so that the following approximation holds, Songsiri et al. (2010):

$$\ell(\mathbf{X}^{m,N};\Sigma,\Lambda)=\frac{N-n}{2}\int -\log|\Sigma-\Lambda|+\langle\Sigma-\Lambda,\hat{\Phi}_m\rangle. \tag{4.11}$$

The optimal values of the regularization parameters λ and γ can be estimated by using a criterion with complexity terms or by using a cross-validation technique. However, both the procedures require to define a 2-dimensional regularization grid. On the contrary, with our approach only a 1-dimensional regularization grid is needed. As a consequence, the number of candidate models that are required to be computed is drastically reduced. We observe that in view of (4.11) and (4.7), constraints $\Phi_m^{-1}=\Sigma-\Lambda$ and $\mathbb{S}(\hat{\Phi}_m||\Phi_m)\leq\delta$ can be rewritten as

$$\ell(\mathbf{X}^{m,N};\Sigma,\Lambda)\leq l_{MAX} \tag{4.12}$$

where $l_{MAX} = (N-n)(\delta + m + \log|\hat{\Phi}_m|)/2$. Accordingly, our problem can also be written as:

$$\begin{aligned} \underset{\Sigma,\Lambda\in\mathcal{Q}_{m,n}}{\arg\min} \quad & \phi_*(\Lambda) + \gamma\phi_\infty(\Sigma) \\ \text{subject to} \quad & \Sigma - \Lambda \succ 0, \;\; \Lambda \succeq 0, \\ & \ell(\mathbf{X}^{m,N}; \Sigma, \Lambda) \leq l_{MAX}. \end{aligned}$$

It is interesting to note that the problem above provides an estimator whose "principle" is in the same spirit of the SPARSEVA (SPARSe Estimation based on a VAlidation criterion) estimator proposed in Ha, Welsh, Rojas, and Wahlberg (2018):

$$\min_{\theta} \phi(\theta) \quad \text{subject to } \ell(\mathrm{y};\theta) \leq l_{MAX}$$

where $\ell(\mathrm{y};\theta)$ represents the negative log-likelihood of the linear regression model $y = A\theta + e$, where e is a random noise, θ is the unknown parameter and ϕ is the regularizer.
Notice also that (4.8) is a relaxed version of the problem:

$$\begin{aligned} \underset{\Sigma,\Lambda\in\mathcal{Q}_{m,n}}{\arg\min} \quad & \phi_*(\Lambda) + \gamma\phi_\infty(\Sigma) \\ \text{subject to} \quad & \Sigma - \Lambda = \hat{\Phi}_m^{-1}, \\ & \Lambda \succeq 0 \end{aligned} \tag{4.13}$$

which can be understood as the dynamic extension of the identification paradigm proposed in Chandrasekaran et al. (2011). The error affecting the estimate $\hat{\Phi}_m$ may destroy the underlying sparse plus low-rank decomposition in (4.13) while the relaxation in problem (4.8) prevents this issue.
Finally, if we constrain Σ to be diagonal in (4.8), then the resulting model can be understood as a dynamic factor model with idiosyncratic noise, see Scherrer and Deistler (1998); Deistler and Zinner (2007); Zorzi and Sepulchre (2015).

4.4 The choice of δ

The choice of the tolerance parameter δ should reflect the accuracy of the estimation $\hat{\Phi}_m$ of Φ_m. This can be accomplished by choosing a desired probability $\alpha \in (0,1)$ and considering a ball of radius δ_α (in the Itakura-Saito topology) centered in $\hat{\Phi}_m$ and containing the true spectrum Φ_m with probability α. To estimate δ_α we proceed in two steps.
First, we consider the periodogram of Φ_m and we rely on a scale invariance property of the Itakura-Saito divergence. To introduce this property we define $\hat{\hat{R}}_j$ as the *estimator* corresponding to $\hat{R}_j$ i.e. the *random matrix* defined analogously to (4.5), but taking

the Gaussian random variables in place of the corresponding realization $\mathbf{x}^m(t)$. The periodogram, understood as an estimator, is defined as

$$\hat{\hat{\Phi}}^{\mathrm{p}}_m := \sum_{j=-N}^{N} e^{-ij\theta}\hat{\hat{R}}_j.$$

The corresponding estimate is denoted by $\hat{\Phi}^{\mathrm{p}}_m$. This convention is used also for other estimates $\hat{\Phi}$, so that $\hat{\hat{\Phi}}$ denotes the corresponding estimator.

Lemma 4.4.1. *Let $\mathbf{x}^m = \{\mathbf{x}^m(t) : t \in \mathbb{Z}\}$ be a zero mean, stationary, full rank, Gaussian process with spectral density Φ_m. Let $\hat{\hat{\Phi}}^p_m$ be the periodogram based on a sample of $\mathbf{x}^m$ of length N. Then, the Itakura-Saito divergence between $\hat{\hat{\Phi}}^p_m$ and Φ_m is a random variable whose distribution depends only on the numerosity N of the sample and on the dimension m of the process.*

Proof. Let $W(e^{i\theta})$ be the minimum phase spectral factor of Φ_m and define the process $\tilde{\mathbf{x}}^m = \{\tilde{\mathbf{x}}^m(t) : t \in \mathbb{Z}\}$ as $\tilde{\mathbf{x}}^m(t) := W(e^{i\theta})^{-1}\mathbf{x}(t)$. Clearly, $\tilde{\mathbf{x}}^m(t)$ is the normalized white Gaussian noise process. Then, we have

$$\begin{aligned}
\hat{\hat{\Phi}}^{\mathrm{p}}_m(e^{i\theta}) &= \sum_{j=-N}^{N} e^{-ij\theta}\hat{\hat{R}}_j \\
&= \sum_{j=-N}^{N} e^{-ij\theta}\left(\frac{1}{N}\sum_{t=0}^{N-j}\mathbf{x}^m(t+j)\mathbf{x}^m(t)^\top\right) \\
&= \sum_{j=-N}^{N} e^{-ij\theta}\left(\frac{1}{N}\sum_{t=0}^{N-j}W(e^{i\theta})\tilde{\mathbf{x}}^m(t+j)\tilde{\mathbf{x}}^m(t)^\top W^*(e^{i\theta})\right) \\
&= W(e^{i\theta})\left(\sum_{j=-N}^{N} e^{-ij\theta}\frac{1}{N}\sum_{t=0}^{N-j}\tilde{\mathbf{x}}^m(t+j)\tilde{\mathbf{x}}^m(t)^\top\right)W^*(e^{i\theta}) \\
&= W(e^{i\theta})\hat{\hat{\Omega}}^{\mathrm{p}}_N(e^{i\theta})W^*(e^{i\theta})
\end{aligned}$$

where $\hat{\hat{\Omega}}^{\mathrm{p}}_N(e^{i\theta}) := \sum_{j=-N}^{N} e^{-ij\theta}\frac{1}{N}\sum_{t=0}^{N-j}\tilde{\mathbf{x}}^m(t+j)\tilde{\mathbf{x}}^m(t)^\top$ is the periodogram (understood as estimator) based on a sample of the normalized white Gaussian noise of length N. Hence, the Itakura-Saito divergence between $\hat{\hat{\Phi}}^{\mathrm{p}}_m$ and Φ_m is

$$\begin{aligned}
\mathbb{S}(\hat{\hat{\Phi}}^{\mathrm{p}}_m||\Phi_m) &= \int \mathrm{tr}\left\{-\log(\Phi_m^{-1}\hat{\hat{\Phi}}^{\mathrm{p}}_m) + \hat{\hat{\Phi}}^{\mathrm{p}}_m\Phi_m^{-1} - I_m\right\} \\
&= \int -\log|\hat{\hat{\Omega}}^{\mathrm{p}}_N(e^{i\theta})| + \mathrm{tr}(\hat{\hat{\Omega}}^{\mathrm{p}}_N(e^{i\theta})) - m,
\end{aligned}$$

where we exploited the fat that $\Phi_m = WW^*$ so that

$$\mathrm{tr}\log(\Phi_m^{-1}\hat{\hat{\Phi}}_m^{\mathrm{p}}) = \log|W^{-*}W^{-1}W\hat{\hat{\Omega}}_N^{\mathrm{p}}W^*| = \log|\hat{\hat{\Omega}}_N^{\mathrm{p}}|$$

and

$$\mathrm{tr}(\hat{\hat{\Phi}}_m^{\mathrm{p}}\Phi_m^{-1}) = \mathrm{tr}(W\hat{\hat{\Omega}}_N^{\mathrm{p}}W^*W^{-*}W^{-1}) = \mathrm{tr}(\hat{\hat{\Omega}}_N^{\mathrm{p}}).$$

■

In view of the above result, we can easily generate a realization of the random variable $\mathbb{S}(\hat{\hat{\Phi}}_m^{\mathrm{p}}||\Phi_m)$ from a realization of the normalized white Gaussian noise process. Accordingly, we can compute numerically $\delta_\alpha^{\mathrm{p}}$ such that $\Pr(\mathbb{S}(\hat{\hat{\Phi}}_m^{\mathrm{p}}||\Phi_m) \leq \delta_\alpha^{\mathrm{p}}) = \alpha$ by a standard Monte Carlo procedure. It is worth noting that Lemma 4.4.1 holds without making any particular assumption of the structure of Φ_m. Under the assumption that the spectral density Φ_m generating the data is autoregressive of order n, the following relation enables us to consider the Burg estimator $\hat{\hat{\Phi}}_m$ in place of the periodogram $\hat{\hat{\Phi}}_m^{\mathrm{P}}$.

Lemma 4.4.2. *If the spectral density Φ_m generating the data is autoregressive of order n, then*

$$Pr(\mathbb{S}(\hat{\hat{\Phi}}_m||\Phi_m) \leq \delta_\alpha) = Pr(\mathbb{S}(\hat{\hat{\Phi}}_m^p||\Phi_m) \leq \delta_\alpha^p)$$

where $\delta_\alpha := \delta_\alpha^p + \int \log|\hat{\Phi}_m^p| - \log|\hat{\Phi}_m|$.

Proof. It suffices to note that the two events are in fact equivalent. Indeed by simple computation, we see that

$$\int -\log|\Phi_m^{-1}| - \log|\hat{\Phi}_m^{\mathrm{p}}| + \mathrm{tr}(\hat{\Phi}_m^{\mathrm{p}}\Phi_m^{-1}) - m \leq \delta_\alpha^{\mathrm{p}}$$

holds if and only if

$$\int -\log|\Phi_m^{-1}| - \log|\hat{\Phi}_m| + \mathrm{tr}(\hat{\Phi}_m\Phi_m^{-1}) - m \leq \delta_\alpha^{\mathrm{p}} + \int \log|\hat{\Phi}_m^{\mathrm{p}}| - \log|\hat{\Phi}_m|$$

where in the last inequality we have used the fact that $\mathrm{tr}(\hat{\Phi}_m^{\mathrm{p}}\Phi_m^{-1})$ depends only on the first n lags of $\hat{\Phi}_m^{\mathrm{p}}$. ■

Note that $\int \log|\hat{\Phi}_m^{\mathrm{p}}| - \log|\hat{\Phi}_m| \leq 0$ because, once fixed the first n covariance lags R_j, $\hat{\Phi}_m$ is the maximum entropy estimate. Thus $\delta_\alpha \leq \delta_\alpha^{\mathrm{p}}$. More generally, the fact that $\hat{\Phi}_m$ is the maximum entropy estimate implies that for any fixed δ

$$Pr(\mathbb{S}(\hat{\hat{\Phi}}_m||\Phi_m) \leq \delta) \geq Pr(\mathbb{S}(\hat{\hat{\Phi}}||\Phi_m) \leq \delta)$$

where $\hat{\hat{\Phi}}$ is any autoregressive spectral density estimator whose first n covariance lags are equal to $\hat{\hat{R}}_j$. This provides a strong motivation for the choice of $\hat{\Phi}_m$ as the "center" of the

ball where the structured estimate of the spectral density is sought. In fact, this choice maximizes the likelihood that the underlying spectral density Φ_m belongs to such a ball. Finally, notice that all the previous results for the computation of δ_α hold also if we consider a smoothed version of the periodogram (i.e. using a windowing method), Stoica and Moses (1997). In this case, Ω_N^{p} is the smoothed periodogram of the normalized white noise.

It is worth noting that if the chosen α is too large with respect to the data length N, the resulting δ_α may be too generous yielding to a diagonal Φ_m obeying $\mathbb{S}(\hat{\Phi}_m||\Phi_m) \leq \delta_\alpha$. In this case problem (4.8) admits the trivial solution $\Lambda = 0$ and $\Sigma = \Phi_m^{-1}$ diagonal. To rule out this trivial case, δ in (4.8) must be strictly smaller than the upper bound

$$\delta_{\max} := \min_{\Sigma \in \mathscr{S}_m^+, \mathrm{ofd}(\Sigma)=0} \mathbb{S}(\hat{\Phi}_m||\Sigma^{-1}).$$

This problem can be easily solved as follows. Since Σ must be diagonal, denoting by σ_i and by $\hat{\phi}_i$ the i-th elements in the diagonal of Σ and of $\hat{\Phi}_m$, respectively, we have

$$\delta_{\max} = \left[\sum_{i=1}^{m} \min_{\sigma_i \in \mathscr{S}_1^+} \mathbb{S}(\hat{\phi}_i||\sigma_i^{-1})\right] + \int \log|\operatorname{diag}^2(\hat{\Phi}_m)\hat{\Phi}_m^{-1}|.$$

Therefore, the solution corresponds to $\sigma_i^{opt}(e^{i\theta}) = (\hat{\phi}_i(e^{i\theta}))^{-1}$, $i = 1,\ldots,m$, and hence $\mathbb{S}(\hat{\phi}_i||(\sigma_i^{opt})^{-1}) = 0$. Accordingly,

$$\delta_{max} = \int \log|\operatorname{diag}^2(\hat{\Phi}_m)\hat{\Phi}_m^{-1}|. \tag{4.14}$$

A more generous upper bound can be derived by assuming that Σ^{-1} is the spectrum of an autoregressive process of order n. However, numerical experiments showed that $\delta_{max} \gg \delta_\alpha$ even in the case that N is relatively small.

4.5 Matrix formulation

To study problem (4.8) it is convenient to introduce the following matrix parameterization for Σ, Λ and $\Sigma - \Lambda$:

$$\begin{aligned} \Sigma - \Lambda = \Delta X \Delta^* &\in \mathscr{Q}_{m,n} \\ \Sigma = \Delta S \Delta^* &\in \mathscr{Q}_{m,n} \\ \Lambda = \Delta(S - X)\Delta^* &\in \mathscr{Q}_{m,n} \end{aligned}$$

where X and S are matrices in $\mathbf{Q}_{m(n+1)}$. The objective is to provide a more convenient formulation of problem (4.8) in terms of X and S. To this end, we have to take into account the following points.

1) Positivity Constraints $\Sigma - \Lambda \succ 0$ and $\Lambda \succeq 0$: It can been shown (see, for example,

(Zorzi and Sepulchre, 2016, Appendix A)) that, for any $\Psi \in \mathscr{Q}_{m,n}$, $\Psi \succeq 0$ if and only if there exists a matrix $P \in \mathbf{Q}_{m(n+1)}$ such that $\Delta P \Delta^*$ and $P \succeq 0$. Therefore, we replace the conditions $\Lambda \succeq 0$ with $S - X \succeq 0$ and the condition $\Sigma - \Lambda \succ 0$ with $X \succeq 0$. Note that the latter only guarantees $\Sigma - \Lambda$ to be positive semi-definite, however we will show that this is sufficient to guarantee that $\Sigma - \Lambda \succ 0$ at the optimum.

2) The Sparsity regularizer: It is not difficult to see that the coefficients of the pseudo-polynomial matrix Σ are the matrices $Y_j := [D(S)]_j$ with $j = 0, \dots, n$. Hence in this parametrization, the sparsity regularizer must induce the same sparsity pattern on the matrices Y_j, $j = 0, \dots, n$. To this aim we consider the following regularizer, Songsiri and Vandenberghe (2010):

$$\phi_\infty(\Sigma) = h_\infty(Y) = \sum_{k>h} \max\{|[Y_0]_{(h,k)}|, \max_{j=1,\dots,n} \{|[Y_j]_{(h,k)}|, |[Y_j]_{(k,h)}|\}\}.$$

3) The Low Rank Regularizer: we consider the following low-rank regularizer, Zorzi and Sepulchre (2016):

$$\phi_*(\Lambda) := \operatorname{tr} \int \Delta(S-X)\Delta^* = \operatorname{tr}\left((S-X)\int \Delta^*\Delta\right) = tr(S-X)$$

where in the last equality we exploited the fact that $\int e^{ij\theta} = 1$ if $j = 0$, and $\int e^{ij\theta} = 0$ otherwise.

4) The Divergence Constraint: A convenient matrix parametrization of the Itakura-Saito divergence $\mathbb{S}(\hat{\Phi}_m || \Phi_m)$ can be obtained by making use of the following facts. First, since $\Sigma - \Lambda = \Delta X \Delta^*$ with $X \succeq 0$, there exists $A \in \mathbb{R}^{m \times m(n+1)}$ such that $X = A^\top A$. Then by using the Jensen-Kolmogorov formula, we obtain

$$\int \log|\Sigma - \Lambda| = \int \log|\Delta A^\top A \Delta^*| = \log|A_0^\top A_0| = \log|X_{00}|,$$

which holds provided that $X_{00} \succ 0$. Second, observe that

$$\int \operatorname{tr}(\hat{\Phi}_m(\Sigma - \Lambda)) = \int \operatorname{tr}(\hat{\Phi}_m(\Delta X \Delta^*)) = \operatorname{tr}\left(X \int \Delta^* \hat{\Phi}_m \Delta\right) = \langle X, T(\hat{R})\rangle,$$

where $\hat{R} := [\hat{R}_0 \dots \hat{R}_n]$ and we have used the fact that, by construction, $\int \Delta^* \hat{\Phi}_m \Delta = T(\hat{R})$.

Summing up, we get the following matrix re-parametrization of problem (4.8) where the

dummy variable $Y \in \mathbf{M}_{m,n}$ has been introduced.

$$\begin{aligned} \underset{S,X\in\mathbf{Q}_{m(n+1)},\, Y\in\mathbf{M}_{m,n}}{\arg\min} \quad & \operatorname{tr}(S-X)+\gamma h_\infty(Y) \\ \text{subject to} \quad & X_{00}\succ 0,\, X\succeq 0,\, S-X\succeq 0, \\ & Y=D(S), \\ & -\log|X_{00}|-\int \log|\hat{\Phi}_m|+\langle X,T(\hat{R})\rangle - m \le \delta. \end{aligned} \tag{4.15}$$

We remark once again that to prove the equivalence between (4.8) and (4.15) we still need to show that $\Sigma-\Lambda\succ 0$ at the optimum: this fact will be established after the variational analysis.

4.6 The dual problem

To derive the dual of problem (4.15) we introduce the Lagrangian

$$\begin{aligned} \mathscr{L}(X,S,Y,\lambda,U,V,Z) =& \operatorname{tr}(S-X)+\gamma h_\infty(Y)-\langle V,X\rangle-\langle U,S-X\rangle+\langle Z,D(S)-Y\rangle \\ &+\lambda\Big(-\log|X_{00}|-\int\log|\hat{\Phi}_m|+\langle X,T(\hat{R})\rangle-m-\delta\Big) \\ =&\langle S,I-U+T(Z)\rangle+\gamma h_\infty(Y)-\langle Z,Y\rangle+\langle X,U-V-I+\lambda T(\hat{R})\rangle \\ &-\lambda\Big(\log|X_{00}|+\int\log|\hat{\Phi}_m|+m+\delta\Big), \end{aligned} \tag{4.16}$$

where $V,U\in\mathbf{Q}_{m(n+1)}$, $V,U\succeq 0$, are the multipliers associated with the constraints on the positive semi-definiteness of X and $S-X$, respectively. $Z\in\mathbf{M}_{m,n}$ is the multiplier associated with the constraint $Y=D(S)$ and $\lambda\in\mathbb{R},\lambda\ge 0$, is the multiplier associated with the constraint on the Itakura-Saito divergence. Note that we have not included the constraint $X_{00}\succ 0$ because, as we will show later on, this condition is automatically met by the solution of the dual problem. The dual objective function is the infimum of $\mathscr{L}$ over Y,S and X.

