Università degli Studi di Padova

Dipartimento di Ingegneria dell'Informazione

Identification of Reciprocal Processes and related Matrix Extension Problem

Ph.D. candidate

Francesca Paola Carli

Advisor

Prof. Giorgio Picci

Ph.D. School in

Information Engineering

Academic Year 2010-2011

Acknowledgements

I would like to thank my advisor, prof. Giorgio Picci, for having introduced me to research, for his guidance, encouragement and constant support during my years as a Ph.D. student. I wish to express my sincere gratitude to prof. Michele Pavon and prof. Augusto Ferrante for the enlightening discussions, for their willingness and the opportunity to work with them. Also I wish to thank prof. Tryphon Georgiou for his constant interest in my research, for his creativity and contagious enthusiasm, and many advices he gave me during my stay at the University of Minnesota.

Contents

List of Figures

List of Tables

Abstract

Stationary *reciprocal processes* defined on a finite interval of the integer line can be seen as a special class of Markov random fields restricted to one dimension. This kind of processes are potentially useful for describing signals which naturally live in a finite region of the time (or space) line. Non-stationary reciprocal processes have been extensively studied in the past especially by Jamison, Krener, Levy and co-workers. The specialization of the non-stationary theory to the stationary case, however, does not seem to have been pursued in sufficient depth in the literature. Moreover, estimation and identification of reciprocal stochastic models starting from observed data seems still to be an open problem. This dissertation addresses these problems showing that maximum likelihood identification of stationary reciprocal processes on the discrete circle leads to a covariance extension problem for block-circulant covariance matrices. This generalizes the famous covariance band extension problem for stationary processes on the integer line. We show that the maximum entropy principle leads to a complete solution of the problem. An efficient algorithm for the computation of the maximum likelihood estimates is also provided.

Sommario

Un processo reciproco su un intervallo finito può essere visto come la naturale riduzione al caso unidimensionale di un campo di Markov. Questo tipo di processi è potenzialmente utile per descrivere segnali che vivono su di un intervallo spaziale o temporale limitato (si pensi ad esempio alle immagini). I processi reciproci non stazionari sono stati studiati in letteratura da B. Jamison, A. J. Krener, B. C. Levy e coautori. La specializzazione di tale teoria al caso stazionario, tuttavia, non sembra essere stata oggetto di sufficiente approfondimento in letteratura. Inoltre i problemi di stima e identificazione per processi reciproci a partire da dati osservati sono tuttora aperti. Il presente lavoro di tesi si è concentrato su tali problematiche. In particolare è stato mostrato come il problema di stima a massima verosimiglianza per processi reciproci stazionari sia riconducibile a un problema di estensione di covarianza per matrici circolanti. Tale problema generalizza il ben noto problema di estensione di covarianza per processi stazionari definiti sull'asse degli interi e non sembra essere stato affrontato in letteratura. Nel corso del lavoro di tesi è stato mostrato come tale problema sia risolubile facendo ricorso a un principio di massimizzazione dell'entropia. Infine, è stato proposto un algoritmo efficiente per il calcolo della soluzione.

Chapter 1

Introduction

Reciprocal processes have been introduced at the beginning of the last century [43, 3, 44] even earlier than the idea of Markov process was formalized by Kolmogorov and are particularly useful to describe processes indexed by space instead of time (think for example to an image). As for Markov processes, the definition of reciprocal process relies on the concept of conditionally independence. Recall that a stochastic process on a linearly ordered time interval $\mathcal{I}$ is said to be Markov if, for any $t_0 \in \mathcal{I}$, the past and the future of the process (with respect to t_0) are conditionally independent given $y(t_0)$. The same process is said to be reciprocal if, given an arbitrary interval (t_0, t_1), the random variables in the interior and exterior of this interval are conditionally independent given $y(t_0)$ and $y(t_1)$. It follows that the class of reciprocal processes is larger than the class of Markov processes: Markov processes are necessarily reciprocal [28], but the converse is not true (example of reciprocal processes that are not Markov can be found in [28], [9], [10], [31]). Moreover, the class of reciprocal processes naturally extends to the multimensional case. In fact multidimensional Markov random fields (which find applications in image processing, geophysical signal processing, oceanography, meteorology, etc.) reduce in one dimension to a reciprocal process and not to Markov process. This gives another strong motivation for the study

reciprocal processes. Reciprocal processes have been extensively studied in the past notably by Jamison, Krener, Levy and co-workers, see [28, 29, 30], [33, 32], [37], [36], [21]. However the specialization of the non-stationary theory to the stationary case, except for a few noticeable exceptions, e.g. [28], [41, 42], does not seem to have been pursued in sufficient depth in the literature. Stationary reciprocal processes can be described by constant coefficient models which are a natural generalization of the Gauss-Markov state space models widely used in engineering and applied sciences. Estimation and identification of these models starting from observed data seems to be a completely open problem and is the object of the present work. In particular, the dissertation is organized as follows.

The rest of this Chapter is devoted to some preliminaries and notations. An interesting characterization of the covariance matrix of a stationary periodic process is also provided.

Chapter 2 introduces *reciprocal processes* on the discrete circle and shows how they can be modelled by means of a double-sided "symmetric" recursion which generalizes auto-regressive (AR) processes on the integer line. A characterization of reciprocal processes in terms of the sparsity pattern of their concentration matrix is also provided. Finally, the *identification* problem is introduced and rephrased in terms of a matrix extension problem for block-circulant covariance matrices, which we shall call the *block-circulant band extension problem*. Matrix extension problems have been heavily studied in the literature (see [16], [25], [18], [24] and references therein). However, the block-circulant band extension problem seems to be a new problem which have not yet been studied.

In Chapter 3 the block-circulant band extension problemn, and hence the maximum likelihood identification problem for reciprocal processes, is solved via a *maximum entropy paradigm*. Moreover, the relationship between the block-circulant band extension problem and the Dempster's covariance

selection problem is highlighted.

In Chapter 4 some of the most popular algorithms for the solution of the covariance selection problem are reviewed and a new, efficient *algorithm for the block-circulant band extension problem* is proposed. Numerical experiments show that it compares favourably with the algorithms in the literature.

Finally, Chapter 5 summarizes the dissertation and discusses possible generalizations and future work.

1.1 Hilbert space of second order random variables

Throughout the dissertation, we work in the wide-sense setting of zero-mean random variables which have finite second moment. Random variables which have finite second moment are commonly called second order random variables. The set of real or complex-valued second-order random variables defined on the same probability space, say $\mathbf{H}$, is obviously a linear vector space under the usual operations of sum and multiplication by real (or complex) numbers. This vector space comes naturally equipped with an inner product

$$\langle \xi, \eta \rangle = \mathbb{E}\, \xi\eta$$

where $\mathbb{E}[\,\cdot\,]$ denotes the mathematical expectation (i.e. the inner product is just the correlation of the two random variables) Note that the norm induced by this inner product is positive, i.e. $\langle \xi, \xi \rangle = 0 \Leftrightarrow \xi = 0$, only if we agree to identify random variables which are equal almost surely, i.e. differ on a set of probability zero. Convergence with respect to this norm is called convergence in mean square. It is well-known that $\mathbf{H}$ is closed with respect to convergence in mean square and is therefore a *Hilbert space*. The correspondence between probabilistic concepts depending only on second

order moments and geometric operations on certain subspaces of the Hilbert space of finite variance random variables was established by Kolmogorov in the early 1940's (see e.g. [17, p. 636-637] for historical remarks on this) and will be assumed henceforth.

Following this correspondence, we say that two random vectors $\mathbf{x} = [x_1, \ldots, x_n]$ and $\mathbf{y} = [y_i, \ldots, y_n]$ are orthogonal, which we shall write $\mathbf{x} \perp \mathbf{y}$, if they are componentwise uncorrelated, i.e. if $\langle x_i, y_i \rangle = \mathbb{E}\, x_i y_i = 0$ for all $i = 1, \ldots, n$. The symbol $\hat{\mathbb{E}}\,[\,\cdot \mid \cdot\,]$ denotes orthogonal projection (conditional expectation in the Gaussian case) onto the subspace spanned by a family of finite variance random variables listed in the second argument. The concept of conditional orthogonality plays a fundamental role on the definition of reciprocal process.

Definition 1.1.1. *Let* $\mathbf{X}$, $\mathbf{Y}$ *and* $\mathbf{Z}$ *be subspaces of zero mean second order random variables in a certain common ambient Hilbert space* $\mathbf{H}$. $\mathbf{X}$ *and* $\mathbf{Y}$ *are said to be* conditionally orthogonal, given $\mathbf{Z}$*, which we shall write as*

$$\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$$

if

$$\left(\mathbf{x} - \hat{\mathbb{E}}\,[\,\mathbf{x} \mid \mathbf{Z}\,]\right) \perp \left(\mathbf{y} - \hat{\mathbb{E}}\,[\,\mathbf{y} \mid \mathbf{Z}\,]\right), \quad \forall\, \mathbf{x} \in \mathbf{X}, \forall\, \mathbf{y} \in \mathbf{Y}\,. \tag{1.1}$$

i.e., conditional orthogonality is orthogonality after subtracting the projections on $\mathbf{Z}$.

Conditional orthogonality is the same as conditional uncorrelatedness (and hence conditional independence) in the Gaussian case. The intuitive meaning of conditional orthogonality is captured by the following Lemma (see, e.g., [38]).

Lemma 1.1.1. $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$ *if and only if one of the following equivalent conditions holds*

(i) $\hat{\mathbb{E}}\,[x \mid \mathbf{Y} \vee \mathbf{Z}] = \hat{\mathbb{E}}\,(x \mid \mathbf{Z})$, $x \in \mathbf{X}$

(ii) $\hat{\mathbb{E}}\,[y \mid \mathbf{X} \vee \mathbf{Z}] = \hat{\mathbb{E}}\,(y \mid \mathbf{Z})$, $y \in \mathbf{Y}$

where $\mathbf{X} \vee \mathbf{Z}$ *(*$\mathbf{Y} \vee \mathbf{Z}$*) denote the smallest closed vector space containing* $\mathbf{X}$ *(*$\mathbf{Y}$*) and* $\mathbf{Z}$*.*

When $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ are generated by finite dimensional random vectors, condition (1.1) can equivalently be rewritten in terms of the generating vectors, which we shall normally do in the following.

In the next Section some basic facts about stationary processes on a finite interval are introduced. An interesting characterization of the covariance matrix of stationary periodic processes is also provided.

1.2 Stationary processes on a finite interval

A m-dimensional stochastic process on a finite interval $[\,1,\,N]$, is just an ordered collection of (zero-mean) random m-vectors $\mathbf{y} := \{\mathbf{y}(k),\, k = 1, 2, \ldots, N\}$ which will be written as a column vector with N, m-dimensional components. We say that $\mathbf{y}$ is *wide–sense stationary* if the covariances $\mathbb{E}\,\mathbf{y}(k)\mathbf{y}(j)^\top$ depend only on the difference of the arguments, namely

$$\mathbb{E}\,\mathbf{y}(k)\mathbf{y}(j)^\top = \Sigma_{k-j}\,, \qquad k, j = 1, \ldots, N, \tag{1.2}$$

In the following, we shall write simply "stationary", omitting the attribute "wide sense". If $\mathbf{y}$ is stationary (namely, if condition 1.2 holds), its covariance

matrix has a symmetric block-Toeplitz structure, i.e.

$$\mathbf{\Sigma}_N := \mathbb{E}\,\mathbf{y}\mathbf{y}^\top = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_{N-1}^\top \\ \Sigma_1 & \Sigma_0 & \Sigma_1^\top & \dots \\ \vdots & \ddots & \ddots & \ddots \\ \Sigma_{N-1} & \dots & \Sigma_1 & \Sigma_0 \end{bmatrix} \tag{1.3}$$

(From now on, we will use boldface capitals, e.g. $\mathbf{I}_N$, $\mathbf{\Sigma}_N$, etc. to denote block matrices made of N blocks, each of dimension $m \times m$). Processes $\mathbf{y}$ which have a positive definite covariance are called of *full rank* (or *minimal*). In this dissertation, we shall usually deal with full rank processes.

Definition 1.2.1. *A* block-circulant *matrix with N blocks, is a finite block-Toeplitz matrix whose block-columns (or equivalently, block-rows) are shifted cyclically. It looks like*

$$\mathbf{C}_N = \begin{bmatrix} C_0 & C_1 & \dots & \dots & C_{N-1} \\ C_{N-1} & C_0 & C_1 & \dots & \dots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & C_1 \\ C_1 & C_2 & \dots & C_{N-1} & C_0 \end{bmatrix}.$$

where $C_k \in \mathbb{R}^{m\times m}$. A block-circulant matrix $\mathbf{C}_N$ is fully specified by its first block-row (or column). It will be denoted by

$$\mathbf{C}_N = \mathrm{Circ}\{C_0, C_1, \dots, C_{N-1}\}. \tag{1.4}$$

For an introduction to circulant matrices see [13] and Appendix A for a generalization of significant results in [13] for *block* matrices.
Consider now a stationary process $\tilde{\mathbf{y}}$ on the integer line $\mathbb{Z}$, which is periodic of period T, i.e. a process satisfying $\tilde{\mathbf{y}}(k + nT) := \tilde{\mathbf{y}}(k)$ (almost surely) for

all $n \in \mathbb{Z}$. We can think of $\tilde{\mathbf{y}}$ as a process indexed on the *discrete circle* group, $\mathbb{Z}_T \equiv \{1, 2, \ldots, T\}$ with arithmetics mod T [1]. Clearly, its covariance function $\tilde{\Sigma}$ must also be periodic of period T, namely, $\tilde{\Sigma}_{k+T} = \tilde{\Sigma}_k$ for all $k \in \mathbb{Z}$. Hence, we may also see the covariance sequence as a function on the isomorphic discrete group $\tilde{\mathbb{Z}}_T \equiv \{\, 0, T-1 \,\}$ with arithmetics mod T. But more must be true.

Proposition 1.2.1. *A (second order) stochastic process* $\mathbf{y}$ *on* $[\,1, T\,]$ *is the restriction to the interval* $[\,1, T\,]$ *of a wide-sense stationary periodic process* $\tilde{\mathbf{y}}$ *of period* T *defined on* $\mathbb{Z}$*, if and only if its covariance matrix* $\mathbf{\Sigma}_T$ *is symmetric block-circulant.*

Proof. (only if) Let $k \in [\,1, T\,]$. By assumption there is an m-dimensional stationary process $\tilde{\mathbf{y}}$ on the integer line $\mathbb{Z}$, which is periodic of period T, satisfying $\tilde{\mathbf{y}}(k + nT) := \mathbf{y}(k)$ (almost surely) for arbitrary $n \in \mathbb{Z}$. By wide-sense stationarity, the covariance function of $\tilde{\mathbf{y}}$ must depend only on the difference of the arguments, namely

$$\tilde{\Sigma}_{k,j} := \mathbb{E}\, \tilde{\mathbf{y}}(k) \tilde{\mathbf{y}}(j)^\top = \tilde{\Sigma}_{k-j}\,, \qquad k, j = 1, \ldots, T.$$

Moreover, it is a well-known fact that, for any wide-sense stationary process the following symmetry relation holds

$$\tilde{\Sigma}_{-\tau} = \tilde{\Sigma}_\tau^\top \qquad \forall \tau \in \mathbb{Z}\,, \tag{1.5}$$

that is the covariance matrix of $\tilde{\mathbf{y}}$ has a symmetric block-Toeplitz structure. Now since $\tilde{\mathbf{y}}$ is periodic of period T, its covariance function must also be periodic of period T; i.e. $\tilde{\Sigma}_{k+nT} = \tilde{\Sigma}_k$ for arbitrary $k, n \in \mathbb{Z}$. Assume, just to fix the ideas, that T is an even number and consider the midpoint $k = \frac{T}{2}$

[1] Whence $T + \tau = \tau$ so that T plays the role of the zero element.

of the interval $[1,\, T]$. The periodicity combined with the symmetry property (1.5) yields that

$$\tilde{\Sigma}_{\frac{T}{2}+\tau} = \tilde{\Sigma}_{\frac{T}{2}+\tau-T} = \tilde{\Sigma}_{\tau-\frac{T}{2}} = \tilde{\Sigma}_{\frac{T}{2}-\tau}^{\top} \qquad \forall \tau \in \mathbb{Z} \tag{1.6}$$

and since (1.6) holds for $\tau = 0, 1, \ldots, \frac{T}{2} - 1$, we can say that the function $\tilde{\Sigma}$ must be *symmetric with respect to the midpoint* $\tau = \frac{T}{2}$ *of the interval*. Hence, we can conclude that the covariance matrix of the process $\tilde{\mathbf{y}}$ restricted to $[\,1,\, T\,]$; that is the covariance $\mathbf{\Sigma}_T$ of $\mathbf{y}$, is a symmetric block-circulant matrix, i.e. it must have the following structure

$$\mathbf{\Sigma}_T = \begin{bmatrix} \tilde{\Sigma}_0 & \tilde{\Sigma}_1^{\top} & \dots & \tilde{\Sigma}_{\tau}^{\top} & \dots & \tilde{\Sigma}_{\tau} & \dots & \tilde{\Sigma}_1 \\ \tilde{\Sigma}_1 & \tilde{\Sigma}_0 & \tilde{\Sigma}_1^{\top} & \ddots & \tilde{\Sigma}_{\tau}^{\top} & \dots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & & \tilde{\Sigma}_{\tau} \\ \tilde{\Sigma}_{\tau} & \dots & \tilde{\Sigma}_1 & \tilde{\Sigma}_0 & \tilde{\Sigma}_1^{\top} & \dots & \ddots & \\ \vdots & \tilde{\Sigma}_{\tau} & \dots & & \tilde{\Sigma}_0 & & \dots & \tilde{\Sigma}_{\tau}^{\top} \\ \tilde{\Sigma}_{\tau}^{\top} & & \ddots & & & & & \vdots \\ \vdots & \ddots & & \ddots & & \ddots & \ddots & \tilde{\Sigma}_1^{\top} \\ \tilde{\Sigma}_1^{\top} & \dots & \tilde{\Sigma}_{\tau}^{\top} & \dots & \tilde{\Sigma}_{\tau} & & \tilde{\Sigma}_1 & \tilde{\Sigma}_0 \end{bmatrix}$$

which we write

$$\mathbf{\Sigma}_T = \mathrm{Circ}\{\tilde{\Sigma}_0,\, \tilde{\Sigma}_1^{\top},\, \ldots,\, \tilde{\Sigma}_{\tau}^{\top}, \ldots,\, \tilde{\Sigma}_{\frac{T}{2}}, \ldots,\, \tilde{\Sigma}_{\tau},\, \ldots,\, \tilde{\Sigma}_1\}\,.$$

Similarly, if T is odd, it must hold that $\tilde{\Sigma}_{\frac{T+1}{2}+\tau} = \tilde{\Sigma}_{\frac{T-1}{2}-\tau}^{\top}$, $\tau = 0, 1, \ldots, \frac{T-1}{2} - 1$ and $\mathbf{\Sigma}_T$ can be written as

$$\mathbf{\Sigma}_T = \mathrm{Circ}\{\tilde{\Sigma}_0,\, \tilde{\Sigma}_1^{\top},\, \ldots,\, \tilde{\Sigma}_{\tau}^{\top}, \ldots,\, \tilde{\Sigma}_{\frac{T-1}{2}}^{\top},\, \tilde{\Sigma}_{\frac{T-1}{2}}, \ldots,\, \tilde{\Sigma}_{\tau},\, \ldots,\, \tilde{\Sigma}_1\}\,,$$

which proves the first part of the statement.