1.) Partial minimization over Y: $\mathscr{L}$ depends on Y only through $\gamma h_\infty(Y)-\langle Z,Y\rangle$ which has been show in Songsiri and Vandenberghe (2010) to be bounded from below only if

$$\operatorname{diag}(Z_j)=0, \quad j=0,\dots,n \quad \text{and} \quad \sum_{j=0}^{n}|(Z_j)_{kh}|+|(Z_j)_{hk}|\le\gamma, \quad k\neq h \tag{4.17}$$

and in this case the infimum is attained at zero. Therefore, the partial minimization of

the Lagrangian with respect to Y is

$$\inf_{Y} \mathscr{L} = \begin{cases} \langle S, I - U + T(Z) \rangle + \langle X, U - V - I + \lambda T(\hat{R}) \rangle & \\ -\lambda \left(\log |X_{00}| + \int \log |\hat{\Phi}_m| + m + \delta \right) & \text{if (4.17) holds} \\ -\infty & \text{otherwise.} \end{cases}$$

2.) Partial minimization over S: $\mathscr{L}$ depends on S only through $\langle S, I - U + T(Z) \rangle$ which is bounded from below only if

$$I - U + T(Z) = 0. \tag{4.18}$$

Thus, we get that

$$\inf_{Y,S} \mathscr{L} = \begin{cases} \langle X, U - V - I + \lambda T(\hat{R}) \rangle & \\ -\lambda \left(\log |X_{00}| + \int \log |\hat{\Phi}_m| + m + \delta \right) & \text{if (4.17), (4.18) hold} \\ -\infty & \text{otherwise.} \end{cases}$$

3.) Partial minimization over X: The terms in X_{00} are bounded from below only if

$$\left[U - V - I + \lambda T(\hat{R}) \right]_{00} \succ 0 \tag{4.19}$$

and are minimized if $\lambda > 0$ and

$$X_{00} = \left(\left[\lambda^{-1}(U - V - I) + T(\hat{R}) \right]_{00} \right)^{-1}.$$

The Lagrangian is linear in the remaining variables X_{lh}, for $(l,h) \neq (0,0)$, and hence bounded from below only if

$$\left[U - V - I + \lambda T(\hat{R}) \right]_{lh} = 0 \qquad \forall (l,h) \neq (0,0). \tag{4.20}$$

Therefore, the minimization of the Lagrangian over Y, S and X is finite if and only if (4.17), (4.18), (4.19), and (4.20) hold in which case

$$\min_{Y,S,X} \mathscr{L} = -\lambda \left(-\log \left| \left[\lambda^{-1}(U - V - I) + T(\hat{R}) \right]_{00} \right| + \int \log \left| \hat{\Phi}_m \right| + \delta \right).$$

Otherwise the Lagrangian has no minimum and its infimum is $-\infty$.

Let us define the closed, convex and bounded set $\mathscr{Z}$ as:

$$\mathscr{Z} := \{ Z \in \mathbf{M}_{m,n} : \operatorname{diag}^2(Z_j) = 0 \; j = 0, \ldots, n, \; \sum_{j=0}^{n} |(Z_j)_{lh}| + |(Z_j)_{hl}| \leq \gamma, \; l \neq h \}.$$

Then, the Lagrangian dual problem of problem (4.15) is

$$\max_{(\lambda,U,V,Z)\in\mathscr{C}} \lambda\Big(\log\big|\big[\lambda^{-1}(U-V-I)+T(\hat{R})\big]_{00}\big|-\int\log|\hat{\Phi}_m|-\delta\Big) \tag{4.21}$$

where the set $\mathscr{C}$ is given by:

$$\begin{aligned}\mathscr{C}:=\{(\lambda,U,V,Z):\,&U,V\in\mathbf{Q}_{m(n+1)},\,U,V\succeq 0,\,Z\in\mathcal{Z},\,\lambda\in\mathbb{R},\,\lambda>0,\,I-U+T(Z)=0,\\ &[U-V-I+\lambda T(\hat{R}))]_{00}\succ 0,\,[U-V-I+\lambda T(\hat{R})]_{lh}=0,\,\forall(l,h)\neq(0,0)\}.\end{aligned}$$

Note that the constraint $\mathrm{diag}^2(Z_j)=0,\ j=0,\dots,n$ implies that $\mathrm{diag}^2(T(Z))=0$ and, in view of (4.18), this further implies that

$$\mathrm{diag}^2(U)=I_{m(n+1)}.$$

Moreover, by constraint (4.18) it also follows that $\mathrm{ofd}(U)=\mathrm{ofd}(T(Z))$ and, using again that $\mathrm{diag}^2(T(Z))=0$, this yields

$$\mathrm{ofd}(U)=T(Z).$$

This implies $\mathrm{diag}^2(U_{lh})=0$, for all $l\neq h$. Hence, we can eliminate the redundant variable U; we also observe that $\big[\lambda^{-1}(T(Z)-V)+T(\hat{R})\big]_{00}=\lambda^{-1}(Z_0-V_{00})+\hat{R}_0$, so that we redefine the set $\mathscr{C}$ accordingly:

$$\begin{aligned}\mathscr{C}:=\{(\lambda,V,Z):\,&V\in\mathbf{Q}_{m(n+1)},\,V\succeq 0,\,Z\in\mathcal{Z},\,I+T(Z)\succeq 0,\,\lambda\in\mathbb{R},\,\lambda>0,\\ &[Z_0-V_{00}+\lambda\hat{R}_0]\succ 0,\,[T(Z)-V+\lambda(T(\hat{R}))]_{lh}=0,\,\forall(l,h)\neq(0,0)\}.\end{aligned}$$

Then, by changing the sign we can rewrite (4.21) as a minimization problem:

$$\min_{(\lambda,V,Z)\in\mathscr{C}} J(\lambda,V,Z) \tag{4.22}$$

where $J(\lambda,V,Z):=\lambda\big(-\log|\lambda^{-1}(Z_0-V_{00})+\hat{R}_0|+\int\log|\hat{\Phi}_m|+\delta\big)$.

4.6.1 Existence of solutions

In this subsection we show that the dual problem (4.22) admits solution. This is done by restricting the set $\mathscr{C}$, which is neither closed nor bounded, to a smaller compact set $\mathscr{C}_C$ over which the minimization problem is equivalent to the one in (4.22). Then, since the objective function is continuous over $\mathscr{C}$, and hence over the restricted compact set $\mathscr{C}_C$, by Weierstrass's theorem we will conclude that problem (4.22) does admit a minimum. To this aim, we need some preliminary results.

Lemma 4.6.1. *Let $\Psi(z)$ be the spectral density of a second-order zero-mean, purely non-deterministic process* $\mathbf{y}$ *and let $R_i := \mathbb{E}\{\mathbf{y}(t+i)\mathbf{y}^\top(t)\}$ be the i-th covariance lag of* $\mathbf{y}$*. Finally, let $\mathcal{T}_n = T([R_0 \dots R_{n-1}])$ and $\mathcal{K}_n := [R_1 \dots R_n]$. Then,*

$$\log|R_0 - \mathcal{K}_n\mathcal{T}_n^{-1}\mathcal{K}_n^\top| \geq \int \log|\Psi|.$$

Moreover, if $\mathbf{y}$ *is autoregressive of order n then the previous formula holds with equality.*

Proof. First, recall the classical result by Wiener and Masani (see, (Lindquist and Picci, 2015, Theorem 4.7.5)) stating that

$$\int \log|\Psi| = \log|Q|,$$

where $Q = \mathbb{E}\{\mathbf{e}(t)\mathbf{e}^\top(t)\}$ is the covariance matrix of the optimal prediction error, i.e., $\mathbf{e}(t) := \mathbf{y}(t) - \hat{\mathbf{y}}(t)$ where $\hat{\mathbf{y}}(t)$ is the minimum variance prediction of $\mathbf{y}(t)$ based on the whole past history $\{\mathbf{y}(t-i)\}_{i=1}^{\infty}$ of $\mathbf{y}$. Clearly $Q \leq Q_n$ where $Q_n = \mathbb{E}\{\mathbf{e}_n(t)\mathbf{e}_n^\top(t)\}$ is the covariance matrix of the optimal n-steps prediction error, i.e., $\mathbf{e}_n(t) := \mathbf{y}(t) - \hat{\mathbf{y}}_n(t)$ where $\hat{\mathbf{y}}_n(t)$ is the minimum variance prediction of $\mathbf{y}(t)$ based only on $\{\mathbf{y}(t-i)\}_{i=1}^{n}$. Also, if $\mathbf{y}$ is autoregressive of order n then the optimal predictor only uses the last n values of $\mathbf{y}$ so that $Q = Q_n$. It remains to prove that $|R_0 - \mathcal{K}_n T_n^{-1}\mathcal{K}_n^\top| = |Q_n|$. To this aim, we resort to the formula (see, (Lindquist and Picci, 2015, page 140))

$$|Q_n| = |\mathcal{T}_n|/|\mathcal{T}_{n-1}|. \tag{4.23}$$

By taking into account the Woodbury determinant formula, we immediately see that $|\mathcal{T}_n| = |\mathcal{T}_{n-1}||R_0 - \mathcal{K}_n\mathcal{T}_n^{-1}\mathcal{K}_n^\top|$ which, plugged in (4.23), yields the result. ∎

Lemma 4.6.2. *Let*

$$M = \begin{bmatrix} M_{00} & M_{01} & \dots & M_{0n} \\ M_{01}^\top & M_{11} & \dots & M_{1n} \\ \vdots & \dots & \dots & \dots \\ M_{0n}^\top & \dots & \dots & M_{nn} \end{bmatrix} \in \mathbf{Q}_{m(n+1)}$$

be a symmetric and positive (semi-) definite matrix partitioned in blocks M_{jl} of dimension

$m \times m$. *Let* $M_{jl,d} := \operatorname{diag}^2(M_{jl})$ *for all* $j,l = 0,\dots,n$ *and*

$$M_d := \begin{bmatrix} M_{00,d} & M_{01,d} & \dots & M_{0n,d} \\ M_{02,d}^\top & M_{11,d} & \dots & M_{1n,d} \\ \vdots & \dots & \dots & \dots \\ M_{0n,d}^\top & \dots & \dots & M_{nn,d} \end{bmatrix}.$$

Then M_d *is positive (semi-) definite.*

Proof. Let $U \in \mathbb{R}^{m(n+1)\times m(n+1)}$ be a permutation (invertible) matrix whose entries are all zero except for those in positions (rows and columns) $(l+m(j-1),(l-1)(n+1)+j)$ for $l = 1,\dots,m$ and $j = 1,\dots,n+1$. Then by direct inspection we can check that $U^\top M_d U$ is a block-diagonal matrix having the block-diagonal elements equal to those of the (in general full) matrix $U^\top MU$ which is clearly positive (semi-) definite because M is such by assumption. ∎

We are ready to show that we can restrict $\mathscr{C}$ to a subset where $\lambda \geq \varepsilon$, with $\varepsilon > 0$ being a positive constant.

Proposition 4.6.3. *Let* $(\lambda^{(k)},V^{(k)},Z^{(k)})_{k\in\mathbb{N}}$ *be a sequence of elements in* $\mathscr{C}$ *such that*

$$\lim_{k\to\infty} \lambda^{(k)} = 0. \tag{4.24}$$

Then, such a sequence cannot be an infimizing sequence.

Proof. We consider two possible scenarios separately.
1) Let $(\lambda^{(k)},V^{(k)},Z^{(k)})$ be such that, besides (4.24), we have

$$\|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})\| \to +\infty.$$

Since we are dealing with symmetric matrices, this is equivalent to the fact that the maximum of the absolute values of the eigenvalues of $(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})$ diverges:

$$\lim_{k\to\infty} \max_{\alpha^{(k)}\in\sigma\left((\lambda^{(k)})^{-1}([Z^{(k)}]_0-[V^{(k)}]_{00})\right)} |\alpha^{(k)}| = +\infty. \tag{4.25}$$

The next step is to show that this in turn implies that

$$\lim_{k\to\infty} \min_{\alpha^{(k)}\in\sigma\left((\lambda^{(k)})^{-1}([Z^{(k)}]_0-[V^{(k)}]_{00})\right)} \alpha^{(k)} = -\infty. \tag{4.26}$$

Indeed, (4.25) implies that at least one of the two conditions (4.26) and

$$\lim_{k\to\infty} \max_{\alpha^{(k)}\in\sigma\left((\lambda^{(k)})^{-1}([Z^{(k)}]_0-[V^{(k)}]_{00})\right)} \alpha^{(k)} = +\infty \tag{4.27}$$

holds. To show that (4.27) implies (4.26), notice that $V^{(k)} \succeq 0$ and $\lambda^{(k)} > 0$ imply that, $\forall k$, $\max\{\alpha^{(k)};\, \alpha^{(k)} \in \sigma\left((\lambda^{(k)})^{-1}[Z^{(k)}]_0\right)\}$ is non-smaller than the argument of the limit in the left-hand side of (4.27) so that

$$\lim_{k\to\infty} \max_{\alpha^{(k)}\in\sigma\left((\lambda^{(k)})^{-1}[Z^{(k)}]_0\right)} \alpha^{(k)} = +\infty. \tag{4.28}$$

Since $(\lambda^{(k)})^{-1}[Z^{(k)}]_0$ has zero trace for all k, this yields,

$$\lim_{k\to\infty} \min_{\alpha^{(k)}\in\sigma\left((\lambda^{(k)})^{-1}[Z^{(k)}]_0\right)} \alpha^{(k)} = -\infty.$$

By the same argument used before, $V^{(k)} \succeq 0$ and $\lambda^{(k)} > 0$ imply that, $\forall k$, $\min\{\alpha^{(k)};\, \alpha^{(k)} \in \sigma\left((\lambda^{(k)})^{-1}[Z^{(k)}]_0\right)\}$ is non-smaller than the argument of the limit in the left-hand side of (4.26). This fact, together with (4.28), implies (4.26). In turn, (4.26) implies that for k sufficiently large $[Z^{(k)}]_0 - [V^{(k)}]_{00} + \lambda^{(k)}\hat{R}_0 = \lambda^{(k)}[(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0]$ has at least a negative eigenvalue so that the sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})$ is not in $\mathscr{C}$.
2) Let us now consider a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k\in\mathbb{N}}$ for which, besides (4.24), we have $\|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})\| \to c$, with $0 \le c < \infty$. Then, it can be easily seen that $\forall \varepsilon > 0$, $\exists \bar{k}$ such that $J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) > -\varepsilon$, for all $k > \bar{k}$. As a matter of fact, since $\|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})\|$ is bounded there exists a real constant l such that for all k it holds that $(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0 \le lI_m$. Therefore,

$$|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0| \le l^m$$

$$\log|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0| \le m\log l$$
$$-\log|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0| \ge -m\log l.$$

Then, we can define a second real constant l_1 as $l_1 := -m\log l + \int \log|\hat{\Phi}_m| + \delta$ and for all k it holds that $J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) \ge \lambda^{(k)} l_1$. Since l_1 is constant $\lambda^{(k)} l_1 \to 0$ so that, by definition, for all $\varepsilon > 0$, $\exists \bar{k}$ such that

$$J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) \ge \lambda^{(k)} l_1 > -\varepsilon, \qquad \forall k > \bar{k}.$$

Therefore, it is sufficient to find a triple $(\bar{\lambda}, \bar{V}, \bar{Z}) \in \mathscr{C}$ such that the $J(\bar{\lambda}, \bar{V}, \bar{Z})$ is strictly negative to conclude that such a sequence is not an infimizing sequence. To this purpose, let us consider $\bar{\lambda}$ to be sufficiently small but strictly greater than zero. Moreover, let $\bar{Z}_j = -\bar{\lambda}\,\mathrm{ofd}(\hat{R}_j)$ for all $j = 0, \ldots, n$. Finally, we need to define $\bar{V}_{lh}$. To this end, let

$\hat{R}_{j,d} := \mathrm{diag}^2(\hat{R}_j)$, $j = 0,\dots,n$, $\hat{R}_d := [\hat{R}_{0,d} \mid \cdots \mid \hat{R}_{n,d}]$ and $\mathcal{T}_{n+1,d} := T(\hat{R}_d)$: $\mathcal{T}_{n+1,d}$ is defined from $T(\hat{R})$ by the same "block by block diagonalization" procedure defined in Lemma 4.6.2 so that it is positive definite. Now we partition $\mathcal{T}_{n+1,d}$ as

$$\mathcal{T}_{n+1,d} = \begin{bmatrix} \hat{R}_{0,d} & \mathcal{K}_d \\ \mathcal{K}_d^\top & \mathcal{T}_{n,d} \end{bmatrix}$$

which defines the matrices $\mathcal{K}_d$ and $\mathcal{T}_{n,d}$. We now set

$$\bar{V} := \bar{\lambda} \begin{bmatrix} \mathcal{K}_d \mathcal{T}_{n,d}^{-1} \mathcal{K}_d^\top & \mathcal{K}_d \\ \mathcal{K}_d^\top & \mathcal{T}_{n,d} \end{bmatrix}.$$

Notice that, in view of Lemma 4.6.2, $\mathcal{T}_{n,d}$ is positive definite, and hence invertible, so that $\bar{V}$ is positive semi-definite. Now, it is not difficult to check that the triple $(\bar{\lambda},\bar{V},\bar{Z})$ just defined is in $\mathscr{C}$ for $\bar{\lambda}$ sufficiently small. It remains to show that $J(\bar{\lambda},\bar{V},\bar{Z})$ is negative. Indeed, by linearity $\mathrm{diag}^2(\hat{\Phi}_m)$ is the power spectral density of the process whose covariance lags are $\hat{R}_{d,j}$ so that, in view of Lemma 4.6.1, we have

$$\begin{aligned} J(\bar{\lambda},\bar{V},\bar{Z}) =& -\bar{\lambda}\log|\hat{R}_{0,d} - \mathcal{K}_d\mathcal{T}_{n,d}^{-1}\mathcal{K}_d| + \bar{\lambda}\int \log|\hat{\Phi}_m| + \bar{\lambda}\delta \\ \leq& -\bar{\lambda}\int \log|\mathrm{diag}^2(\hat{\Phi}_m)| + \bar{\lambda}\int \log|\hat{\Phi}_m| + \bar{\lambda}\delta \\ =& \bar{\lambda}\left(\delta - \int \log|\mathrm{diag}^2(\hat{\Phi}_m)\hat{\Phi}_m^{-1}|\right) \\ =& \bar{\lambda}\Big(\delta - \delta_{\max}\Big) < 0, \end{aligned}$$

where, in the last equality we have taken into account the expression (4.14) of $\delta_{\max}$ and the last inequality follows from the assumption $\delta_{\max} > \delta$.

This suffices to conclude the proof. Indeed, the only possible remaining case is that for which $\lim_{k\to\infty} \|\lambda^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00})\|$ does not exists. However, in this case it is always possible to consider a subsequence $(\lambda^{(k_j)},V^{(k_j)},Z^{(k_j)})$ for which the limit exists (finite or infinite) and we can therefore reduce to one of the previous cases. ■

As a consequence minimizing the dual objective function over the set $\mathscr{C}$ is equivalent to minimize it over the set

$$\begin{aligned} \mathscr{C}_1 := \{(\lambda,V,Z) : V \in \mathbf{Q}_{m(n+1)}, V \succeq 0, Z \in \mathcal{Z}, I + T(Z) \succeq 0, \lambda \in \mathbb{R}, \lambda \geq \varepsilon, \\ (Z_0 - V_{00} + \lambda\hat{R}_0) \succ 0, [T(Z) - V + \lambda T(\hat{R})]_{lh} = 0, \forall (l,h) \neq (0,0)\} \end{aligned}$$

for a certain $\varepsilon > 0$.

Next we show that we can restrict the search for the optimal solution to a subset of $\mathscr{C}_1$ in which both V and λ cannot diverge.

Proposition 4.6.4. *Let $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k\in\mathbb{N}}$ be a sequence of elements in $\mathscr{C}_1$ such that either*

$$\lim_{k\to\infty} \|V^{(k)}\| = +\infty \tag{4.29}$$

or

$$\lim_{k\to\infty} \lambda^{(k)} = +\infty \tag{4.30}$$

or both. Then, such a sequence cannot be an infimizing sequence.

Proof. We first observe that (4.30) holds if and only if (4.29) holds as well. In fact, we are assuming that the estimated model has non-trivial *dynamic*, i.e. there exists $i \neq 0$ such that $\hat{R}_i \neq 0$. Hence, there exist $(\bar{l}, \bar{h}) \neq (0,0)$ such that $[T(\hat{R})]_{\bar{l}\bar{h}} \neq 0$ so that, since $T(Z^{(k)})$ is bounded, $[T(Z^{(k)}) - V^{(k)} + \lambda^{(k)} T(\hat{R})]_{\bar{l}\bar{h}} = 0$ (which is one of the conditions for the sequence to be in $\mathscr{C}_1$) implies that (4.30) holds if and only if $[V^{(k)}]_{\bar{l}\bar{h}}$ diverges. Since (4.29) holds if and only if $[V^{(k)}]_{lh}$ diverges for some (l,h), it remains to show that if

$$\lim_{k\to\infty} \|[V^{(k)}]_{00}\| = +\infty \tag{4.31}$$

then (4.30) holds. Since $V^{(k)} \succeq 0$ for all k, (4.31) implies that at least one of the eigenvalues of $[V^{(k)}]_{00}$ tends to $+\infty$ as k diverges. Since $\hat{R}_0$ is fixed and $Z^{(k)}$ is bounded, this implies that $([Z^{(k)}]_0 - [V^{(k)}]_{00} + \lambda^{(k)}\hat{R}_0) \succ 0$ can hold for all k only if (4.30) holds. So far we have seen that (4.31) implies (4.29) which is equivalent to (4.30). We now show that (4.29) implies not only (4.31) but the stronger condition

$$\lim_{k\to\infty} \frac{\|[V^{(k)}]_{00}\|}{\lambda^{(k)}} \neq 0. \tag{4.32}$$

In fact, we assume by contradiction that (4.32) does not hold, i.e. that $\frac{\|[V^{(k)}]_{00}\|}{\lambda^{(k)}}$ tends to zero and we show that the corresponding sequence cannot belong to $\mathscr{C}_1$ as the constraint on the positive semi-definiteness of $V^{(k)}$ fails for k sufficiently large. Indeed, a symmetric matrix is positive semi-definite if and only if every principal minor is non-negative. Thus, let us consider the principal minor of order 2 obtained as follows. Select a block $\hat{R}_h$, with $h \neq 0$, and an element in position (p,q), such that $[\hat{R}_h]_{(p,q)} \neq 0$. Note that it is always possible to find such an element because we are assuming that the process has non-trivial dynamic. Then, consider the following 2×2 sub-matrix of $V^{(k)}$:

$$\begin{bmatrix} [[V^{(k)}]_{00}]_{(p,p)} & \lambda^{(k)}[\hat{R}_h]_{(p,q)} + [[Z^{(k)}]_h]_{(p,q)} \\ \lambda^{(k)}[\hat{R}_h]_{(p,q)} + [[Z^{(k)}]_h]_{(p,q)} & \lambda^{(k)}[\hat{R}_0]_{(q,q)} \end{bmatrix} \tag{4.33}$$

where $[\hat{R}_0]_{(q,q)} > 0$ and the off-diagonal terms are obtained by employing the constraint

$[T(Z^{(k)}) - V^{(k)} + \lambda^{(k)} T(\hat{R})]_{lh} = 0$ to express the corresponding entry of $V^{(k)}$ in terms of $Z^{(k)}$ and $\lambda^{(k)}$. Since $Z^{(k)}$ is bounded and we are assuming (by contradiction) that $\frac{\|[V^{(k)}]_{00}\|}{\lambda^{(k)}}$ tends to zero, we immediately see that the determinant of (4.33) is negative for k sufficiently large. Therefore, the constraint on the positive semi-definiteness of $V^{(k)}$ fails. Thus, the proof reduces to ruling out the following two possible cases.

1. Consider the case of a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k\in\mathbb{N}}$ such that, besides (4.30), we also have $\|(\lambda^{(k)})^{-1}[V^{(k)}]_{00}\| \to \infty$ as k tends to ∞. Then, it follows that the largest singular value of $(\lambda^{(k)})^{-1}[V^{(k)}]_{00}$ tends to $+\infty$ as $k \to \infty$ and, since $\lambda^{(k)}$ and $V^{(k)}$ are positive, this implies that the largest eigenvalue of $(\lambda^{(k)})^{-1}[V^{(k)}]_{00}$ tends to $+\infty$ as $k \to \infty$. In turn, this implies that positivity of $([Z^{(k)}]_0 - [V^{(k)}]_{00} + \lambda^{(k)}\hat{R}_0) = \lambda^{(k)}[(\lambda^{(k)})^{-1}[Z^{(k)}]_0 - (\lambda^{(k)})^{-1}[V^{(k)}]_{00} + \hat{R}_0]$ fails for k sufficiently large, which rules out this case.

2. Finally, consider a sequence for which $\|[V^{(k)}]_{00}\| \to \infty$ at the same speed of $\lambda^{(k)}$ and $\|[V^{(k)}]_{lh}\|$. Note that, since $((\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0) \succ 0$, it holds that:

$$\begin{aligned}[V^{(k)}]_{00} &= \lambda^{(k)}(\hat{R}_0 - C^{(k)}) + [Z^{(k)}]_0 \\ &= \lambda^{(k)}[(\hat{R}_0 - C^{(k)}) + (\lambda^{(k)})^{-1}[Z^{(k)}]_0]\end{aligned}$$

for a certain $C^{(k)} \succ 0$. Since $Z^{(k)}$ is bounded element wise, as $k \to \infty$, $(\lambda^{(k)})^{-1}[Z^{(k)}]_0$ tends to zero as $1/\lambda^{(k)}$ or faster. Then, we have:

$$(\lambda^{(k)})^{-1}V^{(k)} = \begin{bmatrix} \hat{R}_0 - C^{(k)} & \mathcal{K} \\ \mathcal{K}^\top & \mathcal{T} \end{bmatrix} + O(\frac{1}{\lambda^{(k)}}),$$

with $\mathcal{K} := [\hat{R}_1 \ \ldots \ \hat{R}_n]$ and $\mathcal{T} := T([\hat{R}_0 \ \ldots \ \hat{R}_{n-1}])$. Since $V^{(k)} \succeq 0$, by using the Schur complement, we get $C^{(k)} \preceq C^{(k)}_{max}$ with

$$C^{(k)}_{max} := \hat{R}_0 - \mathcal{K}\mathcal{T}^{-1}\mathcal{K}^\top + O\left(\frac{1}{\lambda^{(k)}}\right).$$

Therefore, $[V^{(k)}]_{00} \geq \lambda^{(k)}[(\hat{R}_0 - C^{(k)}_{max}) + (\lambda^{(k)})^{-1}[Z^{(k)}]_0]$. Hence, in view of Lemma 4.6.1 which holds with equality:

$$\begin{aligned}J^{(k)} &:= J(\lambda^{(k)}, V^{(k)}, Z^{(k)}) \\ &= \lambda^{(k)}\left(\int -\log|(\lambda^{(k)})^{-1}[[Z^{(k)}]_0 - [V^{(k)}]_{00}] + \hat{R}_0| + \log|\hat{\Phi}_m| + \delta\right) \\ &\geq \lambda^{(k)}\left(\int \log|\hat{R}_0 - \mathcal{K}\mathcal{T}^{-1}\mathcal{K}^\top| + \delta - \log|C^{(k)}_{max}|\right) \\ &= \lambda^{(k)}[\delta + O(\frac{1}{\lambda^{(k)}})] \to +\infty, \qquad \text{as } k \to \infty\end{aligned}$$

Thus $(\lambda^{(k)}, V^{(k)}, Z^{(k)})$ cannot be an infimizing sequence. ■
Then, we can further restrict the set $\mathscr{C}_1$ to:

$$\mathscr{C}_2 := \{(\lambda, V, Z) : V \in \mathbf{Q}_{m(n+1)},\, \alpha I \succeq V \succeq 0,\, Z \in \mathcal{Z},\, I + T(Z) \succeq 0,\, \lambda \in \mathbb{R},\, \gamma \geq \lambda \geq \varepsilon,$$
$$(Z_0 - V_{00} + \lambda \hat{R}_0) \succ 0,\, [T(Z) - V + \lambda T(\hat{R})]_{lh} = 0,\, \forall (l,h) \neq (0,0)\}$$

for certain α and γ, with $\alpha, \gamma < +\infty$.
Finally, consider a sequence $(\lambda^{(k)}, V^{(k)}, Z^{(k)})_{k\in\mathbb{N}} \in \mathscr{C}_2$ such that $((\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0)$ tends to be singular as $k \to \infty$. This implies that $|(\lambda^{(k)})^{-1}([Z^{(k)}]_0 - [V^{(k)}]_{00}) + \hat{R}_0|$ tends to zero and hence $J \to +\infty$. Therefore, such a sequence cannot be an infimizing sequence. The final set $\mathscr{C}_C$ is, therefore:

$$\mathscr{C}_C := \{(\lambda, V, Z) : V \in \mathbf{Q}_{m(n+1)},\, \alpha I \succeq V \succeq 0,\, Z \in \mathcal{Z},\, I + T(Z) \succeq 0,\, \lambda \in \mathbb{R},\, \gamma \geq \lambda \geq \varepsilon,$$
$$(Z_0 - V_{00} + \lambda \hat{R}_0) \succeq \beta I,\, [T(Z) - V + \lambda T(\hat{R})]_{lh} = 0,\, \forall (l,h) \neq (0,0)\},$$

where $\alpha, \beta, \varepsilon$ and γ are positive constants. It is evident that $\mathscr{C}_C$ is compact, hence there exists a solution minimizing the dual objective function.

4.7 Equivalence between the original problem and the matrix formulation

We are now ready to show that the original problem (4.8) and the matrix formulation in (4.15) are equivalent. To this aim we need the following result, see Songsiri et al. (2010).

Lemma 4.7.1. *Let $Z \in \mathbf{M}_{m,n}$ and $W \in \mathbf{Q}_m$. If $W \succ 0$ is such that*

$$T(Z) \succeq \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix}$$

then $T(Z) \succ 0$ and the unique solution to the Yule-Walker equation

$$\begin{cases} T(Z)B^\top = \begin{bmatrix} W \\ 0 \end{bmatrix}, \quad B \in \mathbf{M}_{m,n} \\ B_0 = I \end{cases}$$

is such that $B\Delta^*$ has zeros inside the unit circle (Δ being the shift operator defined in (4.2)).

Let $(\lambda^\circ, V^\circ, Z^\circ)$ be a solution of (4.22) and (S°, X°) be the corresponding solution of (4.15). We observe that the primal problem is strictly feasible (which can be seen by taking, for instance, $X = T(\hat{R})$) and hence the following extremality relation holds, see

e.g. (Ekeland and Temam, 1999, Theorem 5.1):

$$\langle V^\circ, X^\circ \rangle = 0. \tag{4.34}$$

It is not difficult to see that

$$V^\circ = T(Z^\circ) + \lambda^\circ T(\hat{R}) - \begin{bmatrix} W^\circ & 0 \\ 0 & 0 \end{bmatrix}$$

where

$$W^\circ := Z_0^\circ - V_{00}^\circ + \lambda^\circ \hat{R}_0 \succ 0. \tag{4.35}$$

Since $V^\circ \succeq 0$ and in view of Lemma 4.7.1, we have that $T(Z^\circ) + \lambda^\circ T(\hat{R}) \succ 0$. Hence, V° has rank at least equal to mn. Moreover, since $V^\circ, X^\circ \succeq 0$ and in view of (4.34), we have that X° has rank at most equal to m. On the other hand, $\text{rank}(X^\circ) \geq m$ because $X_{00}^\circ = \lambda^\circ (W^\circ)^{-1}$ is positive definite. Therefore, $\text{rank}(X^\circ) = m$ and there exists $A \in \mathbb{R}^{m \times m(n+1)}$ full-row rank such that $X^\circ = A^\top A$ with $X_{00}^\circ = A_0^\top A_0$. By (4.34) and the fact that $V^\circ, X^\circ \succeq 0$, it follows that

$$\left(T(Z^\circ) + \lambda^\circ T(\hat{R}) - \begin{bmatrix} W^\circ & 0 \\ 0 & 0 \end{bmatrix} \right) A^\top = 0. \tag{4.36}$$

Finally, we set $B := A_0^{-1} A \in \mathbf{M}_{m,n}$, and Equation (4.36) reads

$$\left(T(Z^\circ) + \lambda^\circ T(\hat{R}) \right) B^\top = \begin{bmatrix} W^\circ \\ 0 \end{bmatrix}, \quad B_0 = I. \tag{4.37}$$

Since $T(Z^\circ) + \lambda^\circ T(\hat{R}) \succ 0$ and in view of Lemma 4.7.1, the Yule-Walker equation (4.37) admits a unique solution such that $B\Delta^*$ has zeros inside the unit circle. Accordingly, X° is such that

$$\Delta X^\circ \Delta^* = \Delta A^\top A \Delta^* = \Delta B^\top (W^\circ)^{-1} B \Delta^* \succ 0.$$

This result leads to the following proposition.

Proposition 4.7.2. *Let (S°, X°) be a solution of (4.15). Then, $\Delta X^\circ \Delta^* \succ 0$. Accordingly, (4.8) and (4.15) are equivalent.*

4.7.1 Uniqueness of the solution of the dual problem

We now provide conditions for the uniqueness of the solution of the dual problem. To start with, as observed in the previous section, it is possible to get rid of the slack variable $V \in \mathbf{Q}_{m(n+1)}$ by introducing the new variable $W \in \mathbf{Q}_m$ defined, similarly to (4.35), as

$W := Z_0 - V_{00} + \lambda \hat{R}_0$. Since we have the constraint $V = T(Z) + \lambda T(\hat{R}) - \begin{bmatrix} W & 0 \\ 0 & 0 \end{bmatrix} \succeq 0$, the matrix W must satisfy $W \preceq W_{\max}$, where $W_{\max}$ can be easily derived by taking into account that $V \succeq 0$ and computing the the Schur complement of the south-east block of V. On the other hand, the functional J may be written in terms of W as $J = \lambda\big(-\log|\lambda^{-1}W| + \int \log|\hat{\Phi}_m| + \delta\big)$ so that at the optimum W is necessarily equal to $W_{\max}$ because the determinant is the product of the eigenvalues and $\lambda > 0$ (see Songsiri and Vandenberghe (2010) for a similar argument).

To derive a simple expression for $W_{\max}$, we introduce the linear operators $T_{0,0} : \mathbf{M}_{m,n} \to \mathbf{Q}_m$, $T_{0,1:n} : \mathbf{M}_{m,n} \to \mathbf{M}_{m,n-1}$ and $T_{1:n,1:n} : \mathbf{M}_{m,n} \to \mathbf{Q}_{mn}$ that, for a given matrix $H \in \mathbf{M}_{m,n}$, construct a symmetric Toeplitz matrix and extract the blocks in position $(0,0)$, $(0,1:n)$ and $(1:n,1:n)$, respectively. With this notation we have

$$T(Z+\lambda\hat{R}) = \begin{bmatrix} T_{0,0}(Z+\lambda\hat{R}) & T_{0,1:n}(Z+\lambda\hat{R}) \\ T_{0,1:n}^{\top}(Z+\lambda\hat{R}) & T_{1:n,1:n}(Z+\lambda\hat{R}) \end{bmatrix}$$

and the following expression of $W_{\max}$ can be easily derived:

$$W_{\max}(\lambda,Z) = T_{0,0}(Z+\lambda\hat{R}) - T_{0,1:n}(Z+\lambda\hat{R})(T_{1:n,1:n}(Z+\lambda\hat{R}))^{-1}T_{0,1:n}^{\top}(Z+\lambda\hat{R}).$$

Therefore the dual problem (4.22) can be stated in terms of the two variables λ and Z only. To this end we introduce the function $F(\lambda,Z) := \lambda(-\log|\lambda^{-1}W_{\max}| + \int \log|\hat{\Phi}_m| + \delta)$ and we observe that minimizing J over $\mathscr{C}$ is equivalent to minimize F over the set $\mathscr{C}_F := \{(\lambda,Z) : \lambda > 0, Z \in \mathcal{Z}, T(Z) \succeq -I, W_{\max}(\lambda,Z) \succ 0\}$. Hence, from now on we can consider the following problem

$$\min_{(\lambda,Z)\in\mathscr{C}_F} F(\lambda,Z) \tag{4.38}$$

F is convex over $\mathscr{C}_F$, however, there exists at least one direction along which the convexity is not strict. Indeed, for any pair (λ,Z) in the interior of $\mathscr{C}_F$, let $\varepsilon > 0$ be such that $((1+\varepsilon)\lambda,(1+\varepsilon)Z) \in \mathscr{C}_F$. Then, since the operators $T_{0,0}$, $T_{0,1:n}$, $T_{1:n,1:n}$ are linear in the pair (λ,Z), by direct computation we get:

$$F((1+\varepsilon)\lambda,(1+\varepsilon)Z) = (1+\varepsilon)F(\lambda,Z),$$

that is, there exits one direction along which F is linear: this direction is given by $(\delta\lambda,\delta Z) = (\alpha\lambda,\alpha Z)$ with $\alpha \neq 0$. This direction does not affect the uniqueness of the solution as it can be shown by resorting to the same argument illustrated in Subsection 2.4.2.

Indeed, notice first that along the considered direction the objective function can be constant if and only if it is zero. Then, *ad absurdum*, assume that (λ°,Z°) and $(\alpha\lambda^\circ,\alpha Z^\circ)$

are two optimal solution. By convexity of the set it follows that the whole segment connecting (λ°, Z°) and $(\alpha\lambda^\circ, \alpha Z^\circ)$ belongs to the set $\mathscr{C}_F$ and by the convexity of the objective function F it follows that all the points along this segment are optimal solutions so that F is constant along the segment. But as just noticed this can happen along the aforementioned direction if and only if the objective function is zero while at the optimum the objective function must be negative as shown in the previous section. Therefore, as claimed, the non-strict convexity along the direction $(\delta\lambda, \delta Z) = (\alpha\lambda, \alpha Z)$ will not affect the uniqueness of the optimal solution. In particular, any optimal solution lies on the boundary of the set $\mathscr{C}_F$.

In order to discuss strict convexity of the objective function along the other directions, we consider the second variation of F twice in the same direction δZ.

The first variation of F with respect to δZ is easily seen to be:

$$\delta F(\lambda, Z; \delta Z) = -\lambda^2 \operatorname{tr}(W_{\max} \delta W_{\max}),$$

with

$$\begin{aligned}\delta W_{\max} =& \delta W_{\max}(\lambda, Z; \delta Z)\\ =& T_{0,0}(\delta Z) - T_{0,1:n}(\delta Z) T_{1:n,1:n}^{-1} T_{0,1:n}^\top + T_{0,1:n} T_{1:n,1:n}^{-1} T_{1:n,1:n}(\delta Z) T_{1:n,1:n}^{-1} T_{0,1:n}^\top\\ & - T_{0,1:n} T_{1:n,1:n}^{-1} T_{0,1:n}^\top(\delta Z).\end{aligned}$$

and where we have introduced the convention that whenever the argument of the operators $T_{1:n,1:n}$ and $T_{0,1:n}$ is omitted it is intended to be equal to $(Z + \lambda \hat{R})$. Using the same convention, the second variation of F with respect to δZ turns out to be:

$$\begin{aligned}\delta^2 F(\lambda, Z; \delta Z) = \lambda^2 \operatorname{tr}[& W_{\max}^{-1} \delta W_{\max} W_{\max}^{-1} \delta W_{\max} + 2 W_{\max}^{-1} (T_{0,1:n}(\delta Z)\\ & - T_{0,1:n} T_{1:n,1:n}^{-1} T_{1:n,1:n}(\delta Z)) T_{1:n,1:n}^{-1} (T_{0,1:n}^\top(\delta Z) - T_{1:n,1:n}(\delta Z) T_{1:n,1:n}^{-1} T_{0,1:n}^\top)].\end{aligned}$$

It is easy to see that $\delta^2 F$ is always non negative, and it is zero if and only if the following two conditions hold:

$$\begin{cases} \delta W_{\max} = 0_m \\ T_{0,1:n}(\delta Z) = T_{0,1:n} T_{1:n,1:n}^{-1} T_{1:n,1:n}(\delta Z). \end{cases} \tag{4.39}$$

Note that the first condition implies that F might be non-strictly convex only at a stationary point. System (4.39) has $nm^2 + m(m+1)/2$ homogeneous linear equations in a much smaller number $(m^2 - m)/2 + n(m^2 - n)$ of unknowns δZ. Based on this heuristic, we expect that, normally, $\delta Z = 0$ is the only solution of (4.39); clearly, in this case the solution of the dual problem is unique. Such uniqueness condition can be easily checked numerically (as (4.39) is a system of *linear* equation in δZ) for a given optimal solution. This leads us to the following result.