(if) We want to prove that if $\mathbf{y}$ is a process defined on a finite interval

$[1, T]$ with a symmetric block-circulant covariance matrix $\mathbf{\Sigma}_T$, then it admits a wide-sense stationary periodic extension, $\tilde{\mathbf{y}}$, defined on $\mathbb{Z}$ of period T.

Let $\tilde{\mathbf{y}}$ be the process obained by periodically extending the process $\mathbf{y}$ to the whole interger line $\mathbb{Z}$ by setting $\tilde{\mathbf{y}}(k + nT) := \mathbf{y}(k)$ for arbitrary $n \in \mathbb{Z}$ and let us denote by $\tilde{\mathbf{\Sigma}}$ its (infinite) covariance matrix. Since $\tilde{\mathbf{\Sigma}}$ is a covariance matrix, it must be positive semidefinite. What we need to show is that it is a symmetric block-Toeplitz matrix. By definition, $\tilde{\mathbf{\Sigma}}$ is the covariance matrix of the infinite column vector formed by stacking $\tilde{\mathbf{y}}(0), \tilde{\mathbf{y}}(1), \ldots, \tilde{\mathbf{y}}(T), \ldots, \tilde{\mathbf{y}}(nT), \ldots$ in that order, it is formed by subblocks which replicate $\mathbf{\Sigma}_T$ to produce a square matrix of infinite size. Since $\mathbf{\Sigma}_T$ is symmetric block-circulant, then $\tilde{\mathbf{\Sigma}}$ is, in particular, symmetric block-Toeplitz, which implies that $\tilde{\mathbf{y}}$ is stationary. This concludes the proof. □

Remark 1.2.1. The periodic extension to the whole line $\mathbb{Z}$ of deterministic signals originally given on a finite interval $[\,1,\, T]$ is a common device in (deterministic) signal processing. This simple periodic extension does however not preserve the structure of a stationary random process since the covariance of a periodically extended process will not be stationary unless the covariance function of the original process on $[\,1,\, T]$ was center-symmetric to start with. This counter-intuitive fact has to do with the quadratic dependence of the covariance of the process on its random variables.

Let for example $\mathbf{y}$ be a scalar process on the finite interval $[1, 4]$; i.e. let $T = 4$ and $m = 1$. Suppose $\mathbf{y}$ has covariance matrix $\mathbf{\Sigma}_T = \text{Toepl}\,\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$, the notation Toepl $\{a\}$ meaning that $\mathbf{\Sigma}_T$ is a symmetric Toeplitz matrix with first column given by the vector a. The upper-left $2T \times 2T$ corner of the co-

variance of the periodic extension of $\mathbf{y}$ is

$$\begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 \\ \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 & \sigma_3 & \sigma_2 & \sigma_1 & \sigma_0 \end{bmatrix}.$$

This matrix is clearly not Toeplitz unless $\sigma_3 = \sigma_1$, in which case $\mathbf{\Sigma}_T$ would be symmetric circulant. Hence the extended process $\tilde{\mathbf{y}}$ is in general not stationary.

Remark 1.2.2. In many applications to signal and image processing, the signals under study naturally live on a finite interval and modeling them as functions defined on the whole line appears just as an artifice introduced in order to use the standard tools of (causal) time-invariant systems and harmonic analysis on the line. It may indeed be more logical to describe these data as stationary processes $\mathbf{y}$ defined on a finite interval $[1, T]$. The covariance function, say $\mathbf{\Sigma}_T$, of such a process will be a symmetric positive definite block-Toeplitz matrix which has in general no block-circulant structure.
It is however always possible to extended the covariance function of $\mathbf{y}$ to a larger interval so as to make it center-symmetric. This can be achieved by simply letting $\Sigma_{T+\tau} := \Sigma_{T-1-\tau}^{\top}$ for $\tau = 0, 1, \ldots, T-1$. In this way $\mathbf{\Sigma}_T$ is extended to a symmetric block-circulant matrix $\tilde{\mathbf{\Sigma}}_T$ of dimension $(2T-1)\times(2T-1)$, but this operation does not necessarily preserve positivity. Positivity of a symmetric, block-circulant extension, however, can always be guaranteed provided the extension is done on a suitably large interval. The details on how to construct such an extension are postponed to Section 3.2,

see the proof of Theorem 3.2.2. The original process $\mathbf{y}$ can then be seen as the restriction to the interval $[1,\, T]$ of an extended process, say $\tilde{\mathbf{y}}$, which lives on an interval $[1,\, N]$ of length $N \geq 2T-1$. Since the extended covariance is, in any case, completely determined by the entries of the original covariance matrix $\mathbf{\Sigma}_T$, any statistical estimate thereof can be computed from the variables of the original process $\mathbf{y}$ in the interval $[1,\, T]$ (or from their sample values). Hence, there is no need to know what the random vectors $\{\tilde{\mathbf{y}}(k)\,;\, k = T+1, \ldots, N\}$ look like. Indeed, as soon as we are given the covariance of the process $\mathbf{y}$ defined on $[\,1,\, T]$, even if we may not ever see (sample values of) the "external" random vectors $\{\tilde{\mathbf{y}}(k)\,;\, k = T+1, \ldots, N\}$, we would in any case have a completely determined second-order description (covariance function) of $\tilde{\mathbf{y}}$.

In this sense, one can think of any stationary process $\mathbf{y}$ given on a finite interval $[1,\, T\,]$ as the restriction to $[1, T]$ of a wide-sense stationary *periodic* process, $\tilde{\mathbf{y}}$, of period $N \geq 2T-1$, defined on the whole integer line $\mathbb{Z}$. This process naturally lives on the "discrete circle" $\mathbb{Z}_N$. Hence dealing in our future study with the periodic extension $\tilde{\mathbf{y}}$, instead of the original process $\mathbf{y}$, will entail no loss of generality. □

Chapter 2

Reciprocal Processes: modeling and identification

In this Chapter, reciprocal processes on the discrete circle are introduced. In particular, it will be shown that full–rank stationary reciprocal processes can be characterized by a double-sided constant coefficients "symmetric" recursion driven by locally correlated noise whose corelation structure depends by the dynamics of the model. The identification issue will also be addressed showing that maximum likelihood identification leads to a matrix extension problem for block–circulant matrices which does not seem to have been studied in the literature.

2.1 Reciprocal Processes

In this section we define reciprocal processes on the discrete circle. The definition is given in terms of conditionally orthogonality (instead of conditionally independence) somewhat extending the common usage which deals only with Gaussian reciprocal processes. The standard definition follows immediately since, for Gaussian processes, conditional orthogonality is the same as conditional independence. Moreover, in the spirit of [21], we will consider

general reciprocal processes *of order* n, standard reciprocal processes in the literature following as a particularization for $n = 1$.

Let n be a natural number such that $N > 2n$. This inequality will be assumed to hold throughout. We introduce the notation $\mathbf{y}_{[t-n,\,t)}$ for the nm-dimensional random vector obtained by stacking $\mathbf{y}(t-\ n), \ldots, \mathbf{y}(t-1)$ in that order. Similarly, $\mathbf{y}_{(t,t+n]}$ is the vector obtained by stacking $\mathbf{y}(t+1), \ldots, \mathbf{y}(t+n)$ in that order. Likewise, the vector $\mathbf{y}_{[t-n,\,t]}$ is obtained by appending $\mathbf{y}(t)$ as last block to $\mathbf{y}_{[t-n,\,t)}$, etc.. The sums $t-k$ and $t+k$ are to be understood modulo N. Consider a subinterval $(t_1,\, t_2\,) \subset [1,\, N]$ where $(t_1,\, t_2\,) := \{t \mid t_1 < t < t_2\}$ and $(t_1,\, t_2)^c$ denotes the complementary set in $[1,\, N]$.

Definition 2.1.1. *A process* $\{\mathbf{y}(t)\}$ *on* $\mathbb{Z}_N$ *is* reciprocal of order n *if, for any interval* $(t_1,\, t_2\,) \subseteq \mathbb{Z}_N$ *the random variables in* $(t_1,\, t_2\,)$ *are conditionally orthogonal to the random variables in* $(t_1,\, t_2)^c$ *, given the* $2n$ *boundary values* $\mathbf{y}_{(t_1-n,\,t_1\,]}$ *and* $\mathbf{y}_{[t_2,\,t_2+n\,)}$*. Equivalently (see Lemma 1.1.1), it must hold that*

$$\hat{\mathbb{E}}\,[\,\mathbf{y}_{(t_1,\,t_2)} \mid \mathbf{y}(s),\, s \in (t_1,\, t_2)^c\,] = \hat{\mathbb{E}}\,[\,\mathbf{y}_{(t_1,\,t_2)} \mid \mathbf{y}_{(t_1-n,\,t_1\,]} \vee \mathbf{y}_{[t_2,\,t_2+n\,)}\,]\,, \quad (2.1)$$

for $t_1,\, t_2 \in \mathbb{Z}_N$*.*

2.1.1 AR-type modeling

In this section the modeling issue for *stationary* reciprocal processes of *order* n is addressed. These models generalize the reciprocal models *of order one* introduced in [37], discussed in [36] and, for the stationary case, especially in [41, 42].

Let $\mathbf{y}$ be a reciprocal process *of order* n on $\mathbb{Z}_N$. If $\mathbf{y}$ is reciprocal, then (2.1) holds, which, particularized to the interval $(t-1, t+1)$, yields

$$\hat{\mathbb{E}}\,[\,\mathbf{y}(t) \mid \mathbf{y}(s),\, s \neq t\,] = \hat{\mathbb{E}}\,[\,\mathbf{y}(t) \mid \mathbf{y}_{[t-n,t\,)} \vee \mathbf{y}_{(t,t+n]}]\,, \quad (2.2)$$

Let $\mathbf{d}(t)$ denote the estimation error

$$\mathbf{d}(t) := \mathbf{y}(t) - \hat{\mathbb{E}}\,[\,\mathbf{y}(t) \mid \mathbf{y}(s),\, s \neq t\,]. \tag{2.3}$$

Clearly $\mathbf{d}(t)$ is orthogonal to all the random variables $\{\mathbf{y}(s),\, s \neq t\,\}$, i.e.

$$\mathbb{E}\,\mathbf{y}\mathbf{d}^\top = \operatorname{diag}\{\Delta_0,\, \ldots,\, \Delta_{N-1}\}\ , \tag{2.4}$$

where $\mathbf{y}$ and $\mathbf{d}$ are the random vectors obtained by stacking $\{\mathbf{y}(1), \ldots, \mathbf{y}(N)\}$ and $\{\mathbf{d}(1), \ldots, \mathbf{d}(N)\}$, respectively, and Δ_t is the variance of the estimation error, $\Delta_t := \mathbb{E}\,\mathbf{d}(t+1)\mathbf{d}(t+1)^\top$. In the spirit of Masani's definition [39], $\mathbf{d}$ is called the (unnormalized) *conjugate process* (or *double-sided innovation*) of the process $\mathbf{y}$. In force of (2.2), the estimation error (2.3) becomes

$$\mathbf{d}(t) = \mathbf{y}(t) - \hat{\mathbb{E}}\,[\,\mathbf{y}(t) \mid \mathbf{y}_{[t-n,t\,)} \vee \mathbf{y}_{(t,t+n\,]}]. \tag{2.5}$$

i.e. $\mathbf{d}(t)$ is a linear combination of $\{\mathbf{y}(t-n), \ldots, \mathbf{y}(t+n)\}$. Thus, in particular, the error at $t+k$, $\mathbf{d}(t+k)$, is a linear combination of the components of the random vector $\mathbf{y}_{[t+k-n,\,t+k+n]}$. By the orthogonality property (2.4), this implies that both $\mathbf{d}(t+k)$ and $\mathbf{d}(t-k)$ are orthogonal to $\mathbf{d}(t)$ as soon as $k > n$, i.e.

$$\mathbb{E}\,\mathbf{d}(t+k)\,\mathbf{d}(t)^\top \;=\; 0 \quad \text{for} \qquad n < |k| < N-n,\; k \in \mathbb{Z}_N. \tag{2.6}$$

which will be referred saying that $\mathbf{d}$ is a locally correlated process of bandwidth n. Relation (2.5) can be seen as specifying a linear double-sided recursion for $\mathbf{y}$ of the form

$$\sum_{k=-n}^{n} F_k(t)\,\mathbf{y}(t-k) \;=\; \mathbf{d}(t)\,, \qquad t \in \mathbb{Z}_N \tag{2.7}$$

where the $F_k(t)$'s are $m \times m$ matrices, in general dependent on t, with $F_0 = I_m$ and the error process $\mathbf{d}$ satisfies the orthogonality property (2.4) and is locally correlated. From the orthogonality condition $\mathbf{d}(t) \perp \mathbf{y}_{[t-n,t)} \vee \mathbf{y}_{(t,t+n]}$, we get that the $\{F_k(t)\}$'s can be determined as the solution of the system

$$\begin{aligned} \begin{bmatrix} F_{-n}(t) & \dots & F_{-1}(t) & F_1(t) & \dots & F_n(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11}(t) & \mathbf{P}_{12}(t) \\ \mathbf{P}_{12}(t)^\top & \mathbf{P}_{22}(t) \end{bmatrix} = \\ = - \begin{bmatrix} \Sigma_n^\top & \dots \Sigma_1^\top & \Sigma_1 & \dots \Sigma_n \end{bmatrix} \end{aligned} \tag{2.8}$$

where

$$\mathbf{P}_{11} = \begin{bmatrix} \mathbb{E}\,\mathbf{y}(t+n)\mathbf{y}(t+n)^\top & \dots & \mathbb{E}\,\mathbf{y}(t+n)\mathbf{y}(t+1)^\top \\ \vdots & & \vdots \\ \mathbb{E}\,\mathbf{y}(t+1)\mathbf{y}(t+n)^\top & \dots & \mathbb{E}\,\mathbf{y}(t+1)\mathbf{y}(t+1)^\top \end{bmatrix}$$

$$\mathbf{P}_{22} = \begin{bmatrix} \mathbb{E}\,\mathbf{y}(t-1)\mathbf{y}(t-1)^\top & \dots & \mathbb{E}\,\mathbf{y}(t-1)\mathbf{y}(t-n)^\top \\ \vdots & & \vdots \\ \mathbb{E}\,\mathbf{y}(t-n)\mathbf{y}(t-1)^\top & \dots & \mathbb{E}\,\mathbf{y}(t-n)\mathbf{y}(t-n)^\top \end{bmatrix} \tag{2.9}$$

$$\mathbf{P}_{12} = \begin{bmatrix} \mathbb{E}\,\mathbf{y}(t+n)\mathbf{y}(t-1)^\top & \dots & \mathbb{E}\,\mathbf{y}(t+n)\mathbf{y}(t-n)^\top \\ \vdots & & \vdots \\ \mathbb{E}\,\mathbf{y}(t+1)\mathbf{y}(t-1)^\top & \dots & \mathbb{E}\,\mathbf{y}(t+1)\mathbf{y}(t-n)^\top \end{bmatrix}.$$

Moreover, writing (2.7) as

$$\sum_{k=-n,\, k\neq 0}^{n} F_k\, \mathbf{y}(t-k) \;=\; \mathbf{d}(t) - \mathbf{y}(t)$$

and multiplying it on the left by $(\mathbf{d}(t) - \mathbf{y}(t))^\top$, the error variance can be

expressed as

$$\mathrm{Var}\{\mathbf{d}(t)\} = \Sigma_0 - \begin{bmatrix} F_-(t) & \dots & F_+(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11}(t) & \mathbf{P}_{12}(t) \\ \mathbf{P}_{12}(t)^\top & \mathbf{P}_{22}(t) \end{bmatrix} \begin{bmatrix} F_-(t) & \dots & F_+(t) \end{bmatrix}^\top \tag{2.10}$$

where $F_-(t)$ and $F_+(t)$ are the matrices obtained by stacking $\{F_{-n}(t), \dots, F_{-1}(t)\}$ and $\{F_1(t), \dots, F_n(t)\}$, respectively, i.e.

$$F_-(t) = \begin{bmatrix} F_{-n}(t) & \dots & F_{-1}(t) \end{bmatrix}, \qquad F_+(t) = \begin{bmatrix} F_1(t) & \dots & F_n(t) \end{bmatrix}.$$

The following lemmas build on (2.8)–(2.10).

Lemma 2.1.1. *If* $\mathbf{y}$ *is stationary, the projection matrices* $\{F_k\}$*'s are independent of* t*, i.e.* $F_k(t) = F_k$ *for all* $k = -n, \dots, n$*. Moreover, if* $\mathbf{y}$ *is full rank, they are uniquely determined by the covariance lags of the process up to order* $2n$*.*

Proof. By stationarity of $\mathbf{y}$, we can drop off the dependence on t in the $\mathbf{P}_{ij}(t)$, so that system (2.8) becomes

$$\begin{aligned} \begin{bmatrix} F_{-n}(t) & \dots & F_{-1}(t) & F_1(t) & \dots & F_n(t) \end{bmatrix} \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{12}^\top & \mathbf{P}_{11} \end{bmatrix} = \\ = - \begin{bmatrix} \Sigma_n^\top & \dots \Sigma_1^\top & \Sigma_1 & \dots \Sigma_n \end{bmatrix} \end{aligned} \tag{2.11}$$

with

$$\mathbf{P}_{11} := \begin{bmatrix} \Sigma_0 & \Sigma_1 & \dots & \Sigma_{n-1} \\ \Sigma_1^\top & \Sigma_0 & \dots & \vdots \\ \vdots & & \ddots & \Sigma_1 \\ \Sigma_{n-1}^\top & \dots & \Sigma_1^\top & \Sigma_0 \end{bmatrix}, \quad \mathbf{P}_{12} := \begin{bmatrix} \Sigma_{n+1} & \Sigma_{n+2} & \dots & \Sigma_{2n} \\ \Sigma_n & \Sigma_{n+1} & \dots & \Sigma_{2n-1} \\ \vdots & & \ddots & \vdots \\ \Sigma_2 & \dots & \Sigma_n & \Sigma_{n+1} \end{bmatrix},$$

which proves the independence of the $F_k(t)$'s from the time index t. The

determinant of the coefficient matrix $\mathbf{P}$ in (2.11) is a principal minor of order $2n$ of $\mathbf{\Sigma}_N$. It follows that, if $\mathbf{y}$ is full rank, it must be nonzero. Thus $\mathbf{P}$ must be invertible and the $\{F_k\}$'s are uniquely determined. □

Lemma 2.1.2. *If* $\mathbf{y}$ *is stationary, then the error variance does not depend on* t*.*

Proof. The conclusion is straightforward by (2.10) observing that, under stationarity of $\mathbf{y}$, neither the $\{F_k\}$'s nor the P_{ij}'s depend on t. □

Let us denote with Δ the constant values taken by the error variance, i.e. $\Delta := \mathbb{E}\,\mathbf{d}(t)\mathbf{d}(t)\top$.

Remark 2.1.1. *In force of lemma 2.1.2, the orthogonality relation* (2.4) *becomes*

$$\mathbb{E}\,\mathbf{y}\mathbf{d}^\top = \operatorname{diag}\{\Delta, \ldots, \Delta\}\ . \tag{2.12}$$

Lemma 2.1.3. *The conjugate process of a stationary reciprocal process is stationary.*

Proof. Equation (2.7) can be written in matrix form as

$$\mathbf{F}_N\mathbf{y} = \mathbf{d} \tag{2.13}$$

where $\mathbf{F}_N$ is the block-banded matrix

$$\begin{bmatrix}
I & F_{-1}(1) & \dots & F_{-n}(1) & 0 & \dots & 0 & F_n(1) & \dots & F_1(1) \\
F_1(2) & I & F_{-1}(2) & \dots & F_{-n}(2) & 0 & \dots & 0 & \ddots & \vdots \\
\vdots & & \ddots & & & \ddots & \ddots & & \ddots & F_n(n) \\
F_n(n+1) & & & \ddots & & & \ddots & \ddots & & 0 \\
0 & \ddots & & & \ddots & & & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & & & \ddots & & & \ddots & 0 \\
0 & & \ddots & \ddots & & & \ddots & & & F_{-n}(N-n) \\
F_{-n}(N-n+1) & \ddots & & \ddots & \ddots & & & \ddots & & \\
\vdots & \ddots & \ddots & & \ddots & \ddots & & & \ddots & F_{-1}(N-1) \\
F_{-1}(N) & \dots & F_{-n}(N) & 0 & \dots & 0 & F_n(N) & \dots & F_1(N) & I
\end{bmatrix}$$

If $\mathbf{y}$ is stationary, we can drop off the dependence on t for the $\{F_k\}$'s (lemma 2.1.1) and $\mathbf{F}_N$ becomes block-circulant. Multiplying (2.13) from the right by $\mathbf{d}^\top$ and taking the espectation, we get $\mathbf{F}_N \mathbb{E}\left\{\mathbf{y}\mathbf{d}^\top\right\} = \mathbb{E}\left\{\mathbf{d}\mathbf{d}^\top\right\}$, which, by (2.12), yields

$$\text{Var}\left\{\mathbf{d}\right\} = \mathbf{F}_N \text{diag}\left\{\Delta, \ldots, \Delta\right\} \tag{2.14}$$

Thus Var $\{\mathbf{d}\}$ is block-circulant, being the product of block-circulant matrices, and the error process is stationary, as claimed. □

The above discussion can be summarized in the following representation theorem.

Theorem 2.1.1. *A stationary reciprocal process, $\mathbf{y}$, of order n on $\mathbb{Z}_N$ satisfies a linear, constant-coefficients difference equation of the type*

$$\sum_{k=-n}^{n} F_k\, \mathbf{y}(t-k) \;=\; \mathbf{d}(t)\,, \qquad t \in [1,\, N]\,, \tag{2.15}$$

associated to the $2n$ cyclic boundary conditions

$$\mathbf{y}(k) = \mathbf{y}(N+k)\,; \qquad k = -n+1, \ldots, n\,, \tag{2.16}$$

where the F_k's are $m \times m$ matrices, with $F_0 = I_m$ and the error process $\mathbf{d}$, besides satisfying the orthogonality property (2.12), is locally correlated. The model can be rewritten in matrix form as

$$\mathbf{F}_N\, \mathbf{y} = \mathbf{d}\,. \tag{2.17}$$

where $\mathbf{F}_N$ is the N-block banded circulant matrix of bandwidth n,

$$\mathbf{F}_N := \text{Circ}\{I,\, F_{-1},\, \ldots,\, F_{-n},\, 0,\, \ldots 0,\, F_n,\, \ldots,\, F_1\}\,. \tag{2.18}$$

If the process is full rank this description is unique.

Up to this point we have derived a system of equations satisfied by $\mathbf{y}$. A natural question to be answered is whether the system (2.15) is well posed, i.e. whether it determines $\mathbf{y}$ uniquely. From (2.17), we see that the well-posedness of (2.15) is equivalent to the invertibility of $\mathbf{F}_N$. An obvious sufficient condition for the model (2.15) to be well posed is the following.

Lemma 2.1.4. *If* $\mathbf{y}$ *is full rank, the model* (2.15) *is well posed.*

From now on we shall focus on full rank reciprocal processes, since for this class we know that the model is well-posed. The following result is useful in connection with realization theory.

Proposition 2.1.1. *A stationary reciprocal process* $\mathbf{y}$ *is full rank if and only if the variance matrix* Δ *of the conjugate process is positive definite.*

Proof. (if) Suppose $\Delta > 0$. Multiplying both members of (2.13) from the right by $\mathbf{y}^\top$ and taking expectations, in virtue of the orthogonality relation (2.12), we get

$$\mathbf{F}_N \, \mathbf{\Sigma}_N = \mathbf{F}_N \, \mathbb{E}\, \mathbf{y}\mathbf{y}^\top = \mathbb{E}\, \mathbf{d}\mathbf{y}^\top = \operatorname{diag}\{\Delta, \dots, \Delta\}. \tag{2.19}$$

Thus $\Delta > 0$ implies that the square matrices $\mathbf{F}_N$ and $\mathbf{\Sigma}_N$ are invertible which, combined with the positive semidefiniteness of $\mathbf{\Sigma}_N$, implies $\mathbf{\Sigma}_N > 0$.

(only if) Suppose now that Δ is only positive semidefinite. This implies that there exists $0 \neq a \in \mathbb{R}^m$ s.t. $\mathbb{E}\, a^\top \mathbf{d}(t)\mathbf{d}(t)^\top a = 0$, i.e. s.t. $a^\top \mathbf{d}(t) = 0$ a.s.. This means that the scalar components of $\mathbf{d}(t)$ are linearly dependent, which, by (2.7), implies that $\mathbf{y}(t-n), \dots, \mathbf{y}(t), \dots, \mathbf{y}(t+n)$ are linearly dependent. Thus $\mathbf{\Sigma}_N$ must be singular, which contradicts the assumption $\mathbf{\Sigma}_N > 0$. □

Remark 2.1.2. *Solving* (2.19) *we get*

$$\mathbf{\Sigma}_N^{-1} = \operatorname{diag}\{\Delta^{-1}, \dots, \Delta^{-1}\}\mathbf{F}_N \tag{2.20}$$

i.e. the inverse of the covariance matrix of a full rank stationary reciprocal process of order n is a banded block-circulant matrix of bandwidth n. As we will see in a few moments, the converse is also true, providing a useful characterization of full rank stationary recirprocal processes.

In the above discussion, we have shown that if a process is reciprocal, it admits an autoregressive model of the form (2.15) associated to the cyclic boundary conditions (2.16), with noise structure (2.6). However we have not yet shown that this model captures completely the structure of a reciprocal processes, i.e. we have not shown that the solution of such a model is necessarily reciprocal. It turns out that this is the case, but instead of considering directly the model (2.15), we introduce a renormalized version of this model which is simpler to analyze.

Multiplying (2.15) by the right for Δ^{-1} and letting $M_k := \Delta^{-1}F_k$, $k = -n, \dots, n$ (and thus, in particular, $M_0 = \Delta^{-1}$), we get

$$\sum_{k=-n}^{n} M_k \, \mathbf{y}(t-k) \;=\; \mathbf{e}(t)\,, \qquad t \in \mathbb{Z}_N \tag{2.21}$$

where $\mathbf{e}(t)$ is the normalized the conjugate process

$$\mathbf{e}(t) := \Delta^{-1}\mathbf{d}(t) \tag{2.22}$$

so that $\mathrm{Var}\{\mathbf{e}(t)\} = \Delta^{-1}$. Equivalently, one can multiply on the left by $\mathrm{diag}\{\Delta^{-1}, \dots, \Delta^{-1}\}$ equation (2.17) to get

$$\mathbf{M}_N \mathbf{y} = \mathbf{e} \tag{2.23}$$

where we have set $\mathbf{e} = \mathrm{diag}\{\Delta^{-1}, \dots, \Delta^{-1}\}\,\mathbf{d}$ and

$$\mathbf{M}_N = \mathrm{diag}\left\{\Delta^{-1}, \dots, \Delta^{-1}\right\}\mathbf{F}_N. \tag{2.24}$$

The information contained in (2.24) is twofold: comparing (2.24) with (2.19) we get that the inverse of the covariance matrix of the reciprocal process $\mathbf{y}$ is in fact the coefficient matrix of the normalized model

$$\mathbf{\Sigma}_N^{-1} = \mathbf{M}_N \,, \tag{2.25}$$

from which we conclude that the (matricial) coefficients of the normalized model must form a *center-symmetric sequence* of bandwidth n

$$M_{-k} = M_k^\top \,, \qquad k = 1, \ldots, n \,, \tag{2.26}$$

i.e. the model (2.7) is *self-adjoint*).

For this symmetrized model, the orthogonality relation (2.12) is replaced by

$$\mathbb{E}\, \mathbf{y}\, \mathbf{e}^\top \,=\, \mathbf{I}_N \,. \tag{2.27}$$

Moreover, the coefficient matrix of the normalized model $\mathbf{M}_N$, besides being the inverse of the covariance matrix of the reciprocal process $\mathbf{y}$, is the covariance matrix of the normalized conjugate process $\mathbf{e}$. In fact, multiplying (2.23) from the right by $\mathbf{e}^\top$ and taking expectations, we get $\mathbf{M}_N\, \mathbb{E}\, \{\mathbf{y}\mathbf{e}^\top\} = \mathbb{E}\, \{\mathbf{e}\mathbf{e}^\top\}$ which, in force of (2.27), yields

$$\mathrm{Var}\, \{\mathbf{e}\} = \mathbf{M}_N \tag{2.28}$$

as announced.

We are now ready to show that model (2.21) (equivalently, (2.23)) captures completely the structure of a reciprocal processes, i.e. that, under the condition that $\mathbf{e}$ is stationary and locally correlate process, the solution of the system $\mathbf{M}_N \mathbf{y} = \mathbf{e}$, is necessarily reciprocal.

Theorem 2.1.2. *Consider the linear system*

$$\mathbf{M}_N \mathbf{y} = \mathbf{e}\,, \tag{2.29}$$

where $\mathbf{M}_N$ *is a symmetric positive-definite banded block-circulant matrix of bandwidth* n *and the process* $\{\mathbf{e}(t)\,;\ t \in \mathbb{Z}_N\}$ *is stationary with covariance matrix* $\mathbf{M}_N$ *(i.e. it is locally correlated of bandwidth* n*). Then there is a unique full rank stationary reciprocal process* $\mathbf{y}$ *of order* n *solution of* (2.29)*. This process satisfies the orthogonality condition* (2.27) *and* $\mathbf{e}$ *is its normalized conjugate process.*

Proof. Pick a finitely correlated process $\mathbf{e}$ with covariance matrix $\mathbf{M}_N$ (we can construct such a, say Gaussian, process on a suitable probability space) and let $\mathbf{y}$ be a solution of (2.29). Since $\mathbf{M}_N$ is invertible, $\mathbf{y}$ is uniquely defined, i.e. there is a unique random vector $\mathbf{y}$, solution of (2.29). Let $\mathbf{\Sigma}_N$ be its covariance matrix. We have

$$\mathbf{\Sigma}_N := \mathbb{E}\left[\mathbf{y}\mathbf{y}^\top\right] = \mathbb{E}\left[\mathbf{M}_N^{-1}\mathbf{e}\mathbf{e}^\top\mathbf{M}_N^{-\top}\right] = \mathbf{M}_N^{-1}\,,$$

i.e. $\mathbf{\Sigma}_N$ is a symmetric positive definite block-circulant matrix, which proves that $\mathbf{y}$ is stationary and full rank on $\mathbb{Z}_N$ (Proposition 1.2.1).

Moreover, by multiplying (2.29) by $\mathbf{e}^\top$ and taking expectations, we find $\mathbf{M}_N\mathbb{E}\{\mathbf{y}\mathbf{e}^\top\} = \mathbf{M}_N$, so that $\mathbb{E}\{\mathbf{y}\mathbf{e}^\top\} = \mathbf{I}_N$, or equivalently $\mathbb{E}\{\mathbf{y}(t)\mathbf{e}(s)^\top\} = \mathbf{I}_m\,\delta_{ts}$. Therefore, the orthogonality (2.27) holds on $\mathbb{Z}_N$.

Next, we need to show that $\mathbf{y}$ is reciprocal of order n. To this end we shall generalize an argument of [42]. Let $s < t$ be two points in $[1,\, N]$, which for the moment we choose such that $t - n > s + n$, which is always possible since by assumption $N > 2n$. The situation is shown in figure 2.1. Expanding (2.21) and rearranging terms, we get equation (2.28), displayed

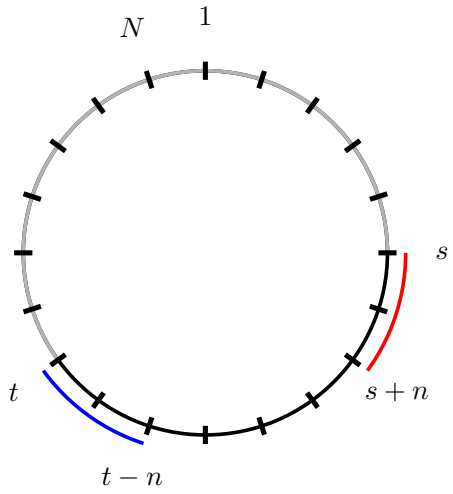


Figure 2.1: Non overlapping intervals $[t-n,\, t)$, $(\, s,\, s+n]$.

at the top of the next page, which can be compactly rewritten as

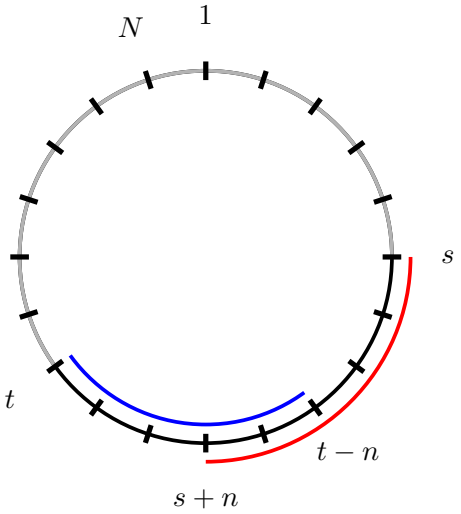
$$\tilde{\mathbf{M}}\, \mathbf{y}_{[t,\, s\,]} = \mathbf{e}_{[t,\, s\,]} - \begin{bmatrix} \mathbf{N} & 0 \\ 0 & 0 \\ 0 & \mathbf{N}^\top \end{bmatrix} \begin{bmatrix} \mathbf{y}_{[t-n,\, t)} \\ \mathbf{y}_{(\, s,\, s+n]} \end{bmatrix} \tag{2.30}$$

with an obvious meaning of the symbols. Note that $\tilde{\mathbf{M}}$ is non-singular, its determinant being a principal minor of $\mathbf{M}_N$, which is positive definite by assumption. Thus we can solve (2.30) and express $\mathbf{y}_{[t,\, s\,]}$ as a sum of two linear functions of $\mathbf{e}_{[t,\, s\,]}$ and of $\mathbf{y}_{(\, s,\, s+n]} \vee \mathbf{y}_{[t-n,\, t)}$. The two random vectors on the right hand side of (2.30) are uncorrelated, since all scalar components of $\mathbf{e}_{[t,\, s\,]}$ are orthogonal to the linear subspace spanned by (the scalar components of) $\{\mathbf{y}(\tau)\,;\, \tau \in [t,\, s\,]^c\}$ and thus, in particular, are orthogonal to the boundary condition vectors $\mathbf{y}_{(\, s,\, s+n]}$, $\mathbf{y}_{[t-n,\, t)}$. It follows that the orthogonal projection of $\mathbf{y}_{[t,\, s\,]}$ onto the linear subspace spanned by (the scalar components of) $\{\mathbf{y}(\tau)\,;\, \tau \in [t,\, s\,]^c\}$ results in a linear function of (the scalar components of) $\mathbf{y}_{[t-n,\, t)} \vee \mathbf{y}_{(\, s,\, s+n]}$ alone, which proves the conditional orthogonality of $\mathbf{y}_{[t,\, s\,]}$ to $\mathbf{y}_{[t,\, s\,]^c}$, given the boundary values $\mathbf{y}_{[t-n,\, t)}$, $\mathbf{y}_{(\, s,\, s+n]}$.

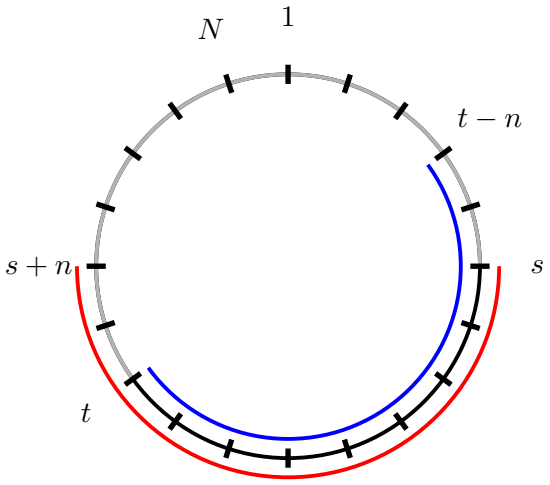
$$
\begin{bmatrix}
M_0 & M_1^\top & \dots & M_n^\top & 0 & \dots & 0 & & & 0 \\
M_1 & M_0 & M_1^\top & \ddots & M_n^\top & 0 & & & & 0 \\
\vdots & & \ddots & & & \ddots & & & & \vdots \\
M_n & \dots & M_1 & M_0 & M_1^\top & \dots & M_n^\top & 0 & & 0 \\
0 & M_n & & \dots & M_0 & \dots & & \ddots & & \\
\vdots & & & \dots & & \ddots & & & & 0 \\
0 & & & \dots & & \dots & & & & M_n^\top \\
 & & & & & & & & & \vdots \\
0 & & & \ddots & & \ddots & & M_1 & M_0 & M_1^\top \\
0 & & & 0 & \dots & 0 & M_n & \dots & M_1 & M_0
\end{bmatrix}
\begin{bmatrix}
\mathbf{y}(t) \\ \mathbf{y}(t+1) \\ \vdots \\ \mathbf{y}(t+n) \\ \vdots \\ \mathbf{y}(s-n) \\ \vdots \\ \mathbf{y}(s-1) \\ \mathbf{y}(s)
\end{bmatrix} =
$$

$$
\begin{bmatrix}
\mathbf{e}(s) \\ \mathbf{e}(s+1) \\ \vdots \\ \mathbf{e}(s+n) \\ \vdots \\ \mathbf{e}(t-n) \\ \vdots \\ \mathbf{e}(t-1) \\ \mathbf{e}(t)
\end{bmatrix} -
\begin{bmatrix}
M_n & \dots & & M_1 & 0 & \dots & 0 \\
0 & M_n & \dots & M_2 & 0 & \dots & 0 \\
0 & \vdots & \ddots & & 0 & \dots & 0 \\
0 & & \ddots & M_n & 0 & \dots & 0 \\
0 & \dots & 0 & \dots & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \dots & 0 & \dots & 0 & \dots & 0 \\
0 & \dots & 0 & & M_n^\top & 0 & \\
0 & \dots & 0 & \dots & \vdots & & 0 \\
0 & \dots & 0 & & M_2^\top & \ddots & 0 \\
0 & \dots & 0 & & M_1^\top & \dots & M_n^\top
\end{bmatrix}
\begin{bmatrix}
\mathbf{y}(t-n) \\ \mathbf{y}(t-n+1) \\ \vdots \\ \mathbf{y}(t-1) \\ \mathbf{y}(s+1) \\ \vdots \\ \mathbf{y}(s+n-1) \\ \mathbf{y}(s+n)
\end{bmatrix}
\tag{2.28}
$$

The argument remains valid also when the non overlapping condition $t-n > s+n$ does not hold, i.e. for an arbitrary interval $[t,\, s\,]$ of the discrete circle $\mathbb{Z}_N$ (see figures 2.2a and 2.2b). In fact, when $[t-n,\, t)$ and $(\, s,\, s+n]$ overlap, clearly we have $[t,\, s\,]^c \subseteq [t-n,\, t) \cup (\, s,\, s+n]$ and hence all random variables in the subspace spanned by $\{\mathbf{y}(\tau)\,;\, \tau \in [t,\, s\,]^c\}$ are contained in the subspace spanned by the boundary conditions, say $\mathcal{C} := \{\mathbf{y}(\tau)\,;\, \tau \in [t-n,\, t) \cup (\, s,\, s+$



(a) $s+n<t$ and $t-n>s$



(b) $s+n>t$ and $t-n<s$

Figure 2.2: Overlapping intervals $[t-n,\, t)$ and $(\, s,\, s+n]$.