Proposition 4.7.3. *Let* (λ°, Z°) *be an optimal solution of* (4.38)*. Then such solution is unique if and only if* $\delta Z = 0$ *is the only solution of the system of equations* (4.39) *evaluated in* (λ°, Z°)*.*

4.8 Recovery of the solution of the primal problem

First, let B be the solution of the *Yule-Walker* equation (4.37), then X° is obtained as $X^\circ = B^\top (W^\circ)^{-1} B$. Note that, if the dual problem admits a unique solution, in view of Lemma 4.7.1 both the matrices W° and B are unique and therefore the uniqueness of X° follows. Consider next the matrix $S^\circ - X^\circ$. In view of the strict feasibility of the primal problem, and of the properties of the primal and dual, the following extremality relation holds:

$$\langle U^\circ, S^\circ - X^\circ \rangle = 0$$

where $U^\circ = T(Z^\circ) + I_{m(n+1)}$. If U° is full rank than $S^\circ = X^\circ$ is the unique solution. Otherwise, let k be the dimension the kernel of U°, then there exists a matrix $G \in \mathbb{R}^{k \times m(n+1)}$ full row rank such that $U^\circ G^\top = 0$. Since $U^\circ, S^\circ - X^\circ \succeq 0$ it follows that $S^\circ - X^\circ = G^\top H G$ where the unknown $H \in \mathbf{Q}_k$ is positive definite. Next, the sparsity pattern of $Y^\circ = D(S^\circ)$ can be recovered as explained in Zorzi and Sepulchre (2016) by setting $(Y_j^\circ)_{lh} = (Y_j^\circ)_{hl} = 0$, $j = 0, \ldots, n$ for all $(l,h) \in \mathcal{I}$, where the set $\mathcal{I}$ is defined as

$$\mathcal{I} := \{(l,h) : l \neq h, \sum_{j=0}^{n} |(Z_j^\circ)_{lh}| + |(Z_j^\circ)_{hl}| < \gamma\}.$$

Finally, recalling that $Y = D(S) = D(S - X + X)$, H can be obtained by solving the following system of linear equations:

$$[[D(G^\top H G + X^\circ)]_j]_{(l,h)} = 0, \quad j = 0, \ldots, n, \quad (l,h) \in \mathcal{I}.$$

If this system admits a unique solution then the resulting S° is also unique.

4.9 Numerical implementation

To solve numerically problem (4.38) we propose an Alternating Direction Method of Multipliers (ADMM) algorithm, Boyd et al. (2011).
To begin with, problem (4.38) needs to be reformulated in a suitable format for ADMM implementation by decoupling the constraints $I + T(Z) \succeq 0$ and $Z \in \mathcal{Z}$. This is achieved by introducing a new variable $Y \in \mathbf{Q}_{m(n+1)}$ which is defined as $Y := I + T(Z)$. The

reformulated problem is:

$$\begin{aligned} \min_{(\lambda,Z)\in\mathscr{C}_{\lambda,Z},Y\in\mathscr{C}_Y} \quad & \lambda(-\log|\lambda^{-1}W_{\max}|+\int\log|\hat{\Phi}_m|+\delta) \\ \text{subject to} \quad & Y=I+T(Z) \end{aligned} \tag{4.40}$$

where $\mathscr{C}_{\lambda,Z}$ and $\mathscr{C}_Y$ are defined, respectively, as:

$$\begin{aligned} \mathscr{C}_{\lambda,Z} :=& \{(\lambda,Z) : \lambda\in\mathbb{R},\lambda>0,Z\in\mathcal{Z},W_{\max}(\lambda,Z)\succ 0\} \\ \mathscr{C}_Y :=& \{Y\in\mathbf{Q}_{m(n+1)} : Y\succeq 0\}. \end{aligned}$$

Hence, the associated *augmented Lagrangian* is:

$$\begin{aligned} \mathscr{L}_\rho(\lambda,Z,Y,M) =& \lambda(-\log|\lambda^{-1}W_{\max}|+\int\log|\hat{\Phi}_m|+\delta) \\ &+\langle M,Y-I-T(Z)\rangle+\frac{\rho}{2}\|Y-I-T(Z)\|_F^2 \end{aligned}$$

and the ADMM updates are:

$$(\lambda^{(k+1)},Z^{(k+1)}) = \underset{(\lambda,Z)\in\mathscr{C}_{\lambda,Z}}{\arg\min}\ \mathscr{L}_\rho(\lambda,Z,Y^{(k)},M^{(k)}) \tag{4.41}$$

$$Y^{(k+1)} = \underset{Y\in\mathscr{C}_Y}{\arg\min}\ \mathscr{L}_\rho(\lambda^{(k+1)},Z^{(k+1)},Y,M^{(k)}) \tag{4.42}$$

$$M^{(k+1)} = M^{(k)}+\rho(Y^{(k+1)}-I-T(Z^{(k+1)})). \tag{4.43}$$

Problem (4.41) does not admit a closed form solution, therefore we approximate the optimal solution by a gradient projection whose update steps are:

$$\begin{aligned} \lambda^{(k+1)} &= \lambda^{(k)}-t_k\nabla_\lambda\mathscr{L}_\rho(\lambda^{(k)},Z^{(k)},Y^{(k)},M^{(k)})) \\ Z^{(k+1)} &= \Pi_{\mathcal{Z}}(Z^{(k)}-t_k\nabla_Z\mathscr{L}_\rho(\lambda^{(k)},Z^{(k)},Y^{(k)},M^{(k)})) \end{aligned}$$

where:

- t_k is chosen according to Armijo's conditions;
- $\Pi_{\mathcal{Z}}$ is the projector onto $\mathcal{Z}$ which reduces to a projector onto the l_1-norm ball Songsiri and Vandenberghe (2010);
- $\nabla_\lambda\mathscr{L}_\rho(\lambda,Z,Y,M) = -\log|\lambda^{-1}W_{\max}|+\int\log|\hat{\Phi}_m|+\delta-\lambda\,\mathrm{tr}[-\lambda^{-1}I_m+\lambda^{-1}(\hat{R}_0-\mathcal{K}_nT_{1:n,1:n}^{-1}T_{0,1:n}^\top+T_{0,1:n}T_{1:n,1:n}^{-1}\mathcal{T}T_{1:n,1:n}^{-1}T_{0,1:n}^\top-T_{0,1:n}T_{1:n,1:n}^{-1}\mathcal{K}_n^\top)]$;
- $\nabla_Z\mathscr{L}_\rho(\lambda,Z,Y,M) = -\lambda D(\begin{bmatrix}I_m & -T_{0,1:n}T_{1:n,1:n}^{-1}\end{bmatrix}^\top W_{\max}^{-1}\begin{bmatrix}I_m & -T_{0,1:n}T_{1:n,1:n}^{-1}\end{bmatrix})$ $-D(M+\rho(Y-I)+\rho T(Z))$.

On the other hand, problem (4.42) admits a closed form solution which can be easily computed as

$$Y^{(k+1)} = \Pi_{\mathscr{C}_Y}\left(-\frac{1}{\rho}M^{(k)} + I + T(Z^{(k+1)})\right)$$

where $\Pi_{\mathscr{C}_Y}$ is the projector onto the cone of symmetric positive semi-definite matrices of size $m(n+1)\times m(n+1)$.
To define the stopping criterion we need to introduce the following quantities:

$$R^P = Y^{(k)} - I - T(Z^{(k)})$$
$$R^D = \rho(Y^{(k)} - Y^{(k-1)})$$

which are referred to as the primal and the dual residual, respectively.
Then, the algorithm stops when the following conditions are met:

$$\|R^P\|_F \leq m(n+1)\varepsilon^{ABS} + \varepsilon^{REL}\max\{\sqrt{m(n+1)}, \|T(Z^{(k)})\|_F, \|Y^{(k)}\|_F\}$$
$$\|R^D\|_F \leq m\sqrt{n+1}\varepsilon^{ABS} + \varepsilon^{REL}\|D(M^{(k)})\|_F$$

where ε^{ABS} and ε^{REL} are the desired absolute and relative tolerances.

4.10 Numerical simulations

In this section we compare the performance of our estimator in (4.8) with the one in (4.9). To this end we consider an autoregressive process of order $n = 1$, with $m = 15$ manifest variables and $l = 1$ latent variable. The sparsity pattern of the model is depicted in Figure 4.1 (on the left). Two samples of numerosity 500 and 500 have been generated from the model: the first sample has been used to compute the maximum entropy spectrum $\hat{\Phi}_m$ and to estimate the (smoothed) periodogram $\hat{\Phi}_m^{\mathrm{p}}$ while the second sample has been

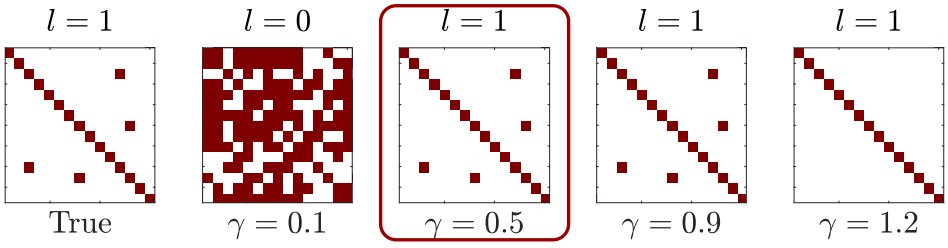


Figure 4.1: Sparsity pattern and number of latent variables for the model generating the data (left) versus sparsity pattern and number of latent variables estimated with the proposed method for $\gamma \in \{0.1, 0.5, 0.9, 1.2\}$ (right). The optimal model, chosen via cross validation, is highlighted by a red square.

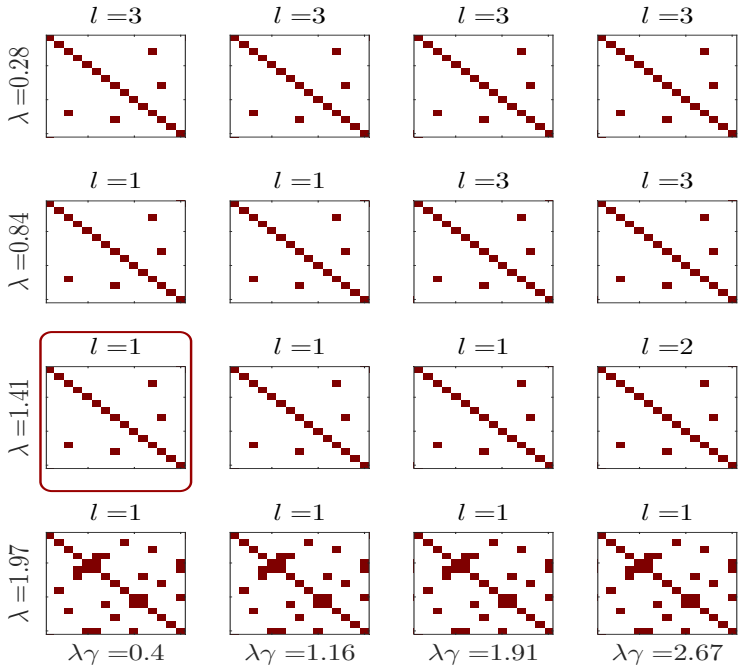


Figure 4.2: Sparsity pattern and number of latent variables estimated with the approach proposed in Zorzi and Sepulchre (2016). Here, $\lambda \in [0.28, 1.97]$ and $\lambda\gamma \in [0.4, 2.67]$. The optimal model is highlighted by a red square.

retained for cross validation. The parameter δ_α has been computed as in Lemma 4.4.2 for $\alpha = 0.5$. Then, problem (4.40) has been solved setting $\varepsilon^{ABS} = 10^{-4}$ and $\varepsilon^{REL} = 10^{-4}$, while the parameter ρ has been updated according to the rule $\rho^{(k)} = \min\{\beta\rho^{(k-1)}, 10^3\}$ for a suitable value of the parameter $\beta > 1$. The results for different values of the regularization parameter γ are depicted in Figure 4.1: as expected increasing the regularization parameter favors less conditional dependence relations among the manifest variables and an increment of the number of latent variables. To discriminate across the different models a cross validation procedure on the second sample of numerosity 500 has been implemented utilizing as ranking criterion $\mathbb{S}(\hat{\Phi}_m^{\mathrm{p}} \| \hat{\Phi}_m^c)$, i.e. the Itakura Saito divergence between the estimated candidate spectrum $\hat{\Phi}_m^c$ and the periodogram $\hat{\Phi}_m^{\mathrm{p}}$. Then, using the same data, we estimate the model by solving problem (4.9), see (Zorzi and Sepulchre, 2016, Section V) for more details. The results for different values of the regularization parameters λ and γ are depicted in Figure 4.2. The same discrimination criteria proposed in Zorzi and Sepulchre (2016) has been adopted. Although the two methods provide the same latent variable graph structure, our approach requires to construct a 1-dimensional regularization path. In contrast, the approach in Zorzi and Sepulchre (2016) requires to construct a 2-dimensional regularization path which requires a more challenging design.

We conclude that our approach brings improvements in terms of computational saving while simplifying the design of the cross validation procedure.

4.11 Concluding remarks

In this chapter we have proposed a new approach for learning latent variable graphical models by using only one regularization parameter. This is accomplished by constructing a confidence set about a finite sample estimate of the spectral density of the process where we search for the optimal model. Such set contains the true model with a prescribed probability and its radius (in the considered topology) depends only on the numerosity of the available data. The resulting procedure turns out to be computationally convenient with respect to the existing approaches: in fact, the cross-validation step for choosing the optimal regularization parameters has been simplified thus reducing the computational burden.
The proposed approach can be also adopted for learning sparse graphical models (without latent variables): in this case the regularization parameter disappears making the identification procedure simpler than the one in Songsiri and Vandenberghe (2010).

Part II

Entropic methods in learning homogeneous random fields

5

Relative entropy for homogeneous random fields

5.1 Introduction

Divergence criteria are ubiquitous in system identification and information theory, Stoorvogel and Van Schuppen (1996), a very popular choice being the *relative entropy* (or *Kullback-Leibler divergence*) interpretable as a measure of similarity between probability distributions. When dealing with discrete-time stochastic processes the natural counterpart of it is the so-called *relative entropy rate* explainable, naïvely, as the rate of growth of the relative entropy. For the case of zero mean stationary Gaussian processes the relative entropy rate can be explicitly expressed in terms of the spectral densities of the processes, Stoorvogel and Van Schuppen (1996), Ihara (1993), we refer to Appendix A for more details. In this sense, the relative entropy rate can be interpreted as a (pseudo-) distance in the cone of positive definite spectral densities, see Baggio, Ferrante, and Sepulchre (2019); Byrnes et al. (2000); Ferrante et al. (2008); Jiang et al. (2012); Georgiou et al. (2009); Ramponi et al. (2009); Chen, Georgiou, Ning, and Tannenbaum (2017) and the references therein for further examples of distances and pseudo-distances between spectral densities and for comparisons among them. Important applications are in spectral estimation and signal process Georgiou and Lindquist (2003); Georgiou (2006); Ferrante et al. (2012); Zorzi (2014, 2015).

In this chapter we consider *discrete, homogeneous, Gaussian random fields*, i.e. stationary Gaussian processes defined over a multidimensional grid: both the case of finite grid (for periodic fields) and infinite grid (for non-periodic fields) are considered. As

it happens for their one-dimensional counterpart, zero-mean homogeneous Gaussian random fields are completely characterized by their spectral density. We remark that, whereas in the one-dimensional setting a rich stream of literature has been developed both in entropy-based methods for spectral estimation and in distances and pseudo-distances between power spectral densities, these topics have received less attention in the multidimensional setting. Among the notable exceptions we mention Georgiou (2006); Ringh et al. (2015); Ringh et al. (2016); Ringh et al. (2017).
Our contribution in this chapter is twofold. On one side, we show that the relative entropy (or relative entropy rate for processes defined over a infinite multidimensional grid) between two homogeneous Gaussian random fields can be computed explicitly in terms of their spectral densities by means of a natural and intuitive formula. The latter can therefore be interpreted as a pseudo-distance in the cone of positive definite multidimensional spectra. On the other side, we show that this formula accounts not only for the relative entropy of the two processes but also for that of the spectral processes associated to them. In the case of one-dimensional infinite grid (discrete line) this specific result specializes to some of the results presented in Ferrante et al. (2012). Our studies are motivated as an effort in the direction of obtaining high-performance algorithms for multidimensional spectral estimation. In fact, to this aim a first fundamental step is that of establishing a sensible distance between multidimensional spectral densities and connecting this distance to the property of the underlying field.

Outline of the chapter

The chapter is organized as follows. In Subsection 5.1.1 some basic facts concerning complex Gaussian random variables are recalled. Thereafter, we present our main results for the case of periodic homogeneous Gaussian random fields in Section 5.2 and for the non periodic case in Section 5.3. In order not to burden the discussion most of the results are derived for the univariate case. The extension to the multivariate case follows, most of the cases, by direct generalization as discussed in Section 5.4 where some other future research directions are also proposed.

Notation conventions

Throughout this chapter we use the convention that when $\mathbf{u}$ and $\mathbf{v}$ are random vectors distributed according to μ and ν, respectively, we refer to the relative entropy between μ and ν as $\mathbb{D}(\mathbf{u}\|\mathbf{v})$.

5.1.1 Review of complex Gaussian random vectors

Let $\mathbf{z}$ be a zero-mean, $\mathbb{C}^n-$valued Gaussian random vector, namely $\mathbf{z} = \mathfrak{Re}[\mathbf{z}] + i\mathfrak{Im}[\mathbf{z}]$, where $\mathfrak{Re}[\mathbf{z}], \mathfrak{Im}[\mathbf{z}]$ are jointly Gaussian. Let $\Gamma_{\mathbf{z}} = \mathbb{E}\{\mathbf{z}\mathbf{z}^*\}$, $C_{\mathbf{z}} := \mathbb{E}\{\mathbf{z}\mathbf{z}^\top\}$ be, respectively, the covariance and relation matrix of $\mathbf{z}$. We denote this situation by $\mathbf{z} \sim \mathscr{N}_{\mathbb{C}}(0, \Gamma_{\mathbf{z}}, C_{\mathbf{z}})$.

The probability density function of $\mathbf{z}$ is the joint probability density function of the $2n-$dimensional vector $[\mathfrak{Re}[\mathbf{z}]^\top \ \mathfrak{Im}[\mathbf{z}]^\top]^\top$. In particular, $[\mathfrak{Re}[\mathbf{z}]^\top \ \mathfrak{Im}[\mathbf{z}]^\top]^\top$ is a zero-mean, $\mathbb{R}^{2n}-$valued, Gaussian random vector whose covariance matrix we denote by $\Sigma_{\mathbf{z}}$. The vector $\mathbf{z}$ is said to be *circularly symmetric* if, for all $\theta \in [-\pi,\pi)$, $\mathbf{z} \sim e^{i\theta}\mathbf{z}$. It can be shown that $\mathbf{z}$ is circularly symmetric if and only if $C_{\mathbf{z}} = 0$. In this case $\mathbb{E}\{\mathfrak{Re}[\mathbf{z}]\mathfrak{Re}[\mathbf{z}]^\top\} = \mathbb{E}\{\mathfrak{Im}[\mathbf{z}]\mathfrak{Im}[\mathbf{z}]^\top\}$.
Let $\mathbf{y} \sim \mathscr{N}_{\mathbb{C}}(0,\Gamma_{\mathbf{y}},C_{\mathbf{y}})$ and $\mathbf{z} \sim \mathscr{N}_{\mathbb{C}}(0,\Gamma_{\mathbf{z}},C_{\mathbf{z}})$. Then, the relative entropy between $\mathbf{z}$ and $\mathbf{y}$ is given by

$$\mathbb{D}(\mathbf{z}\|\mathbf{y}) = \frac{1}{2}\left[\log\det(\Sigma_{\mathbf{z}}^{-1}\Sigma_{\mathbf{y}}) + \mathrm{tr}(\Sigma_{\mathbf{y}}^{-1}\Sigma_{\mathbf{z}}) - 2n\right].$$

If $\mathbf{z}$, $\mathbf{y}$ are circularly symmetric, the expression for the relative entropy simplifies as follows

$$\mathbb{D}(\mathbf{z}\|\mathbf{y}) = \left[\log\det(\Gamma_{\mathbf{z}}^{-1}\Gamma_{\mathbf{y}}) + \mathrm{tr}(\Gamma_{\mathbf{y}}^{-1}\Gamma_{\mathbf{z}}) - n\right].$$

To deal with the complex case we need the following lemma proven in Ferrante et al. (2012).