$n]\}$. This means that $\hat{\mathbb{E}}\,[\,\mathbf{y}(\tau) \mid \mathcal{C}\,] = \mathbf{y}(\tau)$, or equivalently that

$$\mathbf{y}(\tau) - \hat{\mathbb{E}}\,[\,\mathbf{y}(\tau) \mid \mathcal{C}\,] = 0\,, \qquad \tau \in [t,\, s\,]^c$$

i.e. the second member in (1.1) is zero and the orthogonality condition trivially holds. □

From this result, we obtain the following fundamental characterization of reciprocal processes on the discrete group $\mathbb{Z}_N$.

Theorem 2.1.3. *A nonsingular $mN \times mN$-dimensional matrix $\mathbf{\Sigma}_N$ is the covariance matrix of a reciprocal process of order n on the discrete group $\mathbb{Z}_N$ if and only if its inverse is a positive-definite symmetric block-circulant matrix banded of bandwidth n.*

We conclude this section with a couple of remarks.

Remark 2.1.3. The second order statistics of both $\mathbf{y}$ and $\mathbf{e}$ are encapsulated in the covariance matrix $\mathbf{M}_N$ (see (2.25) and (2.28)), which means that the whole auto-regressive model of $\mathbf{y}$ is defined in terms of the matrix $\mathbf{M}_N$. This result makes the stochastic realization problem for reciprocal processes of order n conceptually trivial. In fact, given the covariance matrix $\mathbf{\Sigma}_N$ (the external description of the process), assuming that it is in fact the covariance matrix of such a process, the model matrix $\mathbf{M}_N$ can be computed by simply inverting $\mathbf{\Sigma}_N$. This is the simplest answer one could hope for. The solution requires however a preliminary criterion to check whether a (full rank) symmetric block-circulant covariance matrix has a banded inverse. There seems to be no simple known answer to this question.

Remark 2.1.4. To make contact with the literature, we note that a full rank reciprocal process of order n can always be represented as a linear memoryless function of a reciprocal process of order 1. This reciprocal process, however,

need not be of full rank. To see that this is the case, introduce the vectors

$$\mathbf{y}_t^+ := \begin{bmatrix} \mathbf{y}(t) \\ \vdots \\ \mathbf{y}(t+n-1) \end{bmatrix}, \quad \mathbf{y}_t^- := \begin{bmatrix} \mathbf{y}(t-n+1) \\ \vdots \\ \mathbf{y}(t) \end{bmatrix}. \tag{2.29}$$

Letting $\mathbf{x}(t)^\top := \begin{bmatrix} (\mathbf{y}_t^-)^\top & (\mathbf{y}_t^+)^\top \end{bmatrix}$, we find the representation

$$\mathbf{x}(t) = \begin{bmatrix} A_+ & 0 \\ 0 & 0 \end{bmatrix} \mathbf{x}(t-1) + \begin{bmatrix} 0 & 0 \\ 0 & A_- \end{bmatrix} \mathbf{x}(t+1) + \tilde{\mathbf{d}}(t) \tag{2.30}$$

$$\mathbf{y}(t) = \begin{bmatrix} 0 & \dots & 0 & 1 & 1 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}(t) \tag{2.31}$$

where A_- and A_+ are the block-companion matrices

$$A_+ := \begin{bmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ & \dots & & & I \\ -F_n & \dots & & & -F_1 \end{bmatrix}$$

$$A_- := \begin{bmatrix} -F_{-1} & \dots & & & -F_{-n} \\ I & 0 & \dots & & 0 \\ 0 & I & 0 & \dots & 0 \\ & \dots & & I & 0 \end{bmatrix}$$

and $\tilde{\mathbf{d}}(t) = \frac{1}{2} \begin{bmatrix} 0 & \dots & 0 & \mathbf{d}(t)^\top & \mathbf{d}(t)^\top & 0 & \dots & 0 \end{bmatrix}^\top$ has a singular covariance matrix. This model is in general non-minimal [42].

2.2 Identification

Assume that T independent realizations of one period of the process $\mathbf{y}$ are available and let us denote by

$$\underline{y} := \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(T)} \end{bmatrix}$$

the collection of these T realizations. To be more precise, $y^{(k)}$ is the $N \cdot m$–dimensional column vector obtained by stacking the N, m-dimensional vectors which build up the k–th realization, while $\underline{y}$ is the $N \cdot m \cdot T$–dimensional column vector obtained by stacking the T realizations one after the other. (Think for example to a "movie" consisting of T close frame of the same image, each frame built up of $m \times N$ pixels: each frame is a realization of the "image" process (the process of the columns of the image), made up of N, m–dimensional samples (the columns of the image) which can be stored in a $N \cdot m$-dimensional vector $y^{(k)}$; finally, the samples building up the T realizations can be collected in the vector $\underline{y}$ as described above).

The problem we are interested in solving is the following.

Problem 2.2.1 (Identification Problem)**.** *Given the observations $\underline{y}$ of a reciprocal process $\mathbf{y}$ of (known) order n, estimate the parameters $\{M_k\}$ of the underlying reciprocal model $\mathbf{M}_N\mathbf{y} = \mathbf{e}$.*

Remark 2.2.1. If we are given $2n+1$ covariance data $\{\Sigma_k\,;\, k = 0, 1, \ldots, 2n\}$, the identification of an order n reciprocal process can be carried out by a linear algorithm, namely by solving the Yule-Walker-type system of linear equations (2.11). This procedure is however unsatisfactory since, due to the symmetry (2.26), there are actually only $n+1$ unknown M_k to be computed. Hence, one would expect only $n+1$ covariance lags to be needed, while the system (2.11) requires solving also for the negative order coefficients.

Moreover, in practice, the Σ_k's will have to be estimated from observed data and estimates of covariances with a large lag k will unavoidably be more uncertain and have a larger variance.

In an attempt to get asymptotically efficient estimates for the M_k's, we consider maximum likelihood estimation subject to the constraint that the inverse of the covariance matrix $\mathbf{M}_N$ has to be block-circulant and banded. To this end, we set up a Gaussian likelihood function (which does not require to assume that $\mathbf{y}$ has a Gaussian distribution, see [27, p. 112]), which uses the density function

$$p_{(M_0,\dots,M_n)}(y) = \frac{1}{\sqrt{(2\pi)^{mN}\det\left(\mathbf{M}_N^{-1}\right)}}\exp\left(-\frac{1}{2}\left(y^{(k)}\right)^\top \mathbf{M}_N\, y^{(k)}\right),$$

where $y^{(k)} \in \mathbb{R}^{mN}$. Assuming that the T sample measurements are independent and neglecting terms which do not depend on the parameters, the log-likelihood function is

$$\begin{aligned} L(M_0,\dots,M_n) &= -\frac{T}{2}\log\det\mathbf{M}_N^{-1} - \frac{1}{2}\sum_{k=1}^{T}\left(y^{(k)}\right)^\top \mathbf{M}_N\, y^{(k)} \\ &= -\frac{T}{2}\log\det\mathbf{M}_N^{-1} - \frac{1}{2}\mathrm{tr}(\mathbf{M}_N \underline{y}\,\underline{y}^\top) \\ &= -\frac{T}{2}\left[\log\det\mathbf{M}_N^{-1} + \mathrm{tr}\left(\mathbf{M}_N\bar{\Sigma}_N\right)\right] \\ &= -\frac{T}{2}\left[\log\,\det\mathbf{M}_N^{-1} + \sum_{k=0}^{n}\mathrm{tr}\left\{M_k\, T_k\left(\underline{y}\right)\right\}\right] \end{aligned} \tag{2.32}$$

where $\bar{\Sigma}_N$ is the sample covariance $\bar{\Sigma}_N = \frac{1}{T}\underline{y}\,\underline{y}^\top$ and each matrix-valued statistic $T_k(\underline{y})$ has the structure of a sample estimate of the lag k covariance

of the process. For example, T_0 and T_1 are given by:

$$\begin{aligned} T_0\left(\underline{y}\right) &= \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{k=0}^{N-1} y^{(t)}(k)\left[y^{(t)}(k)\right]^\top\right\} \\ T_1\left(\underline{y}\right) &= \frac{2}{T}\sum_{t=1}^{T}\left\{\sum_{k=1}^{N-1} y^{(t)}(k-1)\left[y^{(t)}(k)\right]^\top\right\} \\ &+ \frac{2}{T}\sum_{t=1}^{T} y^{(t)}(N-1)\left[y^{(t)}(0)\right]^\top \end{aligned}$$

From exponential class theory [1], we see that the T_k's are (matrix-valued) sufficient statistics. Indeed, we have the well-known characterization that the (suitably normalized) statistics $T_0, T_1, \ldots, T_n$ are maximum likelihood estimators of their expected values, namely

$$\begin{aligned} \hat{\Sigma}_0 &:= \frac{1}{N}T_0 = \text{ M.L. Estimator of } \mathbb{E}\,\mathbf{y}(k)\mathbf{y}(k)^\top \\ &\vdots \\ \hat{\Sigma}_n &:= \frac{1}{N}T_n = \text{ M.L. Estimator of } \mathbb{E}\,\mathbf{y}(k+n)\mathbf{y}(k)^\top . \end{aligned} \tag{2.33}$$

Let us now consider the following matrix completion problem, which, from now on, will be referred to as the *block-circulant band extension problem.*

Problem 2.2.2 (**Block-Circulant Band Extension Problem**). *Given* $n+1$ *initial data* $m \times m$ *matrices* $\hat{\Sigma}_0, \ldots, \hat{\Sigma}_n$*, arranged in a way consistent with a symmetric block circulant structure, i.e. given the partially specified*

block–circulant matrix

$$\begin{bmatrix}
\Sigma_0 & \Sigma_1^\top & \dots & \Sigma_n^\top & ? & \dots & ? & \Sigma_n & \dots & \Sigma_1 \\
\Sigma_1 & \Sigma_0 & \Sigma_1^\top & \ddots & \Sigma_n^\top & ? & & ? & \ddots & \vdots \\
\vdots & & \ddots & & & \ddots & \ddots & & \ddots & \Sigma_n \\
\Sigma_n & \dots & \Sigma_1 & \Sigma_0 & \Sigma_1^\top & \dots & \Sigma_n^\top & \ddots & & ? \\
? & \Sigma_n & & \dots & \Sigma_0 & \dots & & \Sigma_n^\top & \ddots & \vdots \\
\vdots & \ddots & \ddots & & & \ddots & & & \ddots & ? \\
? & & & & & & & & & \Sigma_n^\top \\
\Sigma_n^\top & \ddots & & & & & & & & \vdots \\
\vdots & \ddots & \ddots & & \ddots & \ddots & & & \ddots & \Sigma_1^\top \\
\Sigma_1^\top & \dots & \Sigma_n^\top & ? & \dots & ? & \Sigma_n & \dots & \Sigma_1 & \Sigma_0
\end{bmatrix}$$

complete it in such a way to form a positive definite symmetric block-circulant matrix $\mathbf{\Sigma}_N$ *with a (block-circulant)* banded inverse of bandwidth n.

Note that the model parameters $(M_0,\, M_1, \dots, M_n)$ are the nonzero blocks of the (banded) inverse of the covariance matrix $\mathbf{\Sigma}_N$ of the process (Theorem 2.1.3). The invariance principle for maximum likelihood estimators [51] leads then to the following statement.

Theorem 2.2.1. *The maximum likelihood estimates of* $(M_0,\, M_1, \dots, M_n)$ *are the nonzero blocks of the banded inverse of the matrix* $\hat{\mathbf{\Sigma}}_N$ *solving the block-circulant band extension problem 2.2.2 with initial data the* $n+1$ *covariance estimates* (2.33).

For the sake of clarity, let us introduce the following notation. Let $\mathcal{I}_b$ be the sets of pairs of block indices consistent with a banded-symmetric circulant structure of bandwidth n, i.e.

$$\begin{aligned}
&|i-j| \le n \Rightarrow (i,j) \in \mathcal{I}_b \\
&(i,j) \in \mathcal{I}_b \Rightarrow (j,i) \in \mathcal{I}_b \\
&(i,j) \in \mathcal{I}_b \Rightarrow \Big((j+1)_{\text{mod N}}, (i+1)_{\text{mod N}} \Big) \in \mathcal{I}_b.
\end{aligned}$$

(the set is represented in figure 2.3a) and $\mathcal{R}_b$ the corresponding set of blocks consistent with a banded-symmetric block-circulant structure, i.e.

$$\mathcal{R}_b := \left\{ r_{ij} \in \mathbb{R}^{m\times m} \mid (i,j) \in \mathcal{I}_b, r_{ij} = r_{ji}^\top, r_{ij} = r_{(j+1)_{\text{mod N}}(i+1)_{\text{mod N}}}^\top \right\}.$$

Morover, let us denote by e_k the $m \times mN$ block-matrix with the identity in the k-th block position,

$$e_k = \begin{bmatrix} 0 & \dots & 0 & I_m & 0 & \dots & 0 \end{bmatrix},$$

by $\mathfrak{S}_N$ the vector space of *symmetric* matrices with $N \times N$ square blocks of dimension $m \times m$, and by $\mathfrak{C}_N$ the set of $N \times N$ m-block-circulant matrices. In the above notation, the ML estimation problem for the M_k's reads as follows:

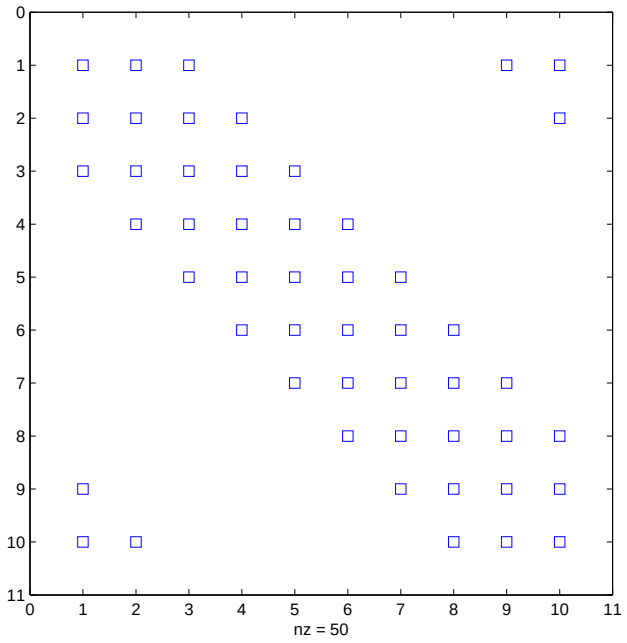
Problem 2.2.3 (**ML estimation problem**).

$$\max \left\{ -\log\det \mathbf{M}_N^{-1} - \operatorname{tr}\left(\mathbf{M}_N \bar{\Sigma}_N\right) \mid \mathbf{M}_N \in \mathfrak{S}_N,\ \mathbf{M}_N > 0 \right\} \tag{2.34a}$$

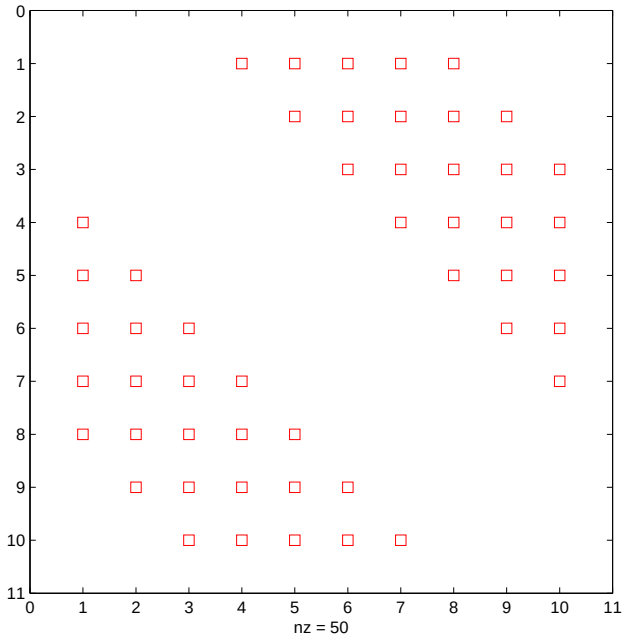
subject to :

$$e_i^\top \mathbf{M}_N e_j = 0,\ \textit{for}\ (i,j) \in \mathcal{I}_b^c. \tag{2.34b}$$

with $\mathcal{I}_b^c$ complement of $\mathcal{I}_b$ in $[1,\dots,N] \times [1,\dots,N]$ (see figure 2.3b), while the block-circulant band extension problem (2.2.2) becomes



(a) Block index set $\mathcal{I}_b$



(b) Block index set $\mathcal{I}_b^c$

Figure 2.3: Block index sets $\mathcal{I}_b$ and $\mathcal{I}_b^c$.

Problem 2.2.4 (**Block-circulant band extension problem**).

$$\text{Find } \mathbf{\Sigma}_N \in \mathfrak{S}_N \text{ such that} \tag{2.35a}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \text{ for } (i,j) \in \mathcal{I}_b \text{ and } r_{ij} \in \mathcal{R}_b, \tag{2.35b}$$

$$\mathbf{\Sigma}_N > 0, \tag{2.35c}$$

$$e_i^\top \mathbf{\Sigma}_N^{-1} e_j = 0, \text{ for } (i,j) \notin \mathcal{I}_b, \tag{2.35d}$$

$$\mathbf{\Sigma}_N \in \mathfrak{C}_N. \tag{2.35e}$$

Up to now, we have shown that solving the ML estimation problem 2.2.3 is equivalent to solve the block-circulant band extension problem 2.2.4. Note, however, that the block-circulant band extension problem 2.2.2 is nonlinear and it is hard to see what is going on by elementary means. Below we give a scalar example.