Lemma 5.1.1. *Let $\mathbf{u}_k$, $\mathbf{v}_k$ be $k-$dimensional, real valued, random vectors.Let $f : \mathbb{R}^k \to \mathbb{R}^h$ be a measurable function and denote by $\mathbf{u}_{kh}$ and $\mathbf{v}_{kh}$, respectively, the augmented vectors $[\mathbf{u}_k^\top \ f(\mathbf{u}_k)^\top]^\top$ and $[\mathbf{v}_k^\top \ f(\mathbf{v}_k)^\top]^\top$. Then*

$$\mathbb{D}(\mathbf{u}_{kh}\|\mathbf{v}_{kh}) = \mathbb{D}(\mathbf{u}_k\|\mathbf{v}_k).$$

5.2 Relative entropy for periodic homogeneous random fields

5.2.1 Multi-level circulant matrices

Given a vector of indices $\mathbf{N} = (N_1,...,N_d) \in \mathbb{N}^d$ a $d-$level circulant matrix $\mathbf{C}$ can be defined recursively as follows. If $d = 1$ then $\mathbf{C}$ is a standard circulant matrix. If $d > 1$ then $\mathbf{C}$ is a block circulant matrix made of $N_1 \times N_1$ blocks each of which is an $(d-1)-$level circulant matrix of indices $(N_2,...,N_d)$. This continues recursively up to the innermost level where each block is made of standard circulant matrices.
We denote by $\mathscr{C}^{\mathbf{N}}$ the set of real multi-level circulant matrices of dimension $|\mathbf{N}| \times |\mathbf{N}|$, with $|\mathbf{N}| := N_1 \times ... \times N_d$.
Multi-level circulant matrices are completely specified by their first row and $\mathbf{N}$. For any given $\mathbf{N} \in \mathbb{N}^d$ we define $\mathbb{N}_{\mathbf{N}}^d := \{(\ell_1,...,\ell_d) : 0 \le \ell_j \le N_j - 1,\ j = 1,...,d\}$. Then for any multi-level matrix $\mathbf{C} \in \mathscr{C}^{\mathbf{N}}$ we address the first row entries as follows: for any $(\ell_1,...,\ell_d) \in \mathbb{N}_{\mathbf{N}}^d$, $\mathbf{C}_{\ell_1,\ell_2,...\ell_d}$ is that element of the first row in the (ℓ_1+1)-th outermost block, in the (ℓ_2+1)-th sub-block of the (ℓ_1+1)-th block and so on up to the innermost sub-block level where it refers to the (ℓ_d+1)-th entry. For a given vector of indices

$\mathbf{N} \in \mathbb{N}_{\mathbf{N}}^d$ we denote by $\mathrm{MCirc}_{\mathbf{N}} : \mathbb{R}^{|\mathbf{N}|} \to \mathscr{C}^{\mathbf{N}}$ the operator mapping a $|\mathbf{N}|-$dimensional vector v into the multi-level circulant matrix whose elements in the first row are the elements of v.

An important property of multi-level circulant matrices is that they can be diagonalized by multidimensional discrete Fourier transform. More precisely, let $\zeta_k := e^{2\pi ik/n}$, $k = 0,...,n-1$, be the $n-$th roots of unity and define the Fourier matrix of order n, which we denote by $\mathbf{F}_n$, as the unitary matrix whose $(k,\ell)-$entry is given by

$$(\mathbf{F}_n)_{(k,\ell)} = \frac{1}{\sqrt{n}} \zeta^{-1}_{(k-1)(\ell-1)}.$$

Then, any matrix $\mathbf{C} \in \mathscr{C}^{\mathbf{N}}$, with $\mathbf{N} = (N_1,...,N_d) \in \mathbb{N}^d$, can be diagonalized by the unitary matrix

$$\mathbf{U}_{\mathbf{N}} := \mathbf{F}_{N_1} \otimes \mathbf{F}_{N_2} \otimes ... \otimes \mathbf{F}_{N_d}.$$

Moreover, the eigenvalues of $\mathbf{C}$ are parametrized by θ by the following formula

$$\sum_{\mathbf{k} \in \mathbb{N}_{\mathbf{N}}^d} \mathbf{C}_{\mathbf{k}} e^{-i(\theta \cdot \mathbf{k})}, \qquad \theta \in \mathbb{T}_{\mathbf{N}}^d, \tag{5.1}$$

where $\mathbb{T}_{\mathbf{N}}^d$ denotes the discrete $d-$dimensional hypercuboid: $\mathbb{T}_{\mathbf{N}}^d = \{(\ell_1 \frac{2\pi}{N_1}, ..., \ell_d \frac{2\pi}{N_d}) : \ell \in \mathbb{N}_{\mathbf{N}}^d\}$ and $(\theta \cdot \mathbf{k})$ denotes the inner product $(\ell_1 k_1 \frac{2\pi}{N_1}, ..., \ell_d k_d \frac{2\pi}{N_d})$.

5.2.2 Spectral representation of periodic homogeneous random fields

Let $\{y(\mathbf{t}); \mathbf{t} \in \mathbb{Z}^d\}$ be a zero-mean, $\mathbb{R}-$valued, $2\mathbf{N}-$periodic, homogeneous random field where $2\mathbf{N} := (2N_1, ..., 2N_d) \in \mathbb{N}^d$ represents the period in each dimension[1], in the sense that $y(\mathbf{t} + \mathbf{n} \cdot 2\mathbf{N}) = y(\mathbf{t})$ almost surely for any $\mathbf{n} \in \mathbb{Z}^d$, where the product $\mathbf{n} \cdot 2\mathbf{N}$ is understood component-wise, and homogeneity generalizes stationarity for the one-dimensional case in the sense that $\mathbb{E}\{y(\mathbf{t}+\mathbf{k})y(\mathbf{t})\}$ is independent of $\mathbf{t}$ for all $\mathbf{k} \in \mathbb{Z}^d$. Given the $2\mathbf{N}-$periodic, homogeneous random field $\{y(\mathbf{t}); \mathbf{t} \in \mathbb{Z}^d\}$, we define the $2^d|\mathbf{N}|-$dimensional column vector

$$\mathbf{y} = [y(0,0,\ldots,0) \mid y(0,0,\ldots,1) \mid \cdots \mid y(2N_1-1, ..., 2N_d-1)]^\top \tag{5.2}$$

obtained by stacking the random variables $y(\mathbf{t})$, $\mathbf{t} \in \mathbb{N}_{\mathbf{2N}}^d$ according to the lexicographic order of the multi-index $\mathbf{t}$. Then, the corresponding $2^d|\mathbf{N}| \times 2^d|\mathbf{N}|$ covariance matrix,

[1] For convenience we assume the period to be even in each dimension. All the results can be re-derived, with minimum variations, for the odd case.

$\Sigma_y := \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\}$, has a multi-level circulant structure:

$$\Sigma_y := \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\} = \mathrm{MCirc}_{2\mathbf{N}}\{c_{(0,\dots,0)},\dots,c_{(2N_1-1,\dots,2N_d-1)}\},$$

where $c_{\mathbf{k}} := \mathbb{E}\{y(\mathbf{t}+\mathbf{k})y(\mathbf{t})\}$ for $\mathbf{k} \in \mathbb{N}^d_{2\mathbf{N}}$ and $c_{\mathbf{k}} = c_{2\mathbf{N}-\mathbf{k}}$.
The spectral density of the process is a non-negative function defined on $\mathbb{T}^d_{2\mathbf{N}}$

$$\Phi_y(\theta) = \sum_{\mathbf{k}\in\mathbb{N}^d_{2\mathbf{N}}} c_{\mathbf{k}} e^{-i(\theta\cdot\mathbf{k})}, \qquad \theta \in \mathbb{T}^d_{2\mathbf{N}}. \tag{5.3}$$

Remark 5.2.1. Comparing formulas (5.1) and (5.3), we observe that the values of the spectral density coincide with the eigenvalues of the multi-level circulant covariance matrix $\Sigma_y := \mathbb{E}\{\mathbf{y}\mathbf{y}^\top\}$.

The spectral process associated to y is

$$\hat{y}(\theta) := \sum_{\mathbf{k}\in\mathbb{N}^d_{2\mathbf{N}}} y(\mathbf{k}) e^{-i(\theta\cdot\mathbf{k})}, \quad \theta \in \mathbb{T}^d_{2\mathbf{N}}. \tag{5.4}$$

It can be easily verified that

$$\mathbb{E}\{\hat{y}(\theta)\} = 0, \quad \theta \in \mathbb{T}^d_{2\mathbf{N}}$$
$$\frac{1}{2^d|\mathbf{N}|}\mathbb{E}\{\hat{y}(\theta)\hat{y}(\eta)^*\} = \Phi_y(\eta)\delta_{\theta\eta}, \quad \theta,\eta \in \mathbb{T}^d_{2\mathbf{N}}.$$

Hence, in analogy with the one-dimensional case Lindquist and Picci (2013), Rozanov (1967), the following spectral representation can be obtained for the process y

$$y(t) = \frac{1}{2^d|\mathbf{N}|}\sum_{\theta\in\mathbb{T}^d_{2\mathbf{N}}} e^{i(\theta\cdot\mathbf{t})}\hat{y}(\theta).$$

5.2.3 Space and spectral domain relative entropy for periodic homogeneous random fields

We are now ready to define the relative entropy for periodic homogeneous random fields.

Definition 5.2.2. *The relative entropy between two be zero-mean, $\mathbb{R}-$valued, $2\mathbf{N}-$periodic, homogeneous random fields $\{y(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ and $\{z(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ is defined by*

$$\mathbb{D}(y\|z) := \mathbb{D}(\mathbf{y}\|\mathbf{z}) \tag{5.5}$$

where $\mathbf{y}$ is the vector associated to y by (5.2) and $\mathbf{z}$ is defined analogously.

The following result provides a formula to compute the relative entropy in the spectral

domain and establishes a natural entropic pseudo-distance in the set of multidimensional discrete spectral densities.

Theorem 5.2.3. *Let* $\{y(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ *and* $\{z(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ *be as in Definition 5.2.2, and* Φ_y *and* Φ_z *be the corresponding spectral densities. Then*

$$\mathbb{D}(y\|z)=\frac{1}{2}\sum_{\theta\in\mathbb{T}^d_{2\mathbf{N}}}\log\left(\Phi_y(\theta)^{-1}\Phi_z(\theta)\right)+\Phi_z(\theta)^{-1}\left(\Phi_y(\theta)-\Phi_z(\theta)\right).$$

Proof. Consider the vector $\mathbf{y}$ associated to $y(\mathbf{t})$ by (5.2) and its covariance matrix $\Sigma_y:=\mathbb{E}\{\mathbf{y}\mathbf{y}^\top\}$ and let the vector $\mathbf{z}$ be associated to $z(\mathbf{t})$ by the same procedure and $\Sigma_z:=\mathbb{E}\{\mathbf{z}\mathbf{z}^\top\}$. Let $\psi_y(\theta),\ \psi_z(\theta),\ \theta\in\mathbb{T}^d_{2\mathbf{N}}$, denote the eigenvalues of Σ_y and Σ_z, respectively, and define $\mathbf{D}_y:=\operatorname{diag}\{\psi_y(\theta);\ \theta\in\mathbb{T}^d_{2\mathbf{N}}\}$, $\mathbf{D}_z:=\operatorname{diag}\{\psi_z(\theta);\ \theta\in\mathbb{T}^d_{2\mathbf{N}}\}$ so that

$$\Sigma_y=\mathbf{U}^*_{2\mathbf{N}}\mathbf{D}_y\mathbf{U}_{2\mathbf{N}},\ \ \Sigma_z=\mathbf{U}^*_{2\mathbf{N}}\mathbf{D}_z\mathbf{U}_{2\mathbf{N}}.$$

Then, by direct computation

$$\begin{aligned}\mathbb{D}(y\|z)=\mathbb{D}(\mathbf{y}\|\mathbf{z})&=\frac{1}{2}\left[\log\det(\Sigma_y^{-1}\Sigma_z)+\operatorname{tr}(\Sigma_z^{-1}\Sigma_y)-2^d|\mathbf{N}|\right]\\&=\frac{1}{2}\left[\log\det(\mathbf{D}_y^{-1}\mathbf{D}_z)+\operatorname{tr}(\mathbf{D}_z^{-1}\mathbf{D}_y)-2^d|\mathbf{N}|\right]\\&=\frac{1}{2}\left[\log\prod_{\theta\in\mathbb{T}^d_{2\mathbf{N}}}\left(\psi_y(\theta)^{-1}\psi_z(\theta)\right)+\sum_{\theta\in\mathbb{T}^d_{2\mathbf{N}}}\left(\psi_z(\theta)^{-1}\psi_y(\theta)\right)-2^d|\mathbf{N}|\right]\\&=\frac{1}{2}\left[\sum_{\theta\in\mathbb{T}^d_{2\mathbf{N}}}\left(\log(\Phi_y(\theta)^{-1}\Phi_z(\theta))+\Phi_z(\theta)^{-1}\Phi_y(\theta)\right)-2^d|\mathbf{N}|\right].\end{aligned}$$

■

We can associate to y and z their spectral processes as in (5.4). This are complex valued processes for which we can naturally define the relative entropy as

$$\mathbb{D}(\hat{y}\|\hat{z}):=\mathbb{D}(\hat{\mathbf{y}}\|\hat{\mathbf{z}})$$

where $\hat{\mathbf{y}}$ is the column vector obtained by stacking the random variables $\hat{y}(\theta)$, $\theta\in\mathbb{T}^d_{2\mathbf{N}}$. Remarkably the relative entropy between y and z equals the relative entropy between their associated spectral processes.

Proposition 5.2.4. *Let* $\{y(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ *and* $\{z(\mathbf{t});\ \mathbf{t}\in\mathbb{Z}^d\}$ *be as in Definition 5.2.2 and let* $\hat{y}(\theta),\hat{z}(\theta),\ \theta\in\mathbb{T}^d_{2\mathbf{N}}$, *be the associated spectral processes defined as in* (5.4)*. Then*

$$\mathbb{D}(\hat{y}\|\hat{z})=\mathbb{D}(y\|z).$$

Proof. Let $\psi_y(\theta),\ \psi_z(\theta),\ \theta \in \mathbb{T}^d_{2\mathbf{N}}$, denote the eigenvalues of Σ_y and Σ_z, respectively. Observe that $\frac{1}{\sqrt{2^d|\mathbf{N}|}}\hat{\mathbf{y}} = \mathbf{U}_{2\mathbf{N}}\mathbf{y}$ and $\frac{1}{\sqrt{2^d|\mathbf{N}|}}\hat{\mathbf{z}} = \mathbf{U}_{2\mathbf{N}}\mathbf{z}$. Therefore

$$\frac{1}{2^d|\mathbf{N}|}\mathbb{E}\{\hat{\mathbf{y}}\hat{\mathbf{y}}^*\} = \mathbf{U}_{2\mathbf{N}}\Sigma_y\mathbf{U}^*_{2\mathbf{N}} = \mathrm{diag}\{\psi_y(\theta); \theta \in \mathbb{T}^d_{2\mathbf{N}}\}$$

$$\frac{1}{2^d|\mathbf{N}|}\mathbb{E}\{\hat{\mathbf{z}}\hat{\mathbf{z}}^*\} = \mathbf{U}_{2\mathbf{N}}\Sigma_z\mathbf{U}^*_{2\mathbf{N}} = \mathrm{diag}\{\psi_z(\theta); \theta \in \mathbb{T}^d_{2\mathbf{N}}\}.$$

Next, notice that for any element $\hat{y}(\theta)$ in $\hat{\mathbf{y}}$ its complex conjugate $\hat{y}(2\mathbf{N}-\theta)$ is also in $\hat{\mathbf{y}}$ and the same holds for $\hat{\mathbf{z}}$. Hence $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$ are not circularly symmetric since $\mathbb{E}\{\hat{y}(\theta)\hat{y}(2\mathbf{N}-\theta)\} = \mathbb{E}\{\hat{y}(\theta)\hat{y}(\theta)^*\} \neq 0$. Let $\mathfrak{T}^d_{2\mathbf{N}} \subset \mathbb{T}^d_{2\mathbf{N}}$ be the set

$$\mathfrak{T}^d_{2\mathbf{N}} := \Big\{ \Big(\ell_1\frac{2\pi}{2N_1},....,\ell_j\frac{2\pi}{2N_j},...,\ell_d\frac{2\pi}{2N_d}\Big) :\ 0 \le \ell_1 < 2N_1,..., \\ 0 \le \ell_j < N_j,....,0 \le \ell_d < 2N_d \Big\}$$

where, in an arbitrary direction j, only half of the indices are considered and define $\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}$ and $\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}$ as the $2^{d-1}|\mathbf{N}|-$dimensional column vectors containing the elements $\{\hat{y}(\theta);\ \theta \in \mathfrak{T}^d_{2\mathbf{N}}\}$ and $\{\hat{z}(\theta);\ \theta \in \mathfrak{T}^d_{2\mathbf{N}}\}$, respectively. Then, by construction, for any element $\hat{y}(\theta)$ in $\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}$ its complex conjugate $\hat{y}(2\mathbf{N}-\theta)$ is not in $\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}$ and the same holds for $\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}$. Hence $\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}$ and $\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}$ are circularly symmetric. Then, recalling that the relative entropy between $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ is the relative entropy between the augmented real random vectors

$$[\mathfrak{Re}[\hat{\mathbf{y}}]^\top\ \ \mathfrak{Im}[\hat{\mathbf{y}}]^\top]^\top = [\mathfrak{Re}[\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}]^\top\ \mathfrak{Re}[\hat{\mathbf{y}}^*_{2^{d-1}\mathbf{N}}]\ \ \mathfrak{Im}[\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}]^\top\ \mathfrak{Im}[\hat{\mathbf{y}}^*_{2^{d-1}\mathbf{N}}]]^\top$$

$$[\mathfrak{Re}[\hat{\mathbf{z}}]^\top\ \ \mathfrak{Im}[\hat{\mathbf{z}}]^\top]^\top = [\mathfrak{Re}[\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}]^\top\ \mathfrak{Re}[\hat{\mathbf{z}}^*_{2^{d-1}\mathbf{N}}]\ \ \mathfrak{Im}[\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}]^\top\ \mathfrak{Im}[\hat{\mathbf{z}}^*_{2^{d-1}\mathbf{N}}]]^\top$$

and applying Lemma 5.1.1 to such augmented vectors it holds that

$$\mathbb{D}(\hat{\mathbf{y}}\|\hat{\mathbf{z}}) = \mathbb{D}(\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}\|\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}).$$

Moreover, we have

$$\mathbb{E}\{\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}\hat{\mathbf{y}}^*_{2^{d-1}\mathbf{N}}\} = \mathrm{diag}\{\psi_y(\theta);\ \theta \in \mathfrak{T}^d_{2\mathbf{N}}\},$$

$$\mathbb{E}\{\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}\hat{\mathbf{z}}^*_{2^{d-1}\mathbf{N}}\} = \mathrm{diag}\{\psi_z(\theta);\ \theta \in \mathfrak{T}^d_{2\mathbf{N}}\}.$$

Therefore, in view of circular symmetry and by direct computation

$$\begin{aligned}\mathbb{D}(\hat{y}\|\hat{z}) &= \mathbb{D}(\hat{\mathbf{y}}_{2^{d-1}\mathbf{N}}\|\hat{\mathbf{z}}_{2^{d-1}\mathbf{N}}) \\ &= \log \prod_{\theta\in\mathfrak{T}^d_{2\mathbf{N}}} (\psi_y(\theta)^{-1}\psi_z(\theta)) + \sum_{\theta\in\mathfrak{T}^d_{2\mathbf{N}}} (\psi_z(\theta)^{-1}\psi_y(\theta)) - 2^{d-1}|\mathbf{N}| \\ &= \frac{1}{2}\left[\sum_{\theta\in\mathbb{T}^d_{2\mathbf{N}}} \log(\Phi_y(\theta)^{-1}\Phi_z(\theta)) + \Phi_z(\theta)^{-1}\Phi_y(\theta) - 2^d|\mathbf{N}|\right].\end{aligned}$$

■

5.3 Relative entropy rate for homogeneous random fields

5.3.1 Multi-level Toeplitz matrices

Given a vector of indices $\mathbf{N} = (N_1,...,N_d) \in \mathbb{N}^d$ a $d-$level Toeplitz matrix $\mathbf{T}$ can be defined recursively as follows. If $d = 1$ then $\mathbf{T}$ is a usual Toeplitz matrix of dimension $N_1 \times N_1$. If $d > 1$ then $\mathbf{T}$ is a block Toeplitz matrix made of $N_1 \times N_1$ blocks each of which is an $(d-1)-$level Toeplitz matrix of indices $(N_2,...,N_d)$. This continues up to the innermost level where each block is made of standard Toeplitz matrices of dimension $N_d \times N_d$.
We denote by $\mathscr{T}^{\mathbf{N}}$ the set of real multi-level Toeplitz matrices of dimension $|\mathbf{N}| \times |\mathbf{N}|$. To address the entries of a multi-level Toeplitz matrix $\mathbf{T}$ a natural notation would be $\mathbf{T}_{(k_1,j_1),...,(k_d,j_d)}$, $k_l, j_l = 1,\ldots,N_l$ where k_1 and j_1 are the row and column indices selecting a block in the outermost level and so on up to to the innermost block for which k_d and j_d that are the row and column indices of a single entry. However, by the multi-level Toeplitz property, any two entries $\mathbf{T}_{(k_1,j_1),...,(k_d,j_d)}$ and $\mathbf{T}_{(\bar{k}_1,\bar{j}_1),...,(\bar{k}_d,\bar{j}_d)}$ are equal whenever $k_l - j_l = \bar{k}_l - \bar{j}_l$ for all $l = 1,\ldots,d$. Therefore we use the more pregnant notation $\mathbf{T}_{\ell_1,...,\ell_d}$, with $\ell_l = k_l - j_l = -(N_l-1),\ldots,(N_l-1)$. For sake of clarity we propose an illustrative example.