Example 2.2.1. *Let $m = 1$, $N = 8$, $n = 2$ and assume we are given the covariance estimates $\hat{\sigma}_0$, $\hat{\sigma}_1$, $\hat{\sigma}_2$, forming a positive definite Toeplitz matrix. The three unknown coefficients in the reciprocal model* (2.21) *of order* 2 *are scalars, denoted m_0, m_1, m_2. Multiplying* (2.23) *from the right by $\mathbf{y}^\top$, we get $\mathbf{M}_N \mathbf{\Sigma}_N = \mathbf{I}_N$, which leads to*

$$\begin{bmatrix} m_0 & m_1 & m_2 & 0 & 0 & 0 & m_2 & m_1 \\ m_1 & m_0 & m_1 & m_2 & 0 & 0 & 0 & m_2 \\ m_2 & m_1 & m_0 & m_1 & m_2 & 0 & 0 & 0 \\ 0 & m_2 & m_1 & m_0 & m_1 & m_2 & 0 & 0 \\ 0 & 0 & m_2 & m_1 & m_0 & m_1 & m_2 & 0 \\ 0 & 0 & 0 & m_2 & m_1 & m_0 & m_1 & m_2 \\ m_2 & 0 & 0 & 0 & m_2 & m_1 & m_0 & m_1 \\ m_1 & m_2 & 0 & 0 & 0 & m_2 & m_1 & m_0 \end{bmatrix} \begin{bmatrix} \hat{\sigma}_0 \\ \hat{\sigma}_1 \\ \hat{\sigma}_2 \\ x_3 \\ x_4 \\ x_3 \\ \hat{\sigma}_2 \\ \hat{\sigma}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ \\ \vdots \\ 0 \\ 0 \end{bmatrix},$$

where $x_3 := \sigma_3 = \sigma_5$ and $x_4 := \sigma_4$ are the unknown extended covariance lags.

Rearranging and eliminating the last three redundant equations, one obtains

$$\begin{aligned} m_0\hat{\sigma}_0 + 2m_1\hat{\sigma}_1 + 2m_2\hat{\sigma}_2 &= 1 \\ m_0\hat{\sigma}_1 + m_1(\hat{\sigma}_0 + \hat{\sigma}_2) + m_2(\hat{\sigma}_1 + x_3) &= 0 \\ m_0\hat{\sigma}_2 + m_1(\hat{\sigma}_1 + x_3) + m_2(\hat{\sigma}_0 + x_4) &= 0 \\ m_0 x_3 + m_1(\hat{\sigma}_2 + x_4) + m_2(\hat{\sigma}_1 + x_3) &= 0 \\ m_0 x_4 + 2m_1 x_3 + 2m_2\hat{\sigma}_2 &= 0 \end{aligned}$$

which is a system of five quadratic equations in five unknowns whose solution already looks non-trivial. It may be checked that, under positivity of the matrix Toepl$\{\hat{\sigma}_0, \hat{\sigma}_1, \hat{\sigma}_2\}$*, it has a unique positive definite solution (i.e. making* $\mathbf{M}_N$ *positive definite).*

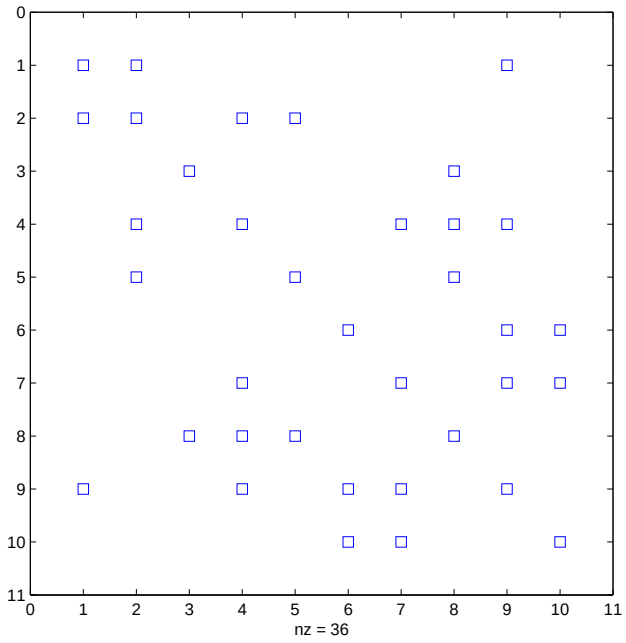
At a first sight, the block-circulant band extension problem 2.2.4 seems to be a particular instance of the general covariance extension problem studied by A. P. Dempster in [16]. Letting $\mathcal{I}$ denote the set of pairs of block indices consistent with a generic (namely not necessarily banded and circulant) symmetric structure, i.e.

$$(i,j) \in \mathcal{I} \Rightarrow (j,i) \in \mathcal{I}$$

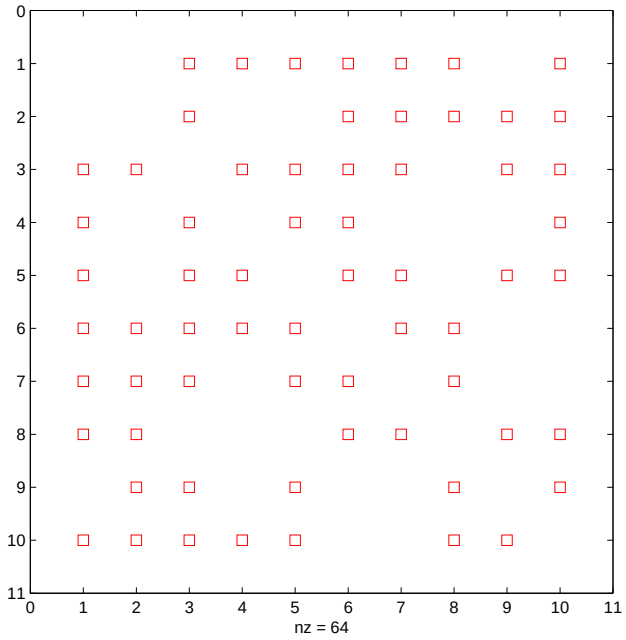
(see also figure 2.4a) and $\mathcal{R}$ be the corresponding set of block-entries they also consistent with a generic symmetric structure, i.e.

$$\mathcal{R} := \left\{ r_{ij} \in \mathbb{R}^{m\times m} \mid (i,j) \in \mathcal{I},\, r_{ij} = r_{ji}^{\top} \right\} ,$$

the problem studied by Dempster reads as follows.



(a) Block index set $\mathcal{I}$



(b) Block index set $\mathcal{I}^c$

Figure 2.4: Block index sets $\mathcal{I}$ and $\mathcal{I}^c$.

Problem 2.2.5 (**Covariance Selection Problem - Dempster '72**).

$$\textit{Find } \mathbf{\Sigma}_N \in \mathfrak{S}_N \textit{ such that} \tag{2.36a}$$

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij}, \ \textit{for } (i,j) \in \mathcal{I} \textit{ and } r_{ij} \in \mathcal{R}, \tag{2.36b}$$

$$\mathbf{\Sigma}_N > 0, \tag{2.36c}$$

$$e_i^\top \mathbf{\Sigma}_N^{-1} e_j = 0, \ \textit{for } (i,j) \notin \mathcal{I}. \tag{2.36d}$$

In words, given a partially specified symmetric matrix find a completion $\mathbf{\Sigma}_N$ *which agrees with the partially specified one in the given positions, is symmetric positive definite and such that its inverse has zeros in the complementary positions of those assigned.*

Even by setting $\mathcal{I} = \mathcal{I}_b$ and $\mathcal{R} = \mathcal{R}_b$ in (2.36b), (2.36d), it is apparent that the linear constraint (2.35e) that enforces the completed matrix to be circulant is not present in the Dempster's setting. Thus, in principle, even if the given data lie on a band symmetric with respect to the main diagonal and on the NE and SW corners and they are consistent with the symmetric block-circulant structure, the solution of the Dempster's problem need not be circulant. Neverthless, Dempster's work contains an observation which will reveal to be important for our purposes and reads as follows.

Proposition 2.2.1 (**Dempster**). *Assume that Problem 2.2.5 is feasibile. Then, for the solution of the covariance selection problem, the following properties hold:*

1. Existence and uniqueness. *If there is a positive definite symmetric completion, then there is* exactly one *with the additional property that its* inverse has zeros in the complementary positions *of those where the elements of the covariance are assigned.*

2. Maximum Entropy. *This extension corresponds to the Gaussian distribution with* maximum entropy.

Recall that the *differential entropy* $H(p)$ of a probability distribution with density p on $\mathbb{R}^n$ is defined by

$$H(p) = -\int_{\mathbb{R}^n} \log(p(x))p(x)dx. \tag{2.37}$$

In the case of a zero-mean Gaussian distribution p with covariance matrix $\mathbf{\Sigma}_N$, we get

$$H(p) = \frac{1}{2}\log(\det \mathbf{\Sigma}_N) + \frac{1}{2}n\left(1 + \log(2\pi)\right). \tag{2.38}$$

Thus, exploiting the maximum entropy principle of Proposition 2.2.1, Problem 2.2.5 can be restated as:

Problem 2.2.6 (**Dempster's maximum entropy problem (DMEP)**)**.**

$$\max\{\log\det \mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in \mathfrak{S}_N,\ \mathbf{\Sigma}_N > 0\} \tag{2.39a}$$

subject to :

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij},\ \textit{for } (i,j) \in \mathcal{I} \textit{ and } r_{ij} \in \mathcal{R}. \tag{2.39b}$$

which, from now on, will be referred to as the Dempster's maximum entropy problem (DMEP, for short). The DMEP has been extensively studied in the literature. A fundamental reference, besides [16], is the work by Grone, Johnson, Sa and Wolkowicz [25], see also [2] for the specialization where the given sparsity pattern corresponds to a chordal graph (we will come back to this in Chapter 4) and [18] and [24] for the case of given data consistent with a banded Toeplitz structure and later generalizations to generic banded algebras.

An independent, simple argument which explains the zero–pattern in the inverse of the solution of the DMEP is the following.

Proof of 2.2.1. Clearly if M_{me} exists, it is uniquely defined. Computing the

Lagrangian

$$\mathcal{L}(M, \lambda_{k,\ell}) := \log\det(M) + \sum_{(k,\ell)\in\mathcal{I}} \lambda_{k,\ell}(m_{k,\ell} - e_k M e_\ell^*) \tag{2.40}$$

and setting the derivative of $\mathcal{L}$ with respect to the entries of M equal to zero, we readily obtain

$$M^{-1} = \sum_{(k,\ell)\in\mathcal{I}} \lambda_{k,\ell} e_k^* e_\ell. \tag{2.41}$$

i.e. the inverse of the maximizer has the claimed zero-pattern. □

See [16] and [25] for alternative proofs.

Inspired by the maximum entropy principle of Dempster's setting, we turn to consider the following problem.

Problem 2.2.7 (**Maximum entropy extension problem for block—circulant matrices (CMEP)**).

$$\max\left\{\log\det\mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in (\mathfrak{S}_N \cap \mathfrak{C}_N),\ \mathbf{\Sigma}_N > 0\right\} \tag{2.42a}$$

subject to :

$$e_i^\top \mathbf{\Sigma}_N e_j = r_{ij},\ \textit{for } (i,j) \in \mathcal{I}_b \textit{ and } r_{ij} \in \mathcal{R}_b. \tag{2.42b}$$

This problem has not been studied in the literature. In the next Chapter, we will solve Problem 2.2.7 and show that it is equivalent to the original block-circulant band extension problem 2.2.4. The relation between the solutions of the DMEP 2.2.6 and of the CMEP 2.2.7 will also be highlighted.

Chapter 3

A maximum entropy solution of the covariance extension problem for reciprocal processes

In this chapter the maximum entropy extension problem for block-circulant matrices (CMEP) is studied. In particular, it will be shown that solving this problem leads to a complete solution of the maximum likelihood identification problem for reciprocal processes. The first step will be to rewrite the problem in such a way as to make it easier to handle (Section 3.1). In Section 3.2 the feasibility issue will be addressed. In Section 3.3 existence and uniqueness of the solution will be proved. In Section 3.4 this unique solution will be shown to have a banded inverse. This is the result we hope for, since it shows that solving the CMEP solves also the maximum likelihood identification problem for reciprocal processes we started with. Moreover, the inverse's bandedness property together with the Dempster's results, implies that the maximum entropy distribution, subject only to moment constraints consistent with the block–circulant structure, is automatically block–circulant. A nontrivial generalization of this result is provided in Section 3.5. This generalization is

accomplished via an independent argument that exploits the invariance of the determinant with respect to the group of transformations that leave circulant matrices invariant. Section 3.6 contains some examples which provide insight into the above developed theory and highlight a connnection with the factorization of certain polynomials in many variables which is facilitated by the circulant structure. Some final remarks are collected in Section 3.7, which concludes the Chapter.

3.1 The maximum entropy extension problem for banded block-circulant matrices

Let $\mathbf{U}_N$ denote the block-circulant "shift" matrix with $N \times N$ blocks,

$$\mathbf{U}_N = \begin{bmatrix} 0 & I_m & 0 & \dots & 0 \\ 0 & 0 & I_m & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_m \\ I_m & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Clearly, $\mathbf{U}_N^\top \mathbf{U}_N = \mathbf{U}_N \mathbf{U}_N^\top = I_{mN}$, i.e. $\mathbf{U}_N$ is orthogonal. Recall that a matrix $\mathbf{C}_N$ with $N \times N$ blocks is block-circulant if and only if it commutes with $\mathbf{U}_N$, namely if and only if it satisfies

$$\mathbf{U}_N^\top \mathbf{C}_N \mathbf{U}_N = \mathbf{C}_N. \tag{3.1}$$

Moreover, let $\mathbf{T}_n \in \mathfrak{S}_{n+1}$ be the Toeplitz matrix of *boundary data*

$$\mathbf{T}_n = \begin{bmatrix} \Sigma_0 & \Sigma_1^\top & \dots & \dots & \Sigma_n^\top \\ \Sigma_1 & \Sigma_0 & \Sigma_1^\top & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \Sigma_1^\top \\ \Sigma_n & \dots & \dots & \Sigma_1 & \Sigma_0 \end{bmatrix} \tag{3.2}$$

and E_n the $N \times (n+1)$ block matrix

$$E_n = \begin{bmatrix} I_m & 0 & \dots & 0 \\ 0 & I_m & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & I_m \\ 0 & \dots & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & \dots & 0 \end{bmatrix}.$$

In this notation, Problem 2.2.7 can be restated as:

Problem 3.1.1 (**Maximum entropy extension problem for block-circulant matrices (CMEP)**)**.**

$$\max \{\log\det \mathbf{\Sigma}_N \mid \mathbf{\Sigma}_N \in \mathfrak{S}_N,\ \mathbf{\Sigma}_N > 0\} \tag{3.3a}$$

subject to :

$$E_n^\top \mathbf{\Sigma}_N E_n = \mathbf{T}_n, \tag{3.3b}$$

$$\mathbf{U}_N^\top \mathbf{\Sigma}_N \mathbf{U}_N = \mathbf{\Sigma}_N. \tag{3.3c}$$

The above problem indeed amounts to finding the maximum entropy Gaussian distribution with a block-circulant covariance, whose first $n+1$ blocks are precisely $\Sigma_0, \dots, \Sigma_n$. We observe that in this problem we are

minimizing a strictly convex function on the intersection of a convex cone (minus the zero matrix) with a linear manifold. Hence, we are dealing with a *convex optimization problem.*

Moreover, note that we are not imposing that the inverse of the solution $\mathbf{\Sigma}_N$ of Problem 3.1.1 should have a banded structure. We shall see that, whenever solutions exist, this property will be *automatically guaranteed.*

The first question to be addressed is feasibility of Problem 3.1.1, which will be the object of the next session.

3.2 Feasibility

In this Section we study feasibility of the CMEP 3.1.1. The first Theorem provides insight into the shape of the set of all positive definite block-circulant completions of a partially specified covariance matrix. Next, we will investigate what conditions on the data must be satisfied for Problem 3.1.1 to be feasible. Theorem 3.2.2 provides a sufficient condition for arbitrary block–size m and bandwidth n. A necessary and sufficient condition for unitary block–size and bandwidth is also established (see Theorem 3.2.4 below).

3.2.1 Structure of the feasible set

Theorem 3.2.1 (**Shape of the feasible set**). *Let* $\mathbf{\Sigma}_N = \text{Circ}\left\{\Sigma_0, \Sigma_1^\top, \dots, \Sigma_1\right\}$ *be a partially specified* $N \times N$ *m-block circulant matrix. The* set of all positive definite block-circulant completions *of* $\mathbf{\Sigma}_N$ *is delimited by the* intersection of the m-order surfaces *defined by the positive semidefiniteness of the matrices*

$$\Psi(e^{-j\frac{2\pi}{N}\ell}) = \sum_{k=0}^{N-1} \Sigma_k^\top e^{-j\frac{2\pi}{N}\ell k}, \qquad \textit{for } \ \ell = 0, 1, \dots, N-1.$$

Proof. The result follows immediately recalling (see Theorem A.0.2) that every block-circulant matrix with $N \times N$ blocks each of size $m \times m$, say

$\mathbf{C}_N = \mathrm{Circ}\{C_0, C_1, \ldots, C_{N-1}\}$, can be diagonalized as

$$(F^* \otimes I_m)\, \mathbf{C}_N\, (F \otimes I_m) = \mathrm{diag}\{\Psi(w^0), \Psi(w^1), \Psi(w^2), \ldots, \Psi(w^{N-1})\}$$

where $\otimes$ denote the Kronecker product, F is the Fourier matrix of order N, i.e. the matrix whose (k,l)-entry is

$$f_{k,l} = \frac{1}{\sqrt{N}} w^{(k-1)(l-1)} \tag{3.4}$$

where $w := e^{-j\frac{2\pi}{N}}$, j denoting the imaginary unit $\sqrt{-1}$, and the $\Psi(w^\ell)$'s are the polynomial matrices

$$\Psi(x) = \sum_{k=0}^{N-1} x^k C_k$$

computed for $x = w^\ell$, $\ell = 0, \ldots, N-1$. □

3.2.2 A sufficient condition for generic block–size and bandwidth

Theorem 3.2.1 throws light on the structure of the feasible set, but what are the conditions on the data for Problem 3.1.1 to be feasible, namely what are the conditions on the matrices $\Sigma_0, \ldots, \Sigma_n$ so that a positive definite, symmetric matrix $\mathbf{\Sigma}_N$ satisfying (3.3b)-(3.3c) exists? Obviously, $\mathbf{T}_n$ positive definite is a necessary condition for the existence of such a $\mathbf{\Sigma}_N$. In general it turns out that, under such a necessary condition, feasibility holds for N large enough. This is, in short, the content of the next theorem. The idea is that for $N \to \infty$, existence of a positive block-circulant extension can be derived from the existence of positive extensions for Toeplitz matrices.

Theorem 3.2.2 (Sufficient condition - block case)**.** *Given the sequence* $\Sigma_i \in$

$\mathbb{R}^{m\times m}$, $i = 0, 1, \ldots, n$, *such that*

$$\mathbf{T}_n = \mathbf{T}_n^\top > 0, \tag{3.5}$$

there exists $\bar{N}$ *such that for* $N \geq \bar{N}$, *the matrix* $\mathbf{T}_n$ *can be extended to an* $N \times N$ *block-circulant, positive-definite symmetric matrix* $\mathbf{\Sigma}_N$.

Proof. A fundamental result in stochastic system theory is the so-called maximum entropy covariance extension. It states that, under condition (3.5), there exists a rational positive real function $\Phi_+(z) = \frac{\Sigma_0}{2} + C(zI - A)^{-1}B$ such that

1. A has spectrum strictly inside the unit circle.

2. $\Sigma_i = CA^{i-1}B$, $i = 1, 2, \ldots, n$.