Example: Let $\mathbf{N} = (3,2)$. The following shows how the multi-index specifies the entries

of a multi-level Toeplitz matrix $\mathbf{T} \in \mathscr{T}^{\mathbf{N}}$:

$$\mathbf{T} = \left[\begin{array}{cc|cc|cc} \mathbf{T}_{0,0} & \mathbf{T}_{0,-1} & \mathbf{T}_{-1,0} & \mathbf{T}_{-1,-1} & \mathbf{T}_{-2,0} & \mathbf{T}_{-2,-1} \\ \mathbf{T}_{0,1} & \mathbf{T}_{0,0} & \mathbf{T}_{-1,1} & \mathbf{T}_{-1,0} & \mathbf{T}_{-2,1} & \mathbf{T}_{-2,0} \\ \hline \mathbf{T}_{1,0} & \mathbf{T}_{1,-1} & \mathbf{T}_{0,0} & \mathbf{T}_{0,-1} & \mathbf{T}_{-1,0} & \mathbf{T}_{-1,-1} \\ \mathbf{T}_{1,1} & \mathbf{T}_{1,0} & \mathbf{T}_{0,1} & \mathbf{T}_{0,0} & \mathbf{T}_{-1,1} & \mathbf{T}_{-1,0} \\ \hline \mathbf{T}_{2,0} & \mathbf{T}_{2,-1} & \mathbf{T}_{1,0} & \mathbf{T}_{1,-1} & \mathbf{T}_{0,0} & \mathbf{T}_{0,-1} \\ \mathbf{T}_{2,1} & \mathbf{T}_{2,0} & \mathbf{T}_{1,1} & \mathbf{T}_{1,0} & \mathbf{T}_{0,1} & \mathbf{T}_{0,0} \end{array}\right].$$

We also consider sequences $\mathbf{T}^{(n)}$ of multi-level Toeplitz matrices of increasing dimension indices $\mathbf{N}^{(n)} = (N_1^{(n)}, ..., N_d^{(n)})$: a matrix of the sequence is obtained from the previous one by adding new blocks at each level (at the innermost level new entries are added instead of new blocks). Thus, given $\mathbf{T}^{(n+1)}$ (of indices $\mathbf{N}^{(n+1)}$), we obtain $\mathbf{T}^{(n)}$ (of indices $\mathbf{N}^{(n)}$) by taking at each level l of $\mathbf{T}^{(n+1)}$ only the first $N_l \times N_l$ blocks.
Sequences of multi-level Toeplitz matrices are naturally associated with multidimensional Fourier series. In fact, to a given multidimensional sequence $\mathbf{t}_{\ell_1,...,\ell_d}$, with $\ell_l \in \mathbb{Z}$, we can associate both its multidimensional Fourier series

$$f(\theta_1, ..., \theta_d) = \sum_{\ell_1=-\infty}^{\infty} ... \sum_{\ell_d=-\infty}^{\infty} \mathbf{t}_{\ell_1,...,\ell_d} e^{-i(\ell_1\theta_1+...+\ell_d\theta_d)} \tag{5.6}$$

and the sequence $\mathbf{T}^{(n)}$ of multi-level Toeplitz matrix of increasing dimension indices $\mathbf{N}^{(n)} = (N_1^{(n)}, ..., N_d^{(n)})$ such that $\mathbf{T}^{(n)}_{\ell_1,...,\ell_d} = \mathbf{t}_{\ell_1,...,\ell_d}$ for $\ell_l = -(N_l^{(n)} - 1), \ldots, (N_l^{(n)} - 1)$. In this case, we say that $\mathbf{T}^{(n)}$ is the sequence of multi-level Toeplitz matrix associated with f.
We recall some useful results on the asymptotic distribution of eigenvalues of multi-level Toeplitz matrices associated with multidimensional Fourier series established in Tyrtyshnikov (1996), Tyrtyshnikov (1994). Let $\{\lambda_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{N}_{\mathbf{N}}^d}$ be a multidimensional sequence, namely a sequence indexed by a multidimensional index $\mathbf{k} \in \mathbb{N}_{\mathbf{N}}^d$. Let h be a Lebesgue integrable function on $[0, 2\pi]^d$. The sequence $\{\lambda_{\mathbf{K}}\}_{\mathbf{k}\in\mathbb{N}_{\mathbf{N}}^d}$ is said to be *distributed as* $h(\theta_1, ..., \theta_d)$ if for every continuous function F with bounded support it holds that Tyrtyshnikov (1996)

$$\lim_{N_1,...,N_d\to\infty} \frac{1}{|\mathbf{N}|} \sum_{\mathbf{k}\in\mathbb{N}_{\mathbf{N}}^d} F(\lambda_{\mathbf{k}}) = \frac{1}{(2\pi)^d} \int_0^{2\pi} ... \int_0^{2\pi} F(h(\theta_1, ..., \theta_d)) d\theta_1 ... d\theta_d.$$

The following extension of the Szegö Theorem to the multi-level case has been established in Tyrtyshnikov (1996) (Theorem 8.2 in Tyrtyshnikov (1996)).

Theorem 5.3.1. *Let* $\mathbf{T}^{(n)}$ *be the sequence of multi-level Toeplitz matrices associated with*

the multidimensional Fourier series (5.6). If $f(\theta_1,...,\theta_2) \in \mathcal{L}^2$ is a real valued function then the sequence of the eigenvalues of $\mathbf{T}^{(n)}$ is distributed as $f(\theta_1,...,\theta_d)$.

The next result on the asymptotic distribution of the eigenvalues of matrices obtained from basic operations on multi-level Toeplitz matrices has been established in Tyrtyshnikov (1994) (Theorem 5.1 in Tyrtyshnikov (1994)).

Theorem 5.3.2. *Let $f(\theta_1,...,\theta_d)$, $g(\theta_1,...,\theta_d) \in \mathcal{L}^\infty$ and let $\mathbf{T}^{(n)}(f)$, $\mathbf{T}^{(n)}(g)$ be the associated sequences of multi-level Toeplitz matrices. Assume that*

$$\inf_{\theta_1,...,\theta_d} g(\theta_1,...,\theta_d) \equiv \gamma > 0.$$

Then the sequence of the eigenvalues of the product $\mathbf{T}^{(n)}(f)\mathbf{T}^{(n)}(g)$ is distributed as $f(\theta_1,...,\theta_d)g(\theta_1,...,\theta_d)$ while the sequence of the eigenvalues of $\mathbf{T}^{(n)}(f)\mathbf{T}^{(n)}(g)^{-1}$ is distributed as $f(\theta_1,...,\theta_d)/g(\theta_1,...,\theta_d)$.

5.3.2 Spectral representation of homogeneous random fields

Let $\{y(\mathbf{t}); \mathbf{t} \in \mathbb{Z}^d\}$ be a zero-mean, $\mathbb{R}-$valued, purely non deterministic, homogeneous random field. Let $\mathbf{N} := (N_1,...,N_d)$ be a set of indices $N_1,...,N_d \in \mathbb{N}$. We define the set $\mathbb{Z}^d_{\mathbf{N}} \subset \mathbb{Z}^d$ as $\mathbb{Z}^d_{\mathbf{N}} := \{(\ell_1,...,\ell_d);\ -N_1 \le \ell_1 \le N_1 - 1,...,-N_d \le \ell_d \le N_d - 1\}$ and we consider the restriction $\{y(\mathbf{t}); \mathbf{t} \in \mathbb{Z}^d_{\mathbf{N}}\}$ of the random field y to the indexes $\mathbf{t} \in \mathbb{Z}^d_{\mathbf{N}}$. Then we construct the $2^d|\mathbf{N}|$-dimensional random vector $\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}$ by stacking the random variables in $\{y(\mathbf{t}); \mathbf{t} \in \mathbb{Z}^d_{\mathbf{N}}\}$ according to the lexicographic order of the index $\mathbf{t}$. As the indices of $\mathbf{N}$ increase, the covariance matrices $\mathbb{E}\{\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}\mathbf{y}^\top_{[-\mathbf{N},\mathbf{N}]}\}$ define the sequence of multi-level (symmetric) Toeplitz matrices associated (in the sense explained before) to the multidimensional Fourier transform of the covariance function $c : \mathbb{Z}^d \to \mathbb{R}$ of y defined by $c_{\mathbf{t}} = \mathbb{E}\{y(\mathbf{t}+\mathbf{k})y(\mathbf{k})\}$, $\mathbf{t},\mathbf{k} \in \mathbb{Z}^d$.
The results on spectral representation of stationary processes generalizes to homogeneous random fields, Cramér and Leadbetter (1967), Adler (2010). Indeed the covariance function c admits the spectral representation

$$c_{\mathbf{t}} = \int_{\mathbb{T}^d} e^{i(\theta\cdot\mathbf{t})} dF_y(\theta)$$

where $\mathbb{T}^d := [-\pi,\pi]^d$ and F is the spectral distribution function of the process y. Analogously, the process y admits itself a spectral representation

$$y(\mathbf{t}) = \int_{\mathbb{T}^d} e^{i(\theta\cdot\mathbf{t})} d\hat{y}(\theta)$$

where $\hat{y}(\theta)$ is a random field with orthogonal increments defined up to an additive random

variable. If this is fixed by taking $\hat{y}(-\pi,...,-\pi)=0$, then

$$\mathbb{E}\{\hat{y}(\theta)\}=0,\quad \mathbb{E}\{|\hat{y}(\theta)|^2\}=F_y(\theta),\quad \mathbb{E}\{|d\hat{y}(\theta)|^2\}=dF_y(\theta).$$

Since the random field y is assumed to be purely non deterministic $dF_y(\theta)=\Phi_y(\theta)d\theta$ where Φ_y is the spectral density of the random field y.

5.3.3 Space and spectral domain relative entropy rate for homogeneous random fields

Let $\{y(\mathbf{t});\mathbf{t}\in\mathbb{Z}^d\}$, $\{z(\mathbf{t});\mathbf{t}\in\mathbb{Z}^d\}$ be two zero-mean, $\mathbb{R}-$valued, purely non deterministic, homogeneous random fields and denote by Φ_y, Φ_z their spectral densities. For a given vector of indices $\mathbf{N}\in\mathbb{N}^d$, $\mathbf{N}:=(N_1,...,N_d)$ let $\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}$, $\mathbf{z}_{[-\mathbf{N},\mathbf{N}]}$ be, as before, the random (column) vectors obtained by considering the restrictions of the processes y and z, respectively, to the indices $\mathbf{t}\in\mathbb{Z}^d_{\mathbf{N}}$ and stacking the corresponding elements in lexicographic order.
In analogy with Definition A.2.1, we define the relative entropy rate between y and z as follows.

Definition 5.3.3. *The relative entropy rate between y and z is defined as*

$$\mathbb{D}_r(y\|z):=\lim_{N_1\to\infty}...\lim_{N_d\to\infty}\frac{1}{2N_1\times...\times 2N_d}\mathbb{D}(\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}\|\mathbf{z}_{[-\mathbf{N},\mathbf{N}]})$$

provided that the limit exist.

The following result establishes an explicit expression for the relative entropy rate between y and z in terms of their spectral densities.

Theorem 5.3.4. *Let $\{y(\mathbf{t});\mathbf{t}\in\mathbb{Z}^d\}$, $\{z(\mathbf{t});\mathbf{t}\in\mathbb{Z}^d\}$, be two zero-mean, $\mathbb{R}-$valued, purely non deterministic, homogeneous random fields and assume that their spectral densities Φ_y, Φ_z are coercive and bounded. Then*

$$\mathbb{D}_r(y\|z)=\frac{1}{2(2\pi)^d}\int_{\mathbb{T}^d}\log\left(\Phi_y(\theta)^{-1}\Phi_z(\theta)\right)+\Phi_z(\theta)^{-1}\left(\Phi_y(\theta)-\Phi_z(\theta)\right)d\theta.$$

Proof. The covariance matrices $\mathbf{T}_{\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}}:=\mathbb{E}\{\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}\mathbf{y}^\top_{[-\mathbf{N},\mathbf{N}]}\}$, $\mathbf{T}_{\mathbf{z}_{[-\mathbf{N},\mathbf{N}]}}:=\mathbb{E}\{\mathbf{z}_{[-\mathbf{N},\mathbf{N}]}\mathbf{z}^\top_{[-\mathbf{N},\mathbf{N}]}\}$ are multi-level Toeplitz.

Hence, in view of theorems 5.3.1 and 5.3.2

$$\begin{aligned}
&\lim_{N_1,\dots,N_d\to\infty}\frac{1}{2^d|\mathbf{N}|}\mathbb{D}(\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}\|\mathbf{z}_{[-\mathbf{N},\mathbf{N}]})\\
&\quad=\lim_{N_1,\dots,N_d\to\infty}\frac{1}{2^d|\mathbf{N}|}\frac{1}{2}\sum_{\mathbf{k}\in\mathbb{Z}^d_{\mathbf{N}}}\Big[-\log\lambda_{\mathbf{k}}(\mathbf{T}_{\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}})+\log\lambda_{\mathbf{k}}(\mathbf{T}_{\mathbf{z}_{[-\mathbf{N},\mathbf{N}]}})\\
&\qquad\qquad+\lambda_{\mathbf{k}}(\mathbf{T}^{-1}_{\mathbf{z}_{[-\mathbf{N},\mathbf{N}]}}\mathbf{T}_{\mathbf{y}_{[-\mathbf{N},\mathbf{N}]}})+1\Big]\\
&\quad=\frac{1}{2(2\pi)^d}\int_{\mathbb{T}^d}\log\left(\Phi_y(\theta)^{-1}\Phi_z(\theta)\right)+\Phi_z(\theta)^{-1}\left(\Phi_y(\theta)-\Phi_z(\theta)\right)d\theta.
\end{aligned}$$

■

In view of the results recalled in Subsection 5.3.2 y and z admit the following spectral representations

$$\begin{aligned}
y(\mathbf{t})&=\int_{\mathbb{T}^d}e^{i(\theta\cdot\mathbf{t})}d\hat{y}(\theta),\quad \mathbb{E}\{d\hat{y}(\theta)d\hat{y}(\theta)^*\}=\Phi_y(\theta)d\theta\\
z(\mathbf{t})&=\int_{\mathbb{T}^d}e^{i(\theta\cdot\mathbf{t})}d\hat{z}(\theta),\quad \mathbb{E}\{d\hat{z}(\theta)d\hat{z}(\theta)^*\}=\Phi_z(\theta)d\theta.
\end{aligned}$$

We consider the following partition of $\mathbb{T}^d$, $\mathbb{T}^d=\Delta\theta_1\sqcup\ldots\sqcup\Delta\theta_{(2N)^d}$, where the $\Delta\theta_j$'s are $d-$dimensional disjoint hypercubes of edges $\frac{\pi}{N}$, $N\in\mathbb{N}$, with $|\Delta\theta_j|=|\Delta\theta_k|\ \forall\ j,k=1,\ldots,(2N)^d$. Observe that, in view of the fact that y and z are real valued, the following symmetry relations hold

$$\hat{y}(\Delta\theta)^*=\hat{y}(-\Delta\theta),\quad \hat{z}(\Delta\theta)^*=\hat{z}(-\Delta\theta)$$

where $-\Delta\theta:=\{\lambda\in\mathbb{T}^d\mid-\lambda\in\Delta\theta\}$. In analogy with Ferrante et al. (2012) we define the random vectors

$$\hat{\mathbf{y}}_{(2N)^d}:=\begin{bmatrix}\hat{y}(\Delta\theta_1)\\ \vdots\\ \hat{y}(\Delta\theta_{(2N)^d})\end{bmatrix},\quad \hat{z}_{(2N)^d}:=\begin{bmatrix}\hat{z}(\Delta\theta_1)\\ \vdots\\ \hat{z}(\Delta\theta_{(2N)^d})\end{bmatrix}.$$

The following appears to be a sensible definition of relative entropy between the spectral processes.

Definition 5.3.5. *The spectral relative entropy between the random fields y and z is defined by the following limit, provided it exists:*

$$\mathbb{D}_r(d\hat{y}\|d\hat{z}):=\lim_{N\to\infty}\frac{1}{(2N)^d}\mathbb{D}(\hat{\mathbf{y}}_{(2N)^d}\|\hat{\mathbf{z}}_{(2N)^d}).$$

The following result connects the relative entropy rate between y and z with the spectral entropy rate between their associated spectral processes.