3. The spectrum $\Phi(z) := \Phi_+(z) + \Phi_+^*(z)$ is coercive, i.e.

$$\exists \varepsilon > 0 \text{ such that } \Phi(\mathrm{e}^{\mathrm{j}\vartheta}) > \varepsilon I, \ \forall \vartheta \in [0, 2\pi). \tag{3.6}$$

In fact $\Phi(z)$ has no zeros on the unit circle since it can be expressed in the form

$$\Phi(z) = L_n(z^{-1})^{-1}\Lambda_n L_n(z)^{-\top}$$

where $\Lambda_n = \Lambda_n^\top > 0$ and $L_n(z^{-1})$ is the n–th Levinson-Whittle matrix polynomial (also called n–th matrix Szegö polynomial) of the block Toeplitz matrix $\mathbf{T}_n$,

$$L_n(z^{-1}) := \sum_{k=0}^{n} A_n(k) z^{-k},$$

where the $A_n(k)$'s are the solution of the Yule-Walker equation

$$\begin{bmatrix} A_n(0) & A_n(1) & \ldots & A_n(n) \end{bmatrix} \mathbf{T}_n^\top = \begin{bmatrix} 0 & 0 & \ldots & 0 \end{bmatrix},$$

with $A_n(0) = I_m$; see [49], [14] and [50].

Let $\Sigma_i := CA^{i-1}B$, $i = n+1, n+2, \ldots$, so that $\Phi_+(z) = \frac{\Sigma_0}{2} + \sum_{i=1}^{\infty} \Sigma_i z^{-i}$, and define

$$\mathbf{\Sigma}_N := \mathrm{Circ}\Big(\Sigma_0, \Sigma_1^\top, \Sigma_2^\top, \ldots, \Sigma_{\frac{N-1}{2}}^\top, \Sigma_{\frac{N-1}{2}}, \Sigma_{\frac{N-1}{2}-1}, \ldots \Sigma_1\Big), \tag{3.7}$$

for N odd, and

$$\mathbf{\Sigma}_N := \mathrm{Circ}\Big(\Sigma_0, \Sigma_1^\top, \Sigma_2^\top, \ldots, \Sigma_{\frac{N-2}{2}}^\top, \Sigma_{\frac{N}{2}}^\top + \Sigma_{\frac{N}{2}}, \Sigma_{\frac{N-2}{2}}, \Sigma_{\frac{N-2}{2}-1}, \ldots \Sigma_1\Big), \tag{3.8}$$

for N even. We need to show that there exists $\bar{N}$ such that $\mathbf{\Sigma}_N > 0$ for $N \geq \bar{N}$. To this aim, notice that $\mathbf{\Sigma}_N$ is, by definition, block-circulant so that, a similarity transformation induced by a unitary matrix $(F \otimes I_m)$ reduces $\mathbf{\Sigma}_N$ to a block-diagonal matrix:

$$(F^* \otimes I_m)\mathbf{\Sigma}_N(F \otimes I_m) = \mathbf{\Psi}_N := \mathrm{diag}\left(\Psi_0, \Psi_1, \ldots, \Psi_{N-1}\right),$$

where $(F \otimes I_m)$ is the Fourier block-matrix whose (k,l)–th block is

$$(F \otimes I_m)_{k,l} = \exp\left[-\mathrm{j}2\pi(k-1)(l-1)/N\right] I_m$$

and Ψ_ℓ are the coefficients of the finite Fourier transform of the first block row of $\mathbf{\Sigma}_N$:

$$\Psi_\ell = \Sigma_0 + \mathrm{e}^{\mathrm{j}\vartheta_\ell}\Sigma_1^\top + \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^2 \Sigma_2^\top + \cdots + \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{N-2} \Sigma_2 + \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{N-1} \Sigma_1, \tag{3.9}$$

with $\vartheta_\ell := -2\pi\ell/N$, see e.g. [47, Sec. 3.4]. Clearly, $\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{N-i} = \left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)^{-i}$ and hence

$$\Psi_\ell = \Phi\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) - \left[\delta\Phi_N\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) + \delta\Phi_N^*\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)\right] \tag{3.10}$$

where,

$$\begin{aligned}\delta\Phi_N(z) &:= \sum_{i=h+1}^{\infty} \Sigma_i z^{-i} = \sum_{i=h+1}^{\infty} CA^{i-1}Bz^{-i} \\ &= z^{-h}CA^h(zI-A)^{-1}B, \end{aligned} \tag{3.11}$$

with

$$h := \begin{cases} \frac{N-1}{2}, & N \text{ odd} \\ N/2, & N \text{ even} \end{cases}$$

Since A is a stability matrix, if N, and hence h, is large enough, $\delta\Phi_N\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) + \delta\Phi_N^*\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right)$ is dominated by εI, i.e. there exists $\bar{N}$ such that

$$\delta\Phi_N\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) + \delta\Phi_N^*\left(\mathrm{e}^{\mathrm{j}\vartheta_\ell}\right) < \varepsilon I, \quad \forall\vartheta_\ell, \quad \forall N \geq \bar{N} \tag{3.12}$$

so that it readily follows from (3.6) and (3.10) that if $N \geq \bar{N}$, $\Psi_\ell > 0$ for all ℓ. □

Remark 3.2.1. We observe that, given $\mathbf{T}_n$, the triple A, B, C can be explicitly computed. In fact, it is well known [19] that if $W(z) = C(zI-A)^{-1}\tilde{B}+\tilde{D}$ is a spectral factor of Φ then the analytic (with rspect to the unit circle) component of Φ is given by

$$\Phi_+(z) = \frac{\Sigma_0}{2} + C(zI-A)^{-1}B$$

where $B = APC^\top + \tilde{B}\tilde{D}^\top$, with P solution of the Lyapunov equation

$$P = APA^\top + \tilde{B}\tilde{B}^\top$$

and $\Sigma_0 = CPC^\top + DD^\top$. Thus, letting

$$\Lambda_n^{-\frac{1}{2}} L_n(z^{-1}) := H(zI-F)^{-1}G + J$$

we get $J = \Lambda_n^{-\frac{1}{2}}$, $H = \Lambda_n^{-\frac{1}{2}} \begin{bmatrix} A_n(n) & A_n(n-1) & \dots & A_n(1) \end{bmatrix}$,

$$G = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I_m \end{bmatrix}, \qquad F = \begin{bmatrix} 0 & I_m & 0 & 0 & \dots & 0 \\ 0 & 0 & I_m & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & I_m \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix},$$

so that the spectral factor $W(z)$ results

$$W(z) = \left(L_n(z^{-1})\right)^{-1} \Lambda_n^{\frac{1}{2}} = (H(zI - F)^{-1}G + J)^{-1} = C(zI - A)^{-1}\tilde{B} + \tilde{D}$$

with $\tilde{D} = \Lambda_n^{\frac{1}{2}}$, $C = -\begin{bmatrix} A_n(n) & A_n(n-1) & \dots & \dots & A_n(1) \end{bmatrix}$

$$A = \begin{bmatrix} 0 & I_m & 0 & 0 & \dots & 0 \\ 0 & 0 & I_m & 0 & \dots & 0 \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & I_m \\ -A_n(n) & -A_n(n-1) & \dots & \dots & \dots & -A_n(1) \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \Lambda_n^{\frac{1}{2}} \end{bmatrix}.$$

Finally, the matrix B is given by $B^\top = APC^\top + \tilde{B}\tilde{D}$ with $P = APA^\top + \tilde{B}\tilde{B}^\top$. Once the matrices A, B and C are known, we can compute ε and $\bar{N}$ for which (3.12) holds. In other words, *Theorem 3.2.2 provides a sufficient condition that can be practically tested.* Similar bounds, but valid only for the scalar case, were derived in [15].

3.2.3 A necessary and sufficient condition for unitary block–size and bandwidth one

In this Section we provide a necessary and sufficient condition of the CMEP 3.1.1 for the special case of unitary block–size and bandwidth. We will start considering a real-valued discrete-time stationary periodic process with circulant covariance matrix and we will write explicitly the covariance samples of such a process. Next we will reformulate the feasibility of the CMEP for such a process in terms of solvability of a system of linear equations and provide necessary and sufficient conditions for the solvability of this system. An intermediate step will be to consider a relaxed version of the feasibility problem where the completion is required to be only positive *semidefinite*.

Lemma 3.2.1. *Let $\{y(t)\}$ be a stationary periodic process of period N taking values in $\mathbb{R}$. Then*

(i) $\{y(t)\}$ can be represented as

$$y(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k e^{jkt\frac{2\pi}{N}}, \tag{3.13}$$

where

$$c_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y(t) e^{-jkt\frac{2\pi}{N}} \tag{3.14}$$

(ii) the covariance samples of $\{y(t)\}$ are given by

$$\sigma_m = \frac{1}{N}\Big\{ \mathbb{E}\left[|c_0|^2\right] + \sum_{k=1}^{\frac{N}{2}-1} \mathbb{E}\left[|c_k|^2\right] 2\cos\left(km\frac{2\pi}{N}\right) + \mathbb{E}\left[\left|c_{\frac{N}{2}}\right|^2\right] \cos(m\pi) \Big\} \tag{3.15}$$

for N even, and

$$\sigma_m = \frac{1}{N}\Big\{\mathbb{E}\left[|c_0|^2\right] + \sum_{k=0}^{\frac{N-1}{2}} \mathbb{E}\left[|c_k|^2\right] 2\cos\left(km\frac{2\pi}{N}\right)\Big\} \tag{3.16}$$

for N odd, where the c_k's are independent random variables.

Proof. (i) The thesis follows immediately by observing that every periodic process $y(t)$ can be seen as the solution of a linear homogeneous constant coefficients difference equation of the type

$$y(t) - y(t-N) = 0.$$

(ii) Equation (3.13) can be written in matrix form as

$$\mathbf{x} = F^*\mathbf{c} \tag{3.17}$$

where, as usual, F denote the Fourier matrix (A.3) and $\mathbf{x}$ and $\mathbf{c}$ are the column vectors obtained by stacking the $x(k)$'s and the c_k's for $k = 0, \ldots, N-1$. From (3.17) we have

$$\mathbb{E}\left[\mathbf{c}\mathbf{c}^*\right] = F\mathbf{\Sigma}_N F^*.$$

which, since $\mathbf{\Sigma}_N$ is circulant (see Proposition A.0.2), implies that $\mathbb{E}\left[\mathbf{c}\mathbf{c}^*\right]$ is diagonal, i.e. the c_k's are linearly independent random variables. By the independence of the c_k's, the covariance samples

$$\sigma_m = \mathbb{E}\left[x(t+m)x(n)^*\right] = \mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} c_k e^{jk(t+m)\frac{2\pi}{N}}\ \frac{1}{\sqrt{N}}\sum_{\ell=0}^{N-1} c_\ell e^{j\ell m\frac{2\pi}{N}}\right],$$

$m = 0, \ldots, N-1$, becomes

$$\sigma_m = \frac{1}{N}\sum_{k=0}^{N-1} \mathbb{E}\left[|c_k|^2\right] e^{jkm\frac{2\pi}{N}},$$

which, exploiting the Hermitian symmetry of the $\{c_k\}$, yields equations (3.15), (3.16). □

Lemma 3.2.1 let us reformulate the feasibility of the CMEP for $m = n = 1$ in terms of the solution of a linear system of equations. This is the content of the Corollary 3.2.1 below.

Corollary 3.2.1. *Let*

$$\mathbf{T}_2 = \begin{bmatrix} 1 & \sigma_1 \\ \sigma_1 & 1 \end{bmatrix}$$

be the matrix of the boundary data, where, without loss of generality, we have assumed $\sigma_0 = 1$*, and let* $\sigma_1 \in \mathbb{R}$ *and* N *be a positive integer,* $N > 3$*. The data matrix* $\mathbf{T}_2$ *admits a positive definite (respectively, a positive semidefinite) circulant completion if and only if the systems*

$$\begin{cases} \frac{1}{N}\left(p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2\,p_k + p_{\frac{N}{2}}\right) = 1 \\ \frac{1}{N}\left(p_0 + \sum_{k=1}^{\frac{N}{2}-1} 2\,p_k\,\cos\left(k\frac{2\pi}{N}\right) - p_{\frac{N}{2}}\right) = \sigma_1 \end{cases} \text{, for } N \text{ even} \tag{3.18}$$

and

$$\begin{cases} \frac{1}{N}\left(p_0 + \sum_{k=1}^{\frac{N-1}{2}} 2\,p_k\right) = 1 \\ \frac{1}{N}\left(p_0 + \sum_{k=1}^{\frac{N-1}{2}} 2\,p_k\,\cos\left(k\frac{2\pi}{N}\right)\right) = \sigma_1 \end{cases} \text{, for } N \text{ odd} \tag{3.19}$$

with the constraints $p_k > 0$ *(respectively,* $p_k \geq 0$*),* $k = 0, \ldots, N-1$*, admit solution.*

Before stating the necessary and sufficient condition for the feasibility CMEP (see Theorem 3.2.4 below), we give the following intermediate result, which answer the question on what conditions on the data must be satisfied for a positive *semidefinite* completion to exist.

Theorem 3.2.3. *The data matrix* $\mathbf{T}_2$ *admits a positive* semidefinite *circulant completion if and only if*

- $|\sigma_1| \leq 1$, *for* N *even;*

- $\cos\left(\frac{N-1}{N}\pi\right) \leq \sigma_1 \leq 1$, *for* N *odd.*

Proof. By Corollary 3.2.1, it suffices to prove that system (3.18) for N even ((3.19) for N odd) with the constraints $p_k \geq 0$, $k = 0, \ldots, N-1$ has solution if and only if $\sigma_1 \in [-1, 1]$ ($\cos\left(\frac{N-1}{N}\pi\right) \leq \sigma_1 \leq 1$, respectively). Necessity is trivial both for N even and N odd. For what concern sufficiency, we must distinguish between the two cases.

Sufficiency for N even: we want to prove that if $|\sigma_1| \leq 1$, then system (3.18) with the constraints $p_k \geq 0$, $k = 0, \ldots, N-1$, has solution.

- If $\sigma_1 = 1$ ($\sigma_1 = -1$), the set $\{p_0 = N\,, p_k = 0 \text{ for } k \neq 0\}$ (respectively, the set $\left\{p_{\frac{N}{2}} = N\,,\, p_k = 0 \text{ for } k \neq \frac{N}{2}\right\}$) is solution of (3.18);

- For $|\sigma_1| < 1$, let $\alpha \in [0, 1]$; if one sets

$$\begin{cases} p_0 = N\alpha \\ p_{\frac{N}{2}} = N\,(1-\alpha) \\ p_k = 0 \qquad\qquad\qquad \forall k \neq 0 \text{ and } k \neq \frac{N}{2} \end{cases}$$

  then the first of (3.18) is satisfied, while the second yields $\sigma_1 = 2\alpha - 1$, which means that as α varies continuously over the interval $[0, 1]$, σ_1 varies continuously over $[-1, 1]$, i.e. system (3.18) is solvable for every $|\sigma_1| \leq 1$.

Sufficiency for N odd: we want to prove that if $\cos\left(\frac{N-1}{N}\pi\right) \leq \sigma_1 \leq 1$, system (3.19) with the constraints $p_k \geq 0$, $k = 0, \ldots, N-1$ admits a solution.

- If $\sigma_1 = 1$ ($\sigma_1 = \cos\left(\frac{N-1}{N}\pi\right)$), the set $\{p_0 = N\,, p_k = 0 \text{ for } k \neq 0\}$ (respectively, the set $\left\{p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}} = \frac{N}{2}\,, p_k = 0 \text{ for } k \neq \frac{N-1}{2}\,, \frac{N+1}{2}\right\}$) is solution of (3.19);

- If $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$, let $\alpha \in [0,1]$; if one chooses

$$\begin{cases} p_0 = N\alpha \\ p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}} = \frac{N(1-\alpha)}{2} \\ p_k = 0 \qquad\qquad \forall k \neq 0,\ k \neq \frac{N-1}{2} \text{ and } k \neq \frac{N+1}{2} \end{cases} \tag{3.20}$$

then the first of (3.19) is satisfied while the second yields

$$\sigma_1 = \left[1 - \cos\left(\frac{N-1}{N}\pi\right)\right]\alpha + \cos\left(\frac{N-1}{N}\pi\right),$$

which implies that as α varies continuously over the interval $[0,1]$, σ_1 varies continuously over $\left[\cos\left(\frac{N-1}{N}\pi\right), 1\right]$, i.e. the system (3.19) is solvable for every $\sigma_1 \in \left[\cos\left(\frac{N-1}{N}\pi\right), 1\right]$, which concludes the proof.

□

We are now ready to state our main result.

Theorem 3.2.4 (**Necessary and sufficient condition - unitary block–size and bandwidth**). *The data matrix* $\mathbf{T}_2$ *admits a* positive definite *circulant completion if and only if*

- $|\sigma_1| < 1$, *for* N *even;*

- $\cos\left(\frac{N-1}{N}\pi\right) \leq \sigma_1 < 1$, *for* N *odd.*

Proof. Once again, by Corollary 3.2.1, it suffices to prove that system (3.18) for N even ((3.19) for N odd) with the constraints $p_k > 0$, $k = 0, \ldots, N-1$ has solution if and only if $|\sigma_1| < 1$ (respectively, if and only if $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$). Necessity is trivial both for N even and N odd. For what concern sufficiency, the two cases must be distinguish.

Sufficiency for N even: we need to prove that if $|\sigma_1| < 1$, then system (3.18) with constraints $p_k > 0$, $k = 0, \dots, N-1$ has solution. In fact, setting

$$\begin{cases} p_0 = N(\alpha - \varepsilon) \\ p_{\frac{N}{2}} = N(1 - \alpha - \varepsilon) \end{cases}$$

with $\alpha \in (0,1)$, $0 < \varepsilon < \min\{\alpha, 1-\alpha\}$, then

$$\begin{cases} \frac{1}{N}\left(p_0 + p_{\frac{N}{2}}\right) = \alpha - \varepsilon + 1 - \alpha - \varepsilon = 1 - 2\varepsilon \\ \frac{1}{N}\left(p_0 - p_{\frac{N}{2}}\right) = \alpha - \varepsilon - 1 + \alpha + \varepsilon = 2\alpha - 1 \end{cases}.$$

Thus if one chooses the remaining p_k in order to satisfy the system

$$\begin{cases} \frac{1}{N}\sum_{k=1}^{\frac{N}{2}-1} 2\, p_k = 2\varepsilon \\ \frac{1}{N}\sum_{k=1}^{\frac{N}{2}-1} 2\, p_k \cos(k\frac{2\pi}{N}) = 0 \end{cases}, \tag{3.21}$$

then $\sigma_0 = 1$ and $\sigma_1 = 2\alpha - 1$, i.e. as α varies continuously over $(0,1)$, σ_1 varies continuously over $(-1,1)$, namely system (3.18) is solvable for every $\sigma_1 \in (-1,1)$. Since $\sum_{k=1}^{\frac{N}{2}-1} \cos\left(k\frac{2\pi}{N}\right) = 0$, it is easy to see that

$$p_k = \frac{N\varepsilon}{(\frac{N}{2}-1)}, \quad k = 1, \dots, \frac{N}{2} - 1$$

is a solution of (3.21) and thus

$$\begin{cases} p_0 = N(\alpha - \varepsilon) \\ p_{\frac{N}{2}} = 1 - \alpha - \varepsilon \\ p_k = \frac{N\varepsilon}{(\frac{N}{2}-1)} \qquad k = 1, \dots, \frac{N}{2} - 1 \end{cases}$$

with $\alpha = \frac{\sigma_1+1}{2}$, $0 < \varepsilon < \alpha$ solves (3.18) with constraints $p_k > 0$, $k = 0, \dots, N-1$. Finally, let's note that if $\sigma_1 = -1$ (i.e. $\alpha = 0$), then $p_0 = -\varepsilon < 0$

and if $\sigma_1 = 1$ (i.e. $\alpha = 1$), then $p_{\frac{N}{2}} = -\varepsilon < 0$, namely (3.18) with constraints $p_k > 0$, $k = 0, \dots, N-1$ does not admit solution for $\sigma_1 = \pm 1$.