Theorem 5.3.6. *Let y and z as defined above. Moreover, assume that both Φ_y and Φ_z are piecewise continuous, coercive, spectral densities. Then, the following holds:*

$$\mathbb{D}_r(y\|z) = \mathbb{D}_r(d\hat{y}\|d\hat{z}).$$

Proof. Observe that for each element $\hat{y}(\Delta\theta_j)$ in $\hat{\mathbf{y}}_{(2N)^d}$ its complex conjugate is also in $\hat{\mathbf{y}}_{(2N)^d}$ and the same holds for $\hat{\mathbf{z}}_{(2N)^d}$. Hence, $\hat{\mathbf{y}}_{(2N)^d}$ and $\hat{\mathbf{z}}_{(2N)^d}$ are not circularly symmetric since $\mathbb{E}\{\hat{y}(\Delta\theta)\hat{y}(-\Delta\theta)\} = \mathbb{E}\{\hat{y}(\Delta\theta)\hat{y}(\Delta\theta)^*\} \neq 0$ and $\mathbb{E}\{\hat{z}(\Delta\theta)\hat{z}(-\Delta\theta)\} = \mathbb{E}\{\hat{z}(\Delta\theta)\hat{z}(\Delta\theta)^*\} \neq 0$.
Let $\hat{\mathbf{y}}_{2^{d-1}N^d}$ be the $2^{d-1}N^d-$dimensional complex random vector containing either the element $\hat{y}(\Delta\theta_j)$ or its complex conjugate and denote by $\mathfrak{J}$ the set of indices $\{j\,;\,\hat{y}(\Delta\theta_j) \in \hat{\mathbf{y}}_{2^{d-1}N^d}\}$. Analogously, let $\hat{\mathbf{z}}_{2^{d-1}N^d}$ be the $2^{d-1}N^d-$dimensional complex random vector with elements $\{\hat{z}(\Delta\theta_j)\,;\,j \in \mathfrak{J}\}$. Then, $\hat{\mathbf{y}}_{2^{d-1}N^d}$ and $\hat{\mathbf{z}}_{2^{d-1}N^d}$ are, by construction, circularly symmetric and have independent elements. In view of Lemma 5.1.1,

$$\mathbb{D}(\hat{\mathbf{y}}_{2\mathbf{N}}\|\hat{\mathbf{z}}_{2\mathbf{N}}) = \mathbb{D}(\hat{\mathbf{y}}_{2^{d-1}N^d}\|\hat{\mathbf{z}}_{2^{d-1}N^d})$$

and, by independence of the elements in $\hat{\mathbf{y}}_{2^{d-1}N^d}$, $\hat{\mathbf{z}}_{2^{d-1}N^d}$, the following decomposition holds

$$\mathbb{D}(\hat{\mathbf{y}}_{2^{d-1}N^d}\|\hat{\mathbf{z}}_{2^{d-1}N^d}) = \sum_{j\in\mathfrak{J}} \mathbb{D}(\hat{y}(\Delta\theta_j)\|\hat{z}(\Delta\theta_j)).$$

By the circular symmetry of $\hat{y}(\Delta\theta_j)$, $\hat{z}(\Delta\theta_j)$

$$\mathbb{D}(\hat{y}(\Delta\theta_j)\|\hat{z}(\Delta\theta_j)) = \log(Q_y^{-1}(\Delta\theta_j)Q_z(\Delta\theta_j)) + \mathrm{tr}(Q_z^{-1}(\Delta\theta_j)Q_y(\Delta\theta_j)) - 1$$

where

$$Q_y(\Delta\theta_j) := \int_{\Delta\theta_j} \Phi_y(\theta)d\theta, \quad Q_z(\Delta\theta_j) := \int_{\Delta\theta_j} \Phi_z(\theta)d\theta.$$

Thus, by piecewise continuity and applying the mean value theorem, we have that, except for a finite number of j's

$$\begin{aligned}\mathbb{D}(\hat{y}(\Delta\theta_j)\|\hat{z}(\Delta\theta_j)) =& \log\left[(\Phi_y(\bar{\theta}_j)|\Delta\theta_j|)^{-1}\Phi_z(\bar{\theta}_j)|\Delta\theta_j|\right] \\ &+ \mathrm{tr}\left[(\Phi_z(\bar{\theta}_j)|\Delta\theta_j|)^{-1}\Phi_y(\bar{\theta}_j)|\Delta\theta_j|\right] - 1\end{aligned}$$

where $\Delta\theta_{j,\ell} \leq \bar{\theta}_j \leq \Delta\theta_{j,u}$ component-wise for $\Delta\theta_{j,\ell}$, $\Delta\theta_{j,u}$ the component-wise lower

and upper boundaries of the $d-$dimensional subinterval $\Delta\theta_j$. Therefore

$$\begin{aligned}
\mathbb{D}_r(d\hat{y}\|d\hat{z}) &= \lim_{N\to\infty} \frac{1}{(2N)^d}\mathbb{D}(\hat{\mathbf{y}}_{2^{d-1}N^d}\|\hat{\mathbf{z}}_{2^{d-1}N^d}) \\
&= \lim_{N\to\infty} \frac{1}{(2N)^d}\sum_{j\in\mathfrak{J}}\mathbb{D}(\hat{y}(\Delta\theta_j)\|\hat{z}(\Delta\theta_j)) \\
&= \lim_{N\to\infty} \frac{1}{(2N)^d}\sum_{j\in\mathfrak{J}}\log\left[\Phi_y(\bar{\theta}_j)^{-1}\Phi_z(\bar{\theta}_j)\right] \\
&\qquad + \operatorname{tr}\left[\Phi_z(\bar{\theta}_j)^{-1}\Phi_y(\bar{\theta}_j)\right] - 1 \\
&= \lim_{N\to\infty} \frac{1}{(2\pi)^d}\frac{(2\pi)^d}{(2N)^d}\sum_{j\in\mathfrak{J}}\log\left[\Phi_y(\bar{\theta}_j)^{-1}\Phi_z(\bar{\theta}_j)\right] \\
&\qquad + \operatorname{tr}\left[\Phi_z(\bar{\theta}_j)^{-1}\Phi_y(\bar{\theta}_j)\right] - 1 \\
&= \lim_{N\to\infty} \frac{1}{2(2\pi)^d}\left(\frac{\pi}{N}\right)^d \sum_{j=1}^{(2N)^d}\log\left[\Phi_y(\bar{\theta}_j)^{-1}\Phi_z(\bar{\theta}_j)\right] \\
&\qquad + \operatorname{tr}\left[\Phi_z(\bar{\theta}_j)^{-1}\Phi_y(\bar{\theta}_j)\right] - 1 \\
&= \frac{1}{2(2\pi)^d}\int_{\mathbb{T}^d}\log\left[\Phi_y(\theta)^{-1}\Phi_z(\theta)\right] \\
&\qquad + \operatorname{tr}\left[\Phi_z(\theta)^{-1}\left(\Phi_y(\theta)-\Phi_z(\theta)\right)\right]d\theta.
\end{aligned}$$

■

5.4 Concluding remarks and future directions

In this chapter we have investigated the connections between space and spectral domain relative entropy for homogeneous (periodic and non-periodic) random fields and established an explicit formula to express the relative entropy in terms of the spectral densities of the fields. This yields a genuine entropic pseudo-distance in the cone of multidimensional spectral densities.

It is worth highlighting that some of the results of Section 5.2 may be seen as special cases of a more general and profound fact. For example, Proposition 5.2.4 can be alternatively established as a consequence of the following more general lemma.

Lemma 5.4.1. *Let* $\mathbf{u}$, $\mathbf{v}$ *be* $k-$*dimensional, real valued, random vectors. Let* F *be a measurable bounded function,* $F:\mathbb{R}^k\to\mathbb{C}^k$ *and assume that there exists a measurable bounded function* $G:\mathbb{C}^k\to\mathbb{C}^k$ *bijective and such that* G^{-1} *is measurable and* $G\,|_{\mathbb{R}^k}=F$. *Then*

$$\mathbb{D}(\mathbf{u}\|\mathbf{v}) = \mathbb{D}(F(\mathbf{u})\|F(\mathbf{v})).$$

Proof. Let $\bar{\mathbf{u}} := \mathbf{u} + if(\mathbf{u})$, $\bar{\mathbf{v}} := \mathbf{v} + if(\mathbf{v})$, where $f \equiv 0$. Clearly, $\mathbb{D}(\mathbf{u}\|\mathbf{v}) = \mathbb{D}(\bar{\mathbf{u}}\|\bar{\mathbf{v}})$. Next, define $\hat{\mathbf{u}} := G(\bar{\mathbf{u}})$, $\hat{\mathbf{v}} := G(\bar{\mathbf{v}})$ and consider the augmented vectors $[\bar{\mathbf{u}}^\top\ \hat{\mathbf{u}}^\top]^\top$, $[\bar{\mathbf{v}}^\top\ \hat{\mathbf{v}}^\top]^\top$. Then the relative entropy between complex random vectors $[\bar{\mathbf{u}}^\top\ \hat{\mathbf{u}}^\top]^\top$, $[\bar{\mathbf{v}}^\top\ \hat{\mathbf{v}}^\top]^\top$ is, by definition, the relative entropy between the real random vectors

$$[\mathfrak{Re}[\bar{\mathbf{u}}]^\top\ \mathfrak{Re}[G(\bar{\mathbf{u}})]^\top\ \mathfrak{Im}[\bar{\mathbf{u}}]^\top\ \mathfrak{Im}[G(\bar{\mathbf{u}})]^\top]^\top,$$

$$[\mathfrak{Re}[\bar{\mathbf{v}}]^\top\ \mathfrak{Re}[G(\bar{\mathbf{v}})]^\top\ \mathfrak{Im}[\bar{\mathbf{v}}]^\top\ \mathfrak{Im}[G(\bar{\mathbf{v}})]^\top]^\top$$

G is measurable, hence so are its real and imaginary parts. Thus, in view of Lemma 5.1.1

$$\mathbb{D}\left(\begin{bmatrix}\bar{\mathbf{u}}\\ \hat{\mathbf{u}}\end{bmatrix}\|\begin{bmatrix}\bar{\mathbf{v}}\\ \hat{\mathbf{v}}\end{bmatrix}\right) = \mathbb{D}(\bar{\mathbf{u}}\|\bar{\mathbf{v}}).$$

On the other hand, G is invertible therefore $[\bar{\mathbf{u}}^\top\ \hat{\mathbf{u}}^\top]^\top = [G^{-1}(\hat{\mathbf{u}})^\top\ \hat{\mathbf{u}}^\top]^\top$ and $[\bar{\mathbf{v}}^\top\ \hat{\mathbf{v}}^\top]^\top = [G^{-1}(\hat{\mathbf{v}})^\top\ \hat{\mathbf{v}}^\top]^\top$ and, in view of the measurability of G^{-1}, from Lemma 5.1.1 it follows that

$$\mathbb{D}\left(\begin{bmatrix}\bar{\mathbf{u}}\\ \hat{\mathbf{u}}\end{bmatrix}\|\begin{bmatrix}\bar{\mathbf{v}}\\ \hat{\mathbf{v}}\end{bmatrix}\right) = \mathbb{D}(\hat{\mathbf{u}}\|\hat{\mathbf{v}}).$$

and therefore $\mathbb{D}(\mathbf{u}\|\mathbf{v}) = \mathbb{D}(F(\mathbf{u})\|F(\mathbf{v}))$ as desired. ∎

A result in the same vein for the infinite dimensional case is much more delicate and is left for future investigation.

As mentioned in the introduction to this chapter in order not to overwhelm the discussion the presented results have been worked out in the univariate setting. However, it is worth observing that most of them generalize directly to the multivariate setting.

In particular the results in Section 5.2 may be extended to the multivariate case in view of the block-diagonalization property of multi-level $m \times m-$block circulant matrices with respect to the unitary matrix $\mathbf{I}_m \otimes \mathbf{U}_\mathbf{N}$. The extension of the contents of Section 5.3 to the multivariate case may be worked out relying on Lemma 4 and Lemma 5 in Oudin and Delmas (2008) which establish a result similar to Theorem 5.3.2 for multi-level block Toeplitz matrices under, however, some different technical assumptions.

6
Summary and outlook

In the first part of this dissertation we have dealt with the problem of robustness in the identification of stochastic models with latent variables. Motivated by this, we have considered the realistic situation in which only a finite sample estimate of the underlying model is available. Then, to account for the uncertainty in the estimation we have proposed a novel approach to construct a confidence region around the given estimate. This has been done relying on certain invariance properties of the relative entropy and of the relative entropy rate.

The proposed paradigm has been detailed for the case of zero-mean Gaussian random variables in Chapter 2 where it has been employed to cope with the problem of robustly estimating the number of factors in finite sample factor analysis. The same paradigm has been applied in Chapter 3 to the robust identification of latent variable graphical models.

Motivated by the problem of robustly identifying latent variable autoregressive graphical models, the proposed paradigm has then been extended to the case of zero-mean Gaussian random processes in Chapter 4.

It is apparent that the applications of this approach are not limited to the problems considered in this dissertation. In fact, it can be applied to a wider class of problems in which one may want to learn a structured second-order model compatible with a finite sample estimate of the process. Further applications may include optimal control in those realistic situations in which only a noisy estimate of the system is available. In fact, in these cases one may want to account for the uncertainty in the estimation by constructing a confidence region about the estimated nominal system and then controlling the "worst system" in the prescribed confidence region.

In the second part of this dissertation we have shown that the relative entropy rate between

two zero-mean homogeneous Gaussian random fields can be computed explicitly in terms of their spectral densities. This has been achieved relying on the properties of multilevel circulant and multilevel Toeplitz matrices. Consequently, a natural entropic pseudo-distance on the cone of positive definite multidimensional spectral densities has been obtained which may be employed in a wide range of applications in multidimensional spectral analysis and estimation.

Among them, a natural one concerns a classical problem in spectral estimation, the celebrated rational covariance extension problem, in which for a given partially specified covariance sequence we seek for an infinite extension such that the corresponding spectral density is non-negative and rational. Among the several variants of this problem that have been studied in the last decades in the unidimensional setting we have the following one: given a partially specified covariance sequence we seek for the spectral density which matches such sequence while minimizing a pseudo-distance with a prior spectral density encoding the a priori information about the system. It is evident that the derived pseudo-distance is a natural candidate for addressing such question in the multidimensional setting. In particular, we observe that for the case of periodic random fields this may lead to a convenient matrix formulation in which we seek for a multi-level circulant covariance completion.

A

Relative entropy & relative entropy rate

A.1 Relative Entropy

The *relative entropy* is a fascinating concept pervading many fields of pure and applied science including information theory, probability, statistics, physics, and data science. In this appendix we briefly recall some of its basic properties. In doing so we refer mainly to Dupuis and Ellis (2011).
Let $\mathscr{X}$ be a Polish space and $\mathcal{B}(\mathscr{X})$ be the $\sigma-$algebra of Borel sets over $\mathscr{X}$. We denote by $\mathscr{P}(\mathscr{X})$ the set of probability measures on $\mathcal{B}(\mathscr{X})$.

Definition A.1.1. *Let $\mu, \nu \in \mathscr{P}(\mathscr{X})$. Then, the relative entropy, or Kullback Leibler divergence, between μ and ν is defined as*

$$\mathbb{D}(\mu\|\nu) = \begin{cases} \int_{\mathscr{X}} \log \frac{d\mu}{d\nu} d\mu, & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases}$$

where $\ll$ denotes absolute continuity of measures.

In particular, for the case in which μ, ν are multivariate normal distributions with zero-mean and non singular covariances matrices Σ, $\Upsilon \in \mathbb{R}^{m\times m}$, respectively, the integral can be explicitly computed and the relative entropy turns out to be given by

$$\mathbb{D}(\mu\|\nu) = \frac{1}{2}\left[\log\det(\Sigma^{-1}\Upsilon) + \mathrm{tr}(\Upsilon^{-1}\Sigma) - m\right].$$

Some properties of the relative entropy can be readily derived by its definition. In fact, if $\mu \ll \nu$ by the Radon-Nykodym Theorem there exists $\frac{d\mu}{d\nu}$ which is uniquely determined $\nu-$almost surely. In this case

$$\mathbb{D}(\mu||\nu) = \int_{\mathscr{X}} \frac{d\mu}{d\nu} \log\left(\frac{d\mu}{d\nu}\right) d\nu.$$

Evidently, $\lim_{x\to 0^+} x\log(x) = 0$. Moreover, $x\log(x) \geq x-1$ for all $x \geq 0$ which holds with equality if and only if $x = 1$. Hence, $\mathbb{D}(\mu||\nu) \geq 0$ and $\mathbb{D}(\mu||\nu) = 0$ if and only if $\mu = \nu$.

The following lemma (Dupuis and Ellis (2011)) collects some appealing properties of the relative entropy, namely convexity, lower semi-continuity, compactness of level sets and approximation by sums. We refer to (Dupuis and Ellis, 2011, Lemma 1.4.3) for a proof of these results.

Lemma A.1.2. *Let $\mathscr{X}$ be a Polish space.*

a.) Let $\mathcal{C}_b(\mathscr{X})$ be the space of bounded continuous functions from $\mathscr{X}$ to $\mathbb{R}$ and let $\Psi_b(\mathscr{X})$ be the space of bounded Borel-measurable functions from $\mathscr{X}$ to $\mathbb{R}$. Then for each $\mu, \nu \in \mathscr{P}(\mathscr{X})$

$$\mathbb{D}(\mu\|\nu) = \sup_{f\in\mathcal{C}_b(\mathscr{X})} \left\{\int_{\mathscr{X}} f d\mu - \log\int_{\mathscr{X}} e^f d\nu\right\} = \sup_{g\in\Psi_b(\mathscr{X})} \left\{\int_{\mathscr{X}} g d\mu - \log\int_{\mathscr{X}} e^g d\nu\right\}.$$

This formula is known as the Donsker-Varadhan variational formula.

b.) $\mathbb{D}(\mu\|\nu)$ is a convex, lower semi-continuous function of $(\mu,\nu) \in \mathscr{P}(\mathscr{X}) \times \mathscr{P}(\mathscr{X})$ and a convex, lower semi-continuous function of each variable separately. Moreover, for a fixed $\nu \in \mathscr{P}(\mathscr{X})$, $\mathbb{D}(\cdot\|\nu)$ is strictly convex on the set $\{\mu \in \mathscr{P}(\mathscr{X}) : \mathbb{D}(\mu\|\nu) < \infty\}$.

c.) For all $\nu \in \mathscr{P}(\mathscr{X})$ $\mathbb{D}(\cdot\|\nu)$ has compact level sets, i.e. for each $M < \infty$ $\{\mu \in \mathscr{P}(\mathscr{X}) : \mathbb{D}(\mu\|\nu) \leq M\}$ is a compact subset of $\mathscr{P}(\mathscr{X})$.

d.) Let Π denote the set of finite measurable partitions of $\mathscr{X}$ [1]*. Then, for all $\mu, \nu \in \mathscr{P}(\mathscr{X})$*

$$\mathbb{D}(\mu||\nu) = \sup_{\pi\in\Pi} \sum_{A\in\pi} \mu(A) \log\left(\frac{\mu(A)}{\nu(A)}\right)$$

where $x\log\left(\frac{x}{y}\right) = 0$ if $x = 0$, $x\log\left(\frac{x}{y}\right) = \infty$ if $x > 0$ and $y = 0$.

[1] A finite measurable partition is a finite sequence of disjoint Borel sets $\pi := \{A_i, i = 1,2...,r\}$ whose union is $\mathscr{X}$.

The next result is closely related with some of the analysis in this dissertation. It establishes the invariance of the relative entropy with respect to bijective and bi-measurable mappings. We refer to (Dupuis and Ellis, 2011, Lemma E.2.1) for a proof of this fact based on the approximation by sums property of the relative entropy.

Lemma A.1.3. *(Dupuis and Ellis, 2011, Lemma E.2.1) Let $\mathscr{X}$ be a Polish space and ψ be a one to one mapping from $\mathscr{X}$ onto $\mathscr{X}$ such that both ψ and its inverse ψ^{-1} are measurable. Let Δ_ψ be the function mapping $\mathscr{P}(\mathscr{X})$ onto $\mathscr{P}(\mathscr{X})$ given by $\Delta_\psi \alpha = \alpha \circ \psi^{-1}$. Then for all probability measures $\mu, \nu \in \mathscr{P}(\mathscr{X})$*

$$\mathbb{D}(\Delta_\psi \mu \| \Delta_\psi \nu) = \mathbb{D}(\mu \| \nu).$$

A.2 Stochastic Processes and Relative Entropy Rate

When dealing with discrete-time stochastic processes a natural generalization of the notion of relative entropy is the so-called relative entropy rate interpretable, naïvely, as the rate of growth of the relative entropy.
Let $\{y(t);\, t \in \mathbb{Z}\}$, $\{z(t);\, t \in \mathbb{Z}\}$ be two zero-mean, jointly Gaussian, stationary, purely non deterministic processes defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $\mathbb{R}^m$. Let $\mathbf{y}_{[-n,n]}$, $\mathbf{z}_{[-n,n]}$ be the random vectors obtained by considering the restrictions of the processes y and z, respectively, to the interval $\{-n, -n+1, ..., 0, ..., n-1, n\}$. Denote by $p_{\mathbf{y}_{[-n,n]}}$ and $p_{\mathbf{z}_{[-n,n]}}$ the corresponding joint probability densities.

Definition A.2.1. *The relative entropy rate between y and z is defined as*

$$\mathbb{D}_r(y\|z) := \lim_{n\to\infty} \frac{1}{2n+1} \mathbb{D}(p_{\mathbf{y}_{[-n,n]}} \| p_{\mathbf{z}_{[-n,n]}})$$

provided that the limit exists.

The following celebrated result provides an explicit formula expressing the relative entropy rate between y and z in terms of their spectral densities. Its proof, based on the asymptotic behaviour of Toeplitz matrices and on Szegoö limit theorem, can be found for example in Stoorvogel and Van Schuppen (1996) and Ihara (1993).

Theorem A.2.2. *Let $\{y(t);\, t \in \mathbb{Z}\}$, $\{z(t);\, t \in \mathbb{Z}\}$ be two zero-mean, $\mathbb{R}^m-$valued, Gaussian, stationary, purely non deterministic processes with spectral densities Φ_y, Φ_z, respectively. Assume moreover that one of the following facts holds:*
a.) $\Phi_y \Phi_z^{-1}$ is bounded;
b.) $\Phi_y \in \mathscr{L}^2(-\pi, \pi)$ and Φ_z is coercive (i.e. $\exists \alpha > 0$ such that $\Phi(e^{i\theta}) - \alpha I_m > 0$ a.e. on $\{e^{i\theta};\, \theta \in [-\pi, \pi]\}$).