Sufficiency for N odd: the aim is to prove that if $\cos\left(\frac{N-1}{N}\pi\right) < \sigma_1 < 1$, then system (3.19) with constraints $p_k > 0$, $k = 0, \dots, N-1$ has solution. In fact, if we subtract the quantities $N\varepsilon$ and $-\frac{N\varepsilon}{2\cos\left(\frac{N-1}{N}\pi\right)}$ from p_0 and $p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}}$ in (3.20), i.e. if we set

$$\begin{cases} p_0 = N(\alpha - \varepsilon) \\ p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}} = N(\frac{1-\alpha}{2} + \frac{\varepsilon}{2\cos\left(\frac{N-1}{N}\pi\right)}) \end{cases} \tag{3.22}$$

and redistribute the total quantity which has been subtracted, namely

$$N\left(\varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)}\right)$$

among all the p_k's, namely if we set

$$\begin{cases} p_0 = N(\alpha - \varepsilon) + \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)} \\ p_{\frac{N-1}{2}} = p_{\frac{N+1}{2}} = N(\frac{1-\alpha}{2} + \frac{\varepsilon}{2\cos\left(\frac{N-1}{N}\pi\right)}) + \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)} \\ p_k = \varepsilon - \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)}, \quad k \neq 0, \frac{N-1}{2}, \frac{N+1}{2} \end{cases} \tag{3.23}$$

with $\alpha \in (0,1)$, $0 < \varepsilon < \min\left\{-\frac{N\alpha}{1-N-\frac{1}{\cos\left(\frac{N-1}{N}\pi\right)}}, \frac{N}{2}\frac{\alpha-1}{\frac{N-2}{2\cos\left(\frac{N-1}{N}\pi\right)}+1}\right\}$, then the first of (3.19) is satisfied and the second yields $\sigma_1 = \left[1 - \cos\left(\frac{N-1}{N}\pi\right)\right]\alpha + \cos\left(\frac{N-1}{N}\pi\right)$, which means that as α varies continuously over the interval $(0,1)$, σ_1 varies continuously over $\left(\cos\left(\frac{N-1}{N}\pi\right), 1\right)$, i.e. the system (3.19) is solvable for every $\sigma_1 \in \left(\cos\left(\frac{N-1}{N}\pi\right), 1\right)$. Moreover if $\sigma_1 = 1$ (i.e. $\alpha = 1$), then $p_{\frac{N-1}{2}} = \frac{\varepsilon}{\cos\left(\frac{N-1}{N}\pi\right)}\left(\frac{N}{2} - 1\right) + \varepsilon < 0$ and if $\sigma_1 = \cos\left(\frac{N-1}{N}\pi\right)$ (i.e. $\alpha = 0$), then $p_0 = \varepsilon\left(1 - N - \frac{1}{\cos\left(\frac{N-1}{N}\pi\right)}\right) < 0$, meaning that (3.19) with constraints

$p_k > 0$, $k = 0, \ldots, N-1$ is not solvable for $\sigma_1 = 1$ and $\sigma_1 = \cos\left(\frac{N-1}{N}\pi\right)$. $\square$

3.3 Variational Analysis

In this section we come back to the main issue, namely existence of a solution for the primal problem 3.1.1. To this aim, we shall introduce a suitable set of "Lagrange multipliers" for our constrained optimization problem. Consider the linear map $A : \mathfrak{S}_{n+1} \times \mathfrak{S}_N \to \mathfrak{S}_N$ defined by

$$A(\Lambda, \Theta) = E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta, \quad (\Lambda, \Theta) \in \mathfrak{S}_{n+1} \times \mathfrak{S}_N$$

and define the set

$$\begin{aligned} \mathcal{L}_+ :=& \{(\Lambda, \Theta) \in (\mathfrak{S}_{n+1} \times \mathfrak{S}_N) \mid (\Lambda, \Theta) \in (\ker(A))^\perp, \\ & \left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) > 0\}. \end{aligned} \tag{3.24}$$

Observe that $\mathcal{L}_+$ is an open, convex subset of $(\ker(A))^\perp$. For each $(\Lambda, \Theta) \in \mathcal{L}_+$, we consider the unconstrained minimization of the *Lagrangian function*

$$\begin{aligned} L(\mathbf{\Sigma}_N, \Lambda, \Theta) :=& -\operatorname{tr} \log \mathbf{\Sigma}_N + \operatorname{tr}\left(\Lambda\left(E_n^\top \mathbf{\Sigma}_N E_n - \mathbf{T}_n\right)\right) \\ & + \operatorname{tr}\left(\Theta\left(\mathbf{U}_N^\top \mathbf{\Sigma}_N \mathbf{U}_N - \mathbf{\Sigma}_N\right)\right) \\ =& -\operatorname{tr} \log \mathbf{\Sigma}_N + \operatorname{tr}\left(E_n \Lambda E_n^\top \mathbf{\Sigma}_N\right) \\ & - \operatorname{tr}\left(\Lambda \mathbf{T}_n\right) + \operatorname{tr}\left(\mathbf{U}_N \Theta \mathbf{U}_N^\top \mathbf{\Sigma}_N\right) \\ & - \operatorname{tr}\left(\Theta \mathbf{\Sigma}_N\right) \end{aligned}$$

over $\mathfrak{S}_{N,+} := \{\mathbf{\Sigma}_N \in \mathfrak{S}_N,\ \mathbf{\Sigma}_N > 0\}$. For $\delta\mathbf{\Sigma}_N \in \mathfrak{S}_N$, we get

$$\begin{aligned} \delta L(\mathbf{\Sigma}_N, \Lambda, \Theta; \delta\mathbf{\Sigma}_N) =& -\operatorname{tr}\left(\mathbf{\Sigma}_N^{-1} \delta\mathbf{\Sigma}_N\right) + \operatorname{tr}\left(E_n \Lambda E_n^\top \delta\mathbf{\Sigma}_N\right) \\ & + \operatorname{tr}\left(\left(\mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) \delta\mathbf{\Sigma}_N\right). \end{aligned}$$

We conclude that $\delta L(\mathbf{\Sigma}_N, \Lambda, \Theta; \delta\mathbf{\Sigma}_N) = 0,\ \forall \delta\mathbf{\Sigma}_N \in \mathfrak{S}_N$ if and only if

$$\mathbf{\Sigma}_N^{-1} = E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta.$$

Thus, for each fixed pair $(\Lambda, \Theta) \in \mathcal{L}_+$, the unique $\mathbf{\Sigma}_N^o$ minimizing the Lagrangian is given by

$$\mathbf{\Sigma}_N^o = \left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right)^{-1}. \tag{3.25}$$

Consider next $L(\mathbf{\Sigma}_N^o, \Lambda, \Theta)$. We get

$$\begin{aligned} L(\mathbf{\Sigma}_N^o, \Lambda, \Theta) = & -\mathrm{tr}\log\left(\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right)^{-1}\right) \\ & + \mathrm{tr}\Big[\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) \\ & \left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right)^{-1}\Big] - \mathrm{tr}(\Lambda \mathbf{T}_n) \\ = & \ \mathrm{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) \\ & + \mathrm{tr} I_{mN} - \mathrm{tr}\left(\Lambda \mathbf{T}_n\right). \end{aligned}$$

This is a strictly concave function on $\mathcal{L}_+$ whose maximization is the *dual problem* of (CMEP). We can equivalently consider the convex problem

$$\min\left\{J(\Lambda, \Theta), (\Lambda, \Theta) \in \mathcal{L}_+\right\}, \tag{3.26}$$

where J (henceforth called dual function) is given by

$$J(\Lambda, \Theta) = \mathrm{tr}\left(\Lambda \mathbf{T}_n\right) - \mathrm{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right). \tag{3.27}$$

3.3.1 Existence for the dual problem

The minimization of the strictly convex function $J(\Lambda, \Theta)$ on the convex set $\mathcal{L}_+$ is a challenging problem as $\mathcal{L}_+$ is an *open* and *unbounded* subset of $(\ker(A))^\perp$.

In fact, situations like those displayed in the one–dimensional examples of Figure 3.1 may arise. Nevertheless, the following existence result in the Byrnes-Lindquist spirit, [23], [7], [20] can be established.

Theorem 3.3.1. *The function J admits a unique minimum point $(\bar{\Lambda}, \bar{\Theta})$ in $\mathcal{L}_+$.*

In order to prove this theorem, we need first to derive a number of auxiliary results. Let $\mathfrak{C}_N$ denote the vector subspace of block-circulant matrices in $\mathfrak{S}_N$. We proceed to characterize the orthogonal complement of $\mathfrak{C}_N$ in $\mathfrak{S}_N$.

Lemma 3.3.1. *Let $M \in \mathfrak{S}_N$. Then $M \in (\mathfrak{C}_N)^\perp$ if and only if it can be expressed as*

$$M = \mathbf{U}_N N \mathbf{U}_N^\top - N \tag{3.28}$$

for some $N \in \mathfrak{S}_N$.

Proof. By (3.1), $\mathfrak{C}_N$ is the kernel of the linear map from $\mathfrak{S}_N$ to $\mathfrak{S}_N$ given by $M \mapsto \mathbf{U}_N^\top M \mathbf{U}_N - M$. Hence, its orthogonal complement is the range of the adjoint map. Since

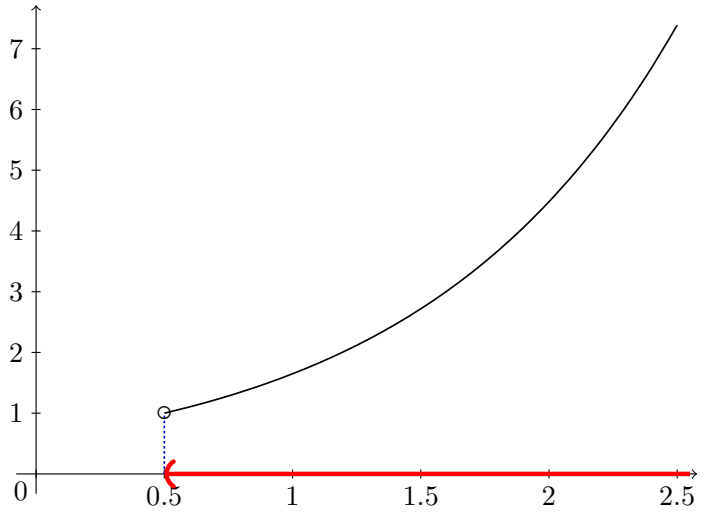
$$\begin{aligned} \operatorname{tr}\left((\mathbf{U}_N^\top M \mathbf{U}_N - M)N\right) &= \langle \mathbf{U}_N^\top M \mathbf{U}_N - M, N\rangle \\ &= \langle M, \mathbf{U}_N N \mathbf{U}_N^\top - N\rangle, \end{aligned}$$

the conclusion follows. □

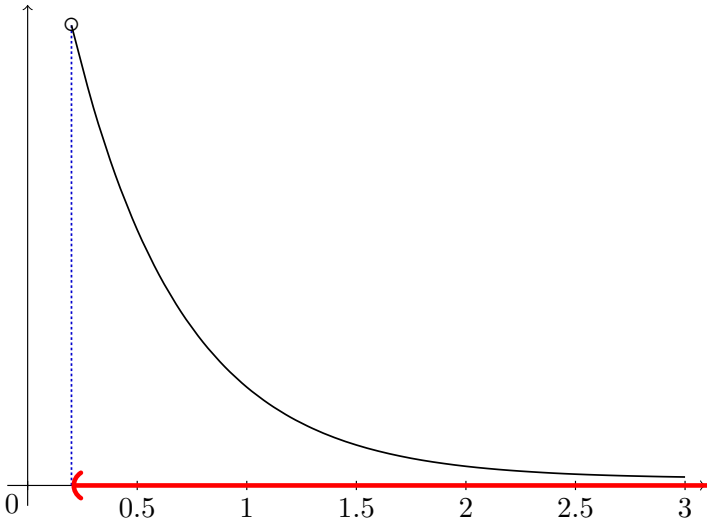
Next we show that, as expected, feasibility of the primal problem (CMEP) implies that the dual function J is bounded below.

Lemma 3.3.2. *Assume that there exists $\bar{\mathbf{\Sigma}}_N \in \mathfrak{S}_{N,+}$ satisfying (3.3b)-(3.3c). Then, for any pair $(\Lambda, \Theta) \in \mathcal{L}_+$, we have*

$$J(\Lambda, \Theta) \geq mN + \operatorname{tr} \log \bar{\mathbf{\Sigma}}_N. \tag{3.29}$$



(a)



(b)

Figure 3.1: One dimensional examples of bounded below strictly convex functions on an open and unbounded interval which do not have a minimum.

Proof. By (3.3b), we have $\mathrm{tr}(\Lambda \mathbf{T}_n) = \mathrm{tr}(\Lambda E_n^\top \bar{\mathbf{\Sigma}}_N E_n) = \mathrm{tr}(E_n \Lambda E_n^\top \bar{\mathbf{\Sigma}}_N)$. Using this fact and Lemma 3.3.1, we can now rewrite the dual function J as follows

$$\begin{aligned} J(\Lambda, \Theta) = &\ \mathrm{tr}\left(\Lambda \mathbf{T}_n\right) - \mathrm{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) \\ = &\ \mathrm{tr}\left[\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right) \bar{\mathbf{\Sigma}}_N\right] \\ &- \mathrm{tr}\log\left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right). \end{aligned}$$

Define $M(\Lambda, \Theta) = \left(E_n \Lambda E_n^\top + \mathbf{U}_N \Theta \mathbf{U}_N^\top - \Theta\right)$ which is positive definite for (Λ, Θ) in $\mathcal{L}_+$. Then

$$J(\Lambda, \Theta) = \mathrm{tr}\left(M(\Lambda, \Theta) \bar{\mathbf{\Sigma}}_N\right) - \mathrm{tr}\log M(\Lambda, \Theta).$$

As a function of M, this is a strictly convex function on $\mathfrak{S}_{N,+}$, whose unique minimum occurs at $M = \bar{\mathbf{\Sigma}}_N^{-1}$ where the minimum value is $\mathrm{tr}(I_{mN}) + \mathrm{tr}\log \bar{\mathbf{\Sigma}}_N$. □

Lemma 3.3.3. *Let $(\Lambda_k, \Theta_k), n \geq 1$ be a sequence of pairs in $\mathcal{L}_+$ such that $\|(\Lambda_k, \Theta_k)\| \to \infty$. Then also $\|A(\Lambda_k, \Theta_k)\| \to \infty$. It then follows that $\|(\Lambda_k, \Theta_k)\| \to \infty$ implies that $J(\Lambda_k, \Theta_k) \to \infty$.*

Proof. Notice that A is a linear operator between finite-dimensional linear spaces. Denote by σ_m the smallest singular value of the restriction of A to $(\ker A)^\perp$ (the orthogonal complement of $\ker A$). Clearly, $\sigma_m > 0$, so that, since each element of the sequence (Λ_k, Θ_k) is in $(\ker A)^\perp$, $\|A(\Lambda_k, \Theta_k)\| \geq \sigma_m \|(\Lambda_k, \Theta_k)\| \to \infty$.

Assume now that

$$\|A(\Lambda_k, \Theta_k)\| = \|\left(E_n \Lambda_k E_n^\top + \mathbf{U}_N \Theta_k \mathbf{U}_N^\top - \Theta_k\right)\| \to \infty.$$

Since these are all positive definite matrices and all matrix norms are equiv-

alent, it follows that

$$\operatorname{tr}\left(E_n\Lambda E_n^\top + \mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right) \to \infty.$$

As a consequence, $\operatorname{tr}\left(\left(E_n\Lambda E_n^\top + \mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right)\bar{\mathbf{\Sigma}}_N\right) \to \infty$ and, finally,

$$J(\Lambda_k, \Theta_k) \to \infty$$

. □

We show next that the dual function tends to infinity also when approaching the boundary of $\mathcal{L}_+$, namely

$$\begin{aligned}\partial\mathcal{L}_+ :=&\{(\Lambda,\Theta) \in (\mathfrak{S}_{n+1} \times \mathfrak{S}_N) | (\Lambda,\Theta) \in (\ker(A))^\perp,\\ &\left(E_n\Lambda E_n^\top + \mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right) \geq 0,\\ &\det\left(E_n\Lambda E_n^\top + \mathbf{U}_N\Theta\mathbf{U}_N^\top - \Theta\right) = 0\}.\end{aligned}$$

Lemma 3.3.4. *Consider a sequence $(\Lambda_k, \Theta_k), k \geq 1$ in $\mathcal{L}_+$ such that the matrix $\lim_k \left(E_n\Lambda_k E_n^\top + \mathbf{U}_N\Theta_k\mathbf{U}_N^\top - \Theta_k\right)$ is singular. Assume also that the sequence (Λ_k, Θ_k) is bounded. Then, $J(\Lambda_k, \Theta_k) \to \infty$.*

Proof. Simply write

$$J(\Lambda_k, \Theta_k) = -\log\det\left(E_n\Lambda_k E_n^\top + \mathbf{U}_N\Theta_k\mathbf{U}_N^\top - \Theta_k\right) + \operatorname{tr}(\Lambda_k\mathbf{T}_k).$$

Since $\operatorname{tr}(\Lambda_k\mathbf{T}_k)$ is bounded, the conclusion follows. □

Proof of Theorem 3.3.1. Observe that the function J is a continuous, bounded below (Lemma 3.3.2) function that tends to infinity both when $\|(\Lambda,\Theta)\|$ tends to infinity (Lemma 3.3.3) and when it tends to the boundary $\partial\mathcal{L}_+$ with $\|(\Lambda,\Theta)\|$ remaining bounded (Lemma 3.3.4). It follows that J is *inf-compact*

on $\mathcal{L}_+$, namely it has compact sublevel sets. By Weierstrass' Theorem[1], it admits at least one minimum point. Since J is strictly convex, the minimum point is unique. □

3.4 Bandedness Property

Let $(\bar{\Lambda}, \bar{\Theta})$ be the unique minimum point of J in $\mathcal{L}_+$ (Theorem 3.3.1). Then $\mathbf{\Sigma}_N^o \in \mathfrak{S}_{N,+}$ given by

$$\mathbf{\Sigma}_N^o = \left(E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta}\right)^{-1} \tag{3.30}$$

satisfies (3.3b) and (3.3c). Hence, it is the unique solution of the primal problem (CMEP). Since it satisfies (3.3c), $\mathbf{\Sigma}_N^o$ is in particular a block-circulant matrix and hence so is

$$(\mathbf{\Sigma}_N^o)^{-1} = \left(E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta}\right).$$

Let $\pi_{\mathfrak{C}_{N,s}}$ denote the orthogonal projection onto the linear subspace of symmetric, block-circulant matrices $\mathfrak{C}_{N,s}$. It follows that, in force of Lemma 3.3.1,

$$\begin{aligned}(\mathbf{\Sigma}_N^o)^{-1} &= \pi_{\mathfrak{C}_{N,s}}((\mathbf{\Sigma}_N^o)^{-1}) \\ &= \pi_{\mathfrak{C}_{N,s}}\left(E_n \bar{\Lambda} E_n^\top + \mathbf{U}_N \bar{\Theta} \mathbf{U}_N^\top - \bar{\Theta}\right) \\ &= \pi_{\mathfrak{C}_{N,s}}\left(E_n \bar{\Lambda} E_n^\top\right). \end{aligned} \tag{3.31}$$

Lemma 3.4.1 (**Bandedness property**). *Let $\mathbf{\Sigma}_N^o$ be the maximum Gaussian entropy covariance given by (3.30). Then $(\mathbf{\Sigma}_N^o)^{-1}$ is a symmetric block-circulant matrix which is banded of bandwidth n.*

[1] A continuous function on a compact set always achieves its maximum and minimum on that set.