Then

$$\mathbb{D}_r(y\|z) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\det\left(\Phi_y(e^{i\theta})^{-1}\Phi_z(e^{i\theta})\right) + \mathrm{tr}\left(\Phi_z(e^{i\theta})^{-1}(\Phi_y(e^{i\theta}) - \Phi_z(e^{i\theta}))\right) d\theta.$$

Observe that the right hand side of the above equation is the celebrated Itakura-Saito divergence widely used in signal processing.

References

Adler R. J. *The geometry of random fields*. SIAM, 2010.

Agarwal A., Negahban S., and Wainwright M. J. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. *The Annals of Statistics*, pages 1171–1197, 2012.

Alpago D., Zorzi M., and Ferrante A. Identification of sparse reciprocal graphical models. *IEEE Control Systems Letters*, 2(4):659–664, 2018a.

Alpago D., Zorzi M., and Ferrante A. A scalable strategy for the identification of latent-variable graphical models. *Submitted*, 2018b.

Anderson B. and Deistler M. Identification of dynamic systems from noisy data: Single factor case. *Mathematics of Control, Signals and Systems*, 6(1):10–29, 1993.

Anderson G., Guionnet A., and Zeitouni O. *An introduction to random matrices*, volume 118. Cambridge university press, 2010.

Anderson T. and Rubin H. Statistical inference in factor analysis. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 5, pages 111–150, 1956.

Avventi E., Lindquist A., and Wahlberg B. ARMA identification of graphical models. *IEEE Transactions Automatic Control*, 58(5):1167–1178, 2013.

Bach F. R. Consistency of trace norm minimization. *Journal of Machine Learning Research*, 9(Jun):1019–1048, 2008.

Baggio G., Ferrante A., and Sepulchre R. Conal distances between rational spectral densities. *IEEE Transactions on Automatic Control*, 64(5):1848–1857, 2019.

Bai J. and Ng S. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Bekker P. A. and de Leeuw J. The rank of reduced dispersion matrices. *Psychometrika*, 52(1):125–135, 1987.

Bertsimas D., Copenhaver M. S., and Mazumder R. Certifiably optimal low rank factor analysis. *Journal of Machine Learning Research*, 18(29):1–53, 2017.

Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

Blei D. M., Kucukelbir A., and McAuliffe J. D. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

Bottegal G. and Picci G. Modeling complex systems by generalized factor analysis. *IEEE Transactions on Automatic Control*, 60(3):759–774, 2015.

Boyd S., Parikh N., Chu E., Peleato B., and Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

Boyd S. and Vandenberghe L. *Convex Optimization*. Cambridge Univ. Press, Cambridge, 2004.

Burg J. *Maximum entropy spectral analysis*. PhD thesis, Stanford University, Dept. of Geophysics, Stanford, CA, 1975.

Burt C. Experimental tests of general intelligence. *British Journal of Psychology, 1904-1920*, 3(1-2):94–177, 1909.

Byrnes C. I., Enqvist P., and Lindquist A. Identifiability and well-posedness of shaping filter parametrizations: A global analysis approach. *SIAM Journal on Control and Optimization*, 41(1):23–59, 2002.

Byrnes C. L., Georgiou T. T., and Lindquist A. A new approach to spectral estimation: a tunable high-resolution spectral estimator. *IEEE Transactions on Signal Processing*, 48(11):3189–3205, 2000.

Chamberlain G. and Rothschild M. Arbitrage, factor structure and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304, 1983.

Chandrasekaran V., Parrilo P., and Willsky A. Latent variable graphical model selection via convex optimization. *Annals of Statistics (with discussion)*, 40(4):1935–2013, 2010.

Chandrasekaran V., Sanghavi S., Parrilo P., and Willsky A. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

Chen Y., Georgiou T. T., Ning L., and Tannenbaum A. Matricial Wasserstein-1 distance. *IEEE Control Systems Letters*, 1(1):14–19, 2017.

Ciccone V. and Ferrante A. Space and spectral domain relative entropy for homogeneous random fields, submitted. 2019.

Ciccone V., Ferrante A., and Zorzi M. Factor analysis with finite data. In *56th IEEE Annual Conference on Decision and Control (CDC)*, pages 4046–4051, 2017.

Ciccone V., Ferrante A., and Zorzi M. Learning latent variable dynamic graphical models by confidence sets selection, submitted. 2018a.

Ciccone V., Ferrante A., and Zorzi M. Robust identification of "sparse plus low-rank" graphical models: An optimization approach. In *57th IEEE Conference on Decision and Control (CDC)*, pages 2241–2246, 2018b.

Ciccone V., Ferrante A., and Zorzi M. An alternating minimization algorithm for factor analysis. *Kybernetika, accepted*, 2019a.

Ciccone V., Ferrante A., and Zorzi M. Factor models with real data: a robust estimation of the number of factors. *IEEE Transactions on Automatic Control*, 64(6):2412–2425, 2019b.

Cramér H. and Leadbetter M. R. Stationary and related stochastic processes, 1967.

Dahlhaus R. Graphical interaction models for multivariate time series. *Metrika*, 51(2): 157–172, 2000.

Deistler M., Scherer W., and Anderson B. The structure of generalized linear dynamic factor models. In *Empirical Economic and Financial Research*, pages 379–400. Springer, 2015.

Deistler M. and Zinner C. Modelling high-dimensional time series by generalized linear dynamic factor models: An introductory survey. *Communications in Information & Systems*, 7(2):153–166, 2007.

Delgado R., Agüero J., and Goodwin G. A rank-constrained optimization approach: Application to factor analysis. *IFAC Proceedings Volumes*, 47(3):10373–10378, 2014.

Della Riccia G. and Shapiro A. Minimum rank and minimum trace of covariance matrices. *Psychometrika*, 47:443–448, 1982.

Dempster A. Covariance selection. *Biometrics*, 28:157–175, 1972.

Dempster A. P., Laird N. M., and Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

Doucet A., De Freitas N., and Gordon N. An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

Dupuis P. and Ellis R. S. *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons, 2011.

Ekeland I. and Temam R. *Convex analysis and variational problems*, volume 28. Siam, 1999.

Ekstrom M. P. *Digital image processing techniques*, volume 2. Academic Press, 2012.

Engle R. and Watson M. A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, 76(376):774–781, 1981.

Fan J., Liao Y., and Mincheva M. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Methodological)*, 75(4):603–680, 2013.

Fazel M. Matrix rank minimization with applications. *Elec. Eng. Dept. Stanford University*, 54:1–130, 2002.

Ferrante A., Masiero C., and Pavon M. Time and spectral domain relative entropy: A new approach to multivariate spectral estimation. *IEEE Transactions on Automatic Control*, 57(10):2561–2575, 2012.

Ferrante A., Pavon M., and Ramponi F. Hellinger versus Kullback-Leibler multivariable spectrum approximation. *IEEE Transactions on Automatic Control*, 53(4): 954–967, 2008.

Forni M. and Lippi M. The generalized dynamic factor model: representation theory. *Econometric theory*, 17(06):1113–1141, 2001.

Georgiou T. T. Relative entropy and the multivariable multidimensional moment problem. *IEEE Transactions on Information Theory*, 52(3):1052–1066, 2006.

Georgiou T. T., Karlsson J., and Takyar M. S. Metrics for power spectra: An axiomatic approach. *IEEE Transactions on Signal Processing*, 57(3):859–867, 2009.

Georgiou T. T. and Lindquist A. Kullback-Leibler approximation of spectral density functions. *IEEE Transactions on Information Theory*, 49(11):2910–2917, 2003.

Geweke J. The dynamic factor analysis of economic time series models. In *Latent Variables in Socio-Economic Models*, SSRI workshop series, pages 365–383. North-Holland, 1977.

Greenwood J. and Williamson J. P. Contact of nominally flat surfaces. *Proceedings of the royal society of London. Series A. Mathematical and physical sciences*, 295(1442): 300–319, 1966.

Guttman L. To what extent can communalities reduce rank? *Psychometrika*, 23(4): 297–308, 1958.

Ha H., Welsh J. S., Rojas C. R., and Wahlberg B. An analysis of the sparseva estimate for the finite sample data case. *Automatica*, 96:141–149, 2018.

Hakim M., Lochard E.-O., Olivier J.-P., and Terouanne E. Sur les traces de spearman (ii) le problème des rangs. *Cahiers du Bureau universitaire de recherche opérationnelle Série Recherche*, 25:23–37, 1976.

Handcock M. S. and Wallis J. R. An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, 89(426): 368–378, 1994.

Hansen L. and Sargent T. *Robustness*. Princeton University Press, 2008.

Heij C., Scherrer W., and Deistler M. System identification by dynamic factor models. *SIAM Journal on Control and Optimization*, 35(6):1924–1951, 1997.

H.Stock J. and W.Watson M. Dynamic factor analysis models. In *Oxford Handbook of Economic Forecasting*. Oxford U.P., 2010.

Hu Y. and Chou R. On the pena-box model. *Journal of Time Series Analysis*, 25(6): 811–830, 2004.

Ihara S. *Information theory for continuous systems*, volume 2. World Scientific, 1993.

Jiang X., Ning L., and Georgiou T. T. Distances and riemannian metrics for multivariate spectral densities. *IEEE Transactions on Automatic Control*, 57(7):1723–1735, 2012.

Kalman R. *Identifiability and problems of model selection in econometrics*. Cambridge University Press, 1983.

Kashyap R. L. and Lapsa P. M. Synthesis and estimation of random fields using long-correlation models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):800–809, 1984.

Kelley T. L. *Crossroads in the Mind of Man: A Study of Differentiable Mental Abilities*. Stanford University Press, 1928.

Lam C. and Yao Q. Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726, 2012.

Lang S. and McClellan J. The extension of Pisarenko's method to multiple dimensions. In *ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 7, pages 125–128. IEEE, 1982a.

Lang S. and McClellan J. Multidimensional MEM spectral estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(6):880–887, 1982b.

Lang S. and McClellan J. Spectral estimation for sensor arrays. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(2):349–358, 1983.

Lauritzen S. *Graphical Models*. Oxford University Press, Oxford, 1996.

Ledermann W. On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika*, 2(2):85–93, 1937.

Lev-Ari H., Parker S. R., and Kailath T. Multidimensional maximum-entropy covariance extension. *IEEE Transactions on Information Theory*, 35(3):497–508, 1989.

Levy B. C. and Nikoukhah R. Robust state space filtering under incremental model perturbations subject to a relative entropy tolerance. *IEEE Transactions on Automatic Control*, 58(3):682–695, 2013.

Levy B. C. and Nikoukhah R. Robust least-squares estimation with a relative entropy constraint. *IEEE Transactions on Information Theory*, 50(1):89–104, 2004.

Li S. Z. Markov random field models in computer vision. In *European conference on computer vision*, pages 361–370. Springer, 1994.

Liégeois R., Mishra B., Zorzi M., and Sepulchre R. Sparse plus low-rank autoregressive identification in neuroimaging time series. In *54th IEEE Conference on Decision and Control (CDC)*, pages 3965–3970, 2015.

Lindquist A. and Picci G. The circulant rational covariance extension problem: the complete solution. *IEEE Transactions of Automatic Control*, 58(11):2848–2861, 2013.

Lindquist A. and Picci G. *Linear Stochastic Systems*. Springer, 2015.

Longuet-Higgins M. S. On the statistical distribution of the height of sea waves. *Journal of Marine Research*, 1952.

Longuet-Higgins M. S. The statistical analysis of a random, moving surface. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 249(966):321–387, 1957.

Maanan S., Dumitrescu B., and Giurcăneanu C. D. Conditional independence graphs for multivariate autoregressive models by convex optimization: Efficient algorithms. *Signal Processing*, 133:122–134, 2017.

Maanan S., Dumitrescu B., and Giurcăneanu C. D. Maximum entropy expectation-maximization algorithm for fitting latent-variable graphical models to multivariate time series. *Entropy*, 20:76, 2018.

Mandelbrot B. B. On the geometry of homogeneous turbulence, with stress on the fractal dimension of the iso-surfaces of scalars. *Journal of Fluid Mechanics*, 72(3): 401–416, 1975a.

Mandelbrot B. B. Stochastic models for the earth's relief, the shape and the fractal dimension of the coastlines, and the number-area rule for islands. *Proceedings of the National Academy of Sciences*, 72(10):3825–3828, 1975b.

Matérn B. *Spatial variation*, volume 36. Springer Science & Business Media, 2013.

McLachlan G. and Krishnan T. The EM algorithm and extensions. Wiley series in probability and statistics, 1997.

Miller K. S. Complex random fields. *Information Sciences*, 9(3):185–225, 1975.

Ning L., Georgiou T. T., Tannenbaum A., and Boyd S. P. Linear models based on noisy data and the Frisch scheme. *SIAM Review*, 57(2):167–197, 2015.

Oudin M. and Delmas J. P. Asymptotic generalized eigenvalue distribution of block multilevel Toeplitz matrices. *IEEE Transactions on Signal Processing*, 57(1):382–387, 2008.

Pavon M. and Ferrante A. On the geometry of maximum entropy problems. *SIAM Review*, 55(3):415–439, 2013.

Pena D. and Box G. Identifying a simplifying in time series. *Journal of the American Statistical Association*, 82(399):836–843, 1987.

Picci G. Parametrization of factor analysis models. *Journal of Econometrics*, 41(1): 17–38, 1989.

Picci G. and Pinzoni S. Dynamic factor-analysis models for stationary processes. *IMA Journal of Mathematical Control and Information*, 3(2-3):185–210, 1986.

Ramponi F., Ferrante A., and Pavon M. A globally convergent matricial algorithm for multivariate spectral estimation. *IEEE Transactions on Automatic Control*, 54(10): 2376–2388, 2009.

Ringh A., Karlsson J., and Lindquist A. The multidimensional circulant rational covariance extension problem: Solutions and applications in image compression. In *54th IEEE Conference on Decision and Control (CDC)*, pages 5320–5327, 2015.

Ringh A., Karlsson J., and Lindquist A. Multidimensional rational covariance extension with applications to spectral estimation and image compression. *SIAM Journal on Control and Optimization*, 54(4):1950–1982, 2016.

Ringh A., Karlsson J., and Lindquist A. Further results on multidimensional rational covariance extension with application to texture generation. In *56th IEEE Conference on Decision and Control (CDC)*, pages 4038–4045, 2017.

Rozanov Y. A. *Stationary random processes*. Holden-Day, 1967.

Sargent T. and Sims C. Business cycle modeling without pretending to have too much a priori economic theory. Technical Report 55, Federal Reserve Bank of Minneapolis, 1977.

Saunderson J., Chandrasekaran V., Parrilo P., and Willsky A. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal Matrix Analysis Applications*, 33(4):1395–1416, 2012.

Scherrer W. and Deistler M. A structure theory for linear dynamic errors-in-variables models. *SIAM Journal on Control and Optimization*, 36(6):2148–2175, 1998.

Shapiro A. Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika*, 47(2):187–199, 1982.

Skrondal A. and Rabe-Hesketh S. Latent variable modelling: a survey. *Scandinavian Journal of Statistics*, 34(4):712–745, 2007.

Songsiri J., Dahl J., and Vandenberghe L. Graphical models of autoregressive processes. *Convex optimization in signal processing and communications*, pages 89–116, 2010.

Songsiri J. and Vandenberghe L. Topology selection in graphical models of autoregressive processes. *Journal of Machine Learning Research*, 11:2671–2705, 2010.

Spearman C. "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904.

Spearman C. and Holzinger K. J. The sampling error in the theory of two factor. *British Journal of Psychology*, 15:17–19, 1924.

Stoica P. and Moses R. *Introduction to spectral analysis*. Pearson Prentice Hall Upper Saddle River, 1997.

Stoorvogel A. A. and Van Schuppen J. H. System identification with information theoretic criteria. In **Bittanti S. and Picci G.**, editors, *Identification, Adaptation, Learning: The Science of Learning Models from Data*, pages 289–338. Springer Verlag, 1996.

Tao M. and Yuan X. Recovering low-rank and sparse components of matrices from incomplete and noisy observations. *SIAM Journal on Optimization*, 21(1):57–81, 2011.

Thurstone L. *The Vectors of the Mind*. University of Chicago Press, 1935.

Tyrtyshnikov E. E. Influence of matrix operations on the distribution of eigenvalues and singular values of Toeplitz matrices. *Linear algebra and its applications*, 207: 225–249, 1994.

Tyrtyshnikov E. E. A unifying approach to some old and new theorems on distribution and clustering. *Linear algebra and its applications*, 232:1–43, 1996.

Van Schuppen J. H. Stochastic realization problems motivated by econometric modeling. In **Byrnes C. and Lindquist A.**, editors, *Modeling Identification and Robust Control*, pages 259–275. North-Holland, 1986.

Vanmarcke E. *Random fields: analysis and synthesis*. World Scientific, 2010.

Varin C. and Vidoni P. A note on composite likelihood inference and model selection. *Biometrika*, 92(3):519–528, 2005.

Watson M. and Engle R. Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23(3): 385–400, 1983.

Whitehouse D. J. and Archard J. The properties of random surfaces of significance in their contact. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 316(1524):97–121, 1970.

Whittle P. On stationary processes in the plane. *Biometrika*, pages 434–449, 1954.

Yaglom A. M. Some classes of random fields in n-dimensional space, related to stationary random processes. *Theory of Probability & Its Applications*, 2(3):273–320, 1957.

Yaglom A. M. *An introduction to the theory of stationary random functions*. Courier Corporation, 2004.

Yaglom A. M. and others . Second-order homogeneous random fields. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory*. The Regents of the University of California, 1961.

Zhou T. and Tao D. Godec: Randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

Zhu B., Ferrante A., Karlsson J., and Zorzi M. Fusion of sensors data in automotive radar systems: A spectral estimation approach. *arXiv preprint arXiv:1908.02504*, 2019.

Zorzi M. A new family of high-resolution multivariate spectral estimators. *IEEE Transactions on Automatic Control*, 59(4):892–904, 2014.

Zorzi M. Multivariate spectral estimation based on the concept of optimal prediction. *IEEE Transactions on Automatic Control*, 60(6):1647–1652, 2015.

Zorzi M. Convergence analysis of a family of robust Kalman filters based on the contraction principle. *SIAM Journal on Control and Optimization*, 2017a.

Zorzi M. On the robustness of the Bayes and Wiener estimators under model uncertainty. *Automatica*, 83:133–140, 2017b.

Zorzi M. Robust Kalman filtering under model perturbations. *IEEE Transactions on Automatic Control*, 62(6):2902–2907, 2017c.

Zorzi M. and Levy B. C. On the convergence of a risk sensitive like filter. In *54th IEEE Conference on Decision and Control (CDC)*, pages 4990–4995, 2015.

Zorzi M. and Sepulchre R. Factor analysis of moving average processes. In *2015 IEEE European Control Conference (ECC)*, pages 3579–3584, 2015.

Zorzi M. and Sepulchre R. AR identification of latent-variable graphical models. *IEEE Transactions on Automatic Control*, 61(9):2327–2340, 2016.

Zorzi M. Empirical bayesian learning in AR graphical models. *Automatica*, 109:108516, 2019.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor Professor Augusto Ferrante for his wise guidance, his everlasting optimism and his vast and profound preparation. I am grateful for everything I have learnt in these years working with him.

I would like to thank Professor Mattia Zorzi, with whom I had the chance to closely collaborate in these years, for his healthy realism and his useful suggestions.

My sincere gratitude goes to Professor Tryphon Georgiou for his scientific curiosity and warm hospitality during my stay at the University of California Irvine.

I wish to sincerely thank Professor Yongxin Chen, Professor Michele Pavon and Professor Giorgio Picci for giving me the great opportunity to work with them.

A special thanks to my mates Alberto, Giuliano, Marco, Matteo and Michele, with whom three years ago I started this journey. Thanks also to all the past and present Ph.D. students of the Automatica group.

On a more personal side, in these few lines I would like to take the chance to apologize to my friends and dear ones for all the times I haven't been present enough in these years. In particular, my deepest gratitude goes to my parents and my brother for their unconditional support and their patience.