Proof. Let

$$\Pi_{\bar{\Lambda}} := \pi_{\mathfrak{C}_{N,s}}\left(E_n\bar{\Lambda}E_n^\top\right) = \begin{bmatrix} \Pi_0 & \Pi_1^\top & \Pi_2^\top & \dots & \Pi_1 \\ \Pi_1 & \Pi_0 & \Pi_1^\top & \dots & \Pi_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \Pi_2^\top & \dots & \Pi_1 & \Pi_0 & \Pi_1^\top \\ \Pi_1^\top & \Pi_2^\top & \dots & \Pi_1 & \Pi_0 \end{bmatrix}$$

be the orthogonal projection of $\left(E_n\bar{\Lambda}E_n^\top\right)$ onto $\mathfrak{C}_{N,s}$. Since $\Pi_{\bar{\Lambda}}$ is symmetric and block-circulant, it is characterized by the orthogonality condition

$$\operatorname{tr}\left[\left(E_n\bar{\Lambda}E_n^\top - \Pi_{\bar{\Lambda}}\right)C\right] = \langle E_n\bar{\Lambda}E_n^\top - \Pi_{\bar{\Lambda}}, C\rangle = 0, \tag{3.32}$$

for all $C \in \mathfrak{C}_{N,s}$. Next observe that, if we write

$$C = \operatorname{Circ}\left[C_0, C_1^\top, C_2^\top, \dots, C_2, C_1\right]$$

and

$$\bar{\Lambda} = \begin{bmatrix} \bar{\Lambda}_{00} & \bar{\Lambda}_{01} & \dots & \bar{\Lambda}_{0n} \\ \bar{\Lambda}_{01}^\top & \bar{\Lambda}_{11} & \dots & \bar{\Lambda}_{1n} \\ \vdots & & \ddots & \vdots \\ \bar{\Lambda}_{0n}^\top & \bar{\Lambda}_{1n}^\top & \dots & \bar{\Lambda}_{nn} \end{bmatrix},$$

with $\bar{\Lambda}_{k,j} = \bar{\Lambda}_{j,k}^\top$, then

$$\begin{aligned} \operatorname{tr}\left[E_n\bar{\Lambda}E_n^\top C\right] &= \operatorname{tr}\left[\bar{\Lambda}E_n^\top CE_n\right] = \operatorname{tr}\left[(\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \ldots + \bar{\Lambda}_{nn})C_0\right. \\ &+ (\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \ldots + \bar{\Lambda}_{n-1,n})C_1 + \ldots + \bar{\Lambda}_{0n}C_n \\ &+ \left.(\bar{\Lambda}_{10} + \bar{\Lambda}_{21} + \ldots, \bar{\Lambda}_{n,n-1})C_1^\top + \ldots + \bar{\Lambda}_{n0}C_n^\top\right]. \end{aligned}$$

On the other hand, recalling that the product of two block-circulant matrices is block-circulant, we have that $\operatorname{tr}\left[\Pi_{\bar{\Lambda}}C\right]$ is simply N times the trace of the

first block row of $\Pi_{\bar{\Lambda}}$ times the first block column of C. We get

$$\begin{aligned}\operatorname{tr}\left[\Pi_{\bar{\Lambda}} C\right] = N \operatorname{tr}\Big[&\Pi_0 C_0 + \Pi_1^\top C_1 + \Pi_2^\top C_2 + \dots \\ &\dots + \Pi_2 C_2^\top + \Pi_1 C_1^\top\Big].\end{aligned}$$

Hence, the orthogonality condition (3.32), reads

$$\begin{aligned}\operatorname{tr}\left[\left(E_n \bar{\Lambda} E_n^\top - \Pi_{\bar{\Lambda}}\right) C\right] = \operatorname{tr}\Big[&\left((\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \dots + \bar{\Lambda}_{nn}) - N\Pi_0\right) C_0 + \\ &+ \left((\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \dots + \bar{\Lambda}_{n-1,n}) - N\Pi_1^\top\right) C_1 \\ &+ \left((\bar{\Lambda}_{10} + \bar{\Lambda}_{21} + \dots, \bar{\Lambda}_{n,n-1}) - N\Pi_1\right) C_1^\top \\ &+ \dots (\bar{\Lambda}_{0n} - N\Pi_1^\top) C_n + (\bar{\Lambda}_{n0} - N\Pi_1) C_n^\top)\Big] \\ &+ N\Pi_{n+1}^\top C_{n+1} + N\Pi_{n+1} C_{n+1}^\top \\ &+ N\Pi_{n+2}^\top C_{n+2} + N\Pi_{n+2} C_{n+2}^\top + \dots \\ = 0.\end{aligned}$$

Since this must hold true for all $C \in \mathfrak{C}_{N,s}$, we conclude that

$$\begin{aligned}\Pi_0 &= \frac{1}{N}\left(\bar{\Lambda}_{00} + \bar{\Lambda}_{11} + \dots + \bar{\Lambda}_{nn}\right), \\ \Pi_1 &= \frac{1}{N}(\bar{\Lambda}_{01} + \bar{\Lambda}_{12} + \dots + \bar{\Lambda}_{n-1,n})^\top, \\ &\vdots \\ \Pi_n &= \frac{1}{N}\bar{\Lambda}_{0n}^\top,\end{aligned}$$

while from the last equation we get $\Pi_i = 0$, forall i in the interval $n+1 \leq i \leq N-n-1$. From this it is clear that the inverse of the covariance matrix solving the primal problem (CMEP), namely $\Pi_{\bar{\Lambda}} = (\mathbf{\Sigma}_N^o)^{-1}$ has a circulant block-banded structure of bandwidth n. □

An immediate consequence of Lemma 3.4.1 is the following.

Theorem 3.4.1. *The CMEP solves the block–circulant band extension pro-*

blem 2.2.4. and hence (Theorem 2.2.1) the maximum likelihood identification problem for reciprocal processes (Problem 2.2.3).

As a further implication of Lemma 3.4.1, note that the solution of the CMEP matches the defining properties of the solution of the Dempster's covariance selection problem, since it agrees with the given partially specified block-circulant matrix in the given positions (central block band and NE and SW corners) and has the property to have a *banded* inverse. As the solution of the Dempster's covariance selection problem is unique (Proposition 2.2.1), the following result can be established.

Theorem 3.4.2. *If the given data lie on a block band symmetric with respect to the main diagonal and in the NE and SW corners and if they are consistent with a block-circulant symmetric structure, then the solution of the CMEP (Problem 2.2.7) and of the DMEP (Problem 2.2.6) coincide. In other words, the maximum entropy distribution, subject only to moment constraints (compatible with the circulant structure) on a block band and on the corners, is necessarily block-circulant, i.e. the underlying process is stationary.*

An alternative proof, as well as a generalization of this result is the object of the next Section.

3.5 Reconciliation with the covariance selection problem

In Theorem 3.4.2 it has been shown that if a *single block band* is specified the constraint that enforces the block-circulant structure when maximizing the determinant is automatically satisfied (see example in Figure 3.2) . In this Section we develop an independent proof of this result which in fact extends naturally to *any number of missing block-bands* as well as *arbitrary*

missing elements in a block-circulant structure (see examples in Figures 3.3a and 3.3b).

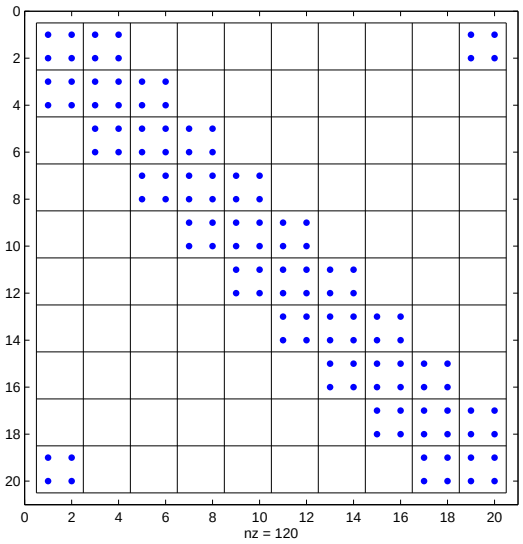
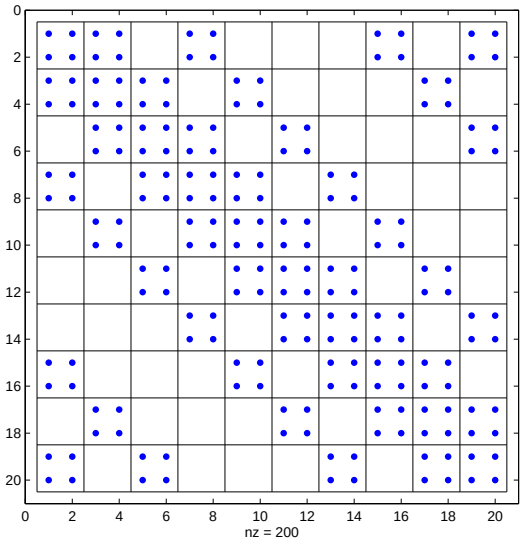
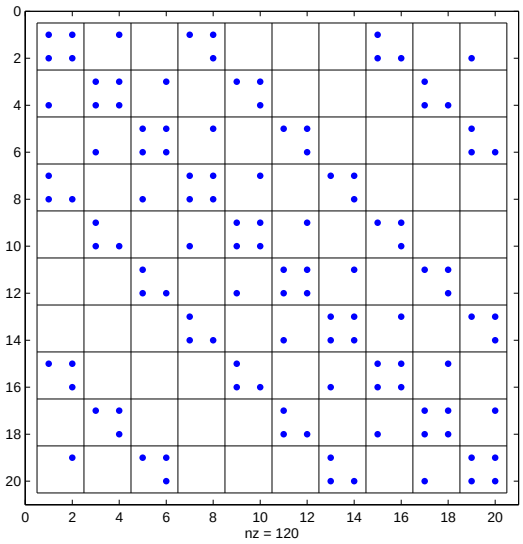


Figure 3.2: Example of sparsity pattern for which, by Theorem 3.4.2, the constraint that enforces the block-circulant structures when maximizing the determinant is automatically satisfied. The partially specified block–circulant matrix is built of 10×10 blocks, each of size 2. Each blue circle corresponds to a complex scalar entry. The black lines bounds different blocks.

We will show that this general fact is a direct consequence of the invariance of the determinant under the group of transformations that leave circulant matrices invariant. More specifically, let $\mathfrak{M}_N$ denote the set of all $N \times N$ m–block matrices over $\mathbb{C}$, $\mathfrak{H}_N \subset \mathfrak{M}_N$ the subset of all Hermitian matrices and $\mathfrak{H}_N^+ \subset \mathfrak{H}_N$ the cone of positive definite $N \times N$ m–block matrices. (not necessarily symmetric). Consider the pair $\mathcal{G} := (\{\text{conj},\ \text{shift}\}, \circ)$ where



(a) Example of given data lying on non consecutive block–bands



(b) Example of given data which in arbitrary missing positions.

Figure 3.3: Examples of sparsity patterns for which, according to the generalization in Theorem 3.5.1, the constraint that enforces the block-circulant structures is still automatically satisfied. In both the examples, the partially specified block–circulant matrix is built of 10×10 blocks, each of size 2.

conj and shift are the maps

$$\begin{aligned} \text{conj} : \mathfrak{M}_N &\longrightarrow \mathfrak{M}_N \\ M &\longmapsto M^* \\ \text{shift} : \mathfrak{M}_N &\longrightarrow \mathfrak{M}_N \\ M &\longmapsto \mathbf{U}_N^\top M \mathbf{U}_N \end{aligned}$$

and $\circ$ denotes the usual composition of maps. It is easy to check that the pair $\mathcal{G}$ is a *commutative group* and the following characterization of block–circulant Hermitian matrices holds.

Proposition 3.5.1. *Let $M \in \mathfrak{M}_N$. Then $M \in \mathfrak{C}_N \cap \mathfrak{H}_N$ if and only if the orbit of M under the action of $\mathcal{G}$ is M itself.*

That is, $M \in \mathfrak{C}_N \cap \mathfrak{H}_N$ if and only if M stays invariant under the action of the group $\mathcal{G}$ and the orbit contains no additional elements.

Before introducing our main result, let us define $\mathcal{I}_c$ to be a set of index-pairs (k, ℓ) *consistent with the block–circulant Hermitian structure*

$$(k, \ell) \in \mathcal{I}_c \Rightarrow (\ell, k) \in \mathcal{I}_c \tag{3.33a}$$

$$(k, \ell) \in \mathcal{I}_c \Rightarrow ((\ell + m)_{\bmod nm}, (k + m)_{\bmod nm}) \in \mathcal{I}_c. \tag{3.33b}$$

(for example one of the sets in Figure 3.3) and

$$\mathcal{R}_c := \{m_{k,\ell} \in \mathbb{C} \mid (k, \ell) \in \mathcal{I}_c\}$$

the set of $m_{k,\ell}$-values *consistent with the block–circulant Hermitian structure* i.e. the set of $m_{k,\ell}$-values such that

$$m_{k,\ell} = m_{\ell,k}^* \tag{3.34a}$$

$$m_{k,\ell} = m_{(\ell+m)_{\bmod nm},(k+m)_{\bmod nm}}^* \tag{3.34b}$$

for all pairs of indices.

Theorem 3.5.1. *Let $\mathcal{I}_c, \mathcal{R}_c$ be the sets of indices and corresponding values consistent with a Hermitian block–circulant structure defined by* (3.33) – (3.34) *and assume that there exists a positive completion (not necessarily circulant), and that this set is bounded. Then the solution of the DMEP 2.2.6 with $\mathcal{I} = \mathcal{I}_c$ and $\mathcal{R} = \mathcal{R}_c$, say $\mathbf{\Sigma}_N^o$, is automatically circulant.*

One rather direct proof can be based on the significance of the Lagrange multipliers as representing the sensitivity of the functional to be maximized (in this case the determinant) with respect to the corresponding constraints. Because the circulant structure dictates that all values linked via (3.33a) and (3.33b) affect the determinant in the same way (since $\det(\mathbf{U}_N^* M^* \mathbf{U}_N) = \det(M)$ and hence the value of the determinant is not affected by action of any $\mathcal{G}$-element), the sensitivity to each value $m_{k,\ell}$ is the same, and therefore the corresponding values for the Lagrange multipliers $\lambda_{k,\ell}$ at the stationary point (see equation (2.40)) are equal. Thus, $(\mathbf{\Sigma}_N^o)^{-1}$ in (2.41) has a circulant structure and so does $\mathbf{\Sigma}_N^o$ by Proposition A.0.3. An alternative and almost immediate proof is given below.

Proof. Once again, observe that for any $M \in \mathfrak{H}_N$ it holds that $\det(\text{shift}M) = \det(\text{conj}M) = \det(M)$ as neither the circulant block-shift nor the conjugation of Hermitian matrices changes the value of the determinant. Furthermore, observe that if M satisfies

$$e_k M e_\ell^* = m_{k,\ell}, \quad \text{for } (k,\ell) \in \mathcal{I}_c \text{ and } m_{k,\ell} \in \mathcal{R}_c, \tag{3.35}$$

then the same is true for shift$M$ as well as conj$M$. This is due to the fact that the constraints are consistent with the block-circulant-Hermitian structure as well. Now since $\det(\cdot)$ is strictly log-concave on $\mathfrak{H}_N^+$, it has a unique maximum subject to (3.35) (disregarding for the moment any restriction for

the maximizer to belong to $\mathfrak{C}_N$). But, this unique maximizing point $\mathbf{\Sigma}_N^o$ must be invariant under the group $\mathcal{G}$ generated by {conj and shift}, for otherwise, there would be multiple maxima. This proves directly that $\mathbf{\Sigma}_N^o$ is in $\mathfrak{C}_N$. □

Remark 3.5.1. The above argument applies to maximizers that may be restricted further by bounding individual elements, or in combination, to lie in a convex set in a way that is consistent with the circulant structure. More specifically and in a very general setting, if a maximizer exists over $\mathfrak{H}_N^+$ and if the constraints, of whatever nature, are consistent with the block–circulant structure, then the maximizer necessarily belongs to $\mathfrak{C}_N$. Thus, the essence of this result is a rather general invariance principle that the maximizer of a concave functional when restricted to points that individually remain invariant under the action of a certain group, it is identical to the unconstrained one —assuming that the domain of the functional is convex and invariant under the group.

3.6 Examples

This Section contains some illustrative examples of the theory developed so far. In the first two examples we compute the feasible set and the maximum entropy completion for different block and completion size as well as for different sparsity patterns. The third example aims to clarify the interconnection between feasibility and the completion size proved in Theorems 3.2.2, 3.2.4. Finally, the last example highlights some connections between the above developed theory and the factorization of certain polynomials in many variables, facilitated by the circulant structure.

Example 3.6.1. Let $N = 7$, $m = 1$ and consider the partially specified

matrix

$$\mathbf{\Sigma}_N = \begin{bmatrix} 2 & 1 & x & y & y & x & 1 \\ 1 & 2 & 1 & x & y & y & x \\ x & 1 & 2 & 1 & x & y & y \\ y & x & 1 & 2 & 1 & x & y \\ y & y & x & 1 & 2 & 1 & x \\ x & y & y & x & 1 & 2 & 1 \\ 1 & x & y & y & x & 1 & 2 \end{bmatrix}$$

where those in red are the given entries, while x and y stands for the entries which has been left unspecified. The eigenvalues of $\mathbf{\Sigma}_N$ are

$$a(w^0) = 2(2 + x + y) \tag{3.36a}$$

$$a(w^1) = 2 - 2y\text{Cos}\left[\frac{\pi}{7}\right] - 2x\text{Sin}\left[\frac{\pi}{14}\right] + 2\text{Sin}\left[\frac{3\pi}{14}\right] \tag{3.36b}$$

$$a(w^2) = -2\left(-1 + x\text{Cos}\left[\frac{\pi}{7}\right] + \text{Sin}\left[\frac{\pi}{14}\right] - y\text{Sin}\left[\frac{3\pi}{14}\right]\right) \tag{3.36c}$$

$$a(w^3) = -2\left(-1 + \text{Cos}\left[\frac{\pi}{7}\right] + y\text{Sin}\left[\frac{\pi}{14}\right] - x\text{Sin}\left[\frac{3\pi}{14}\right]\right) \tag{3.36d}$$

$$a(w^4) = a(w^4),\ a(w^5) = a(w^2),\ a(w^6) = a(w^1), \tag{3.36e}$$

and the feasible set is the interior of the polyhedron shown in Figure 3.4 as the intersection of half–planes. The maximum entropy completion $\mathbf{\Sigma}_N^o$ has $x = 0.5321$ and $y = 0.3473$ (we will see how to compute it in Chapter 4). Its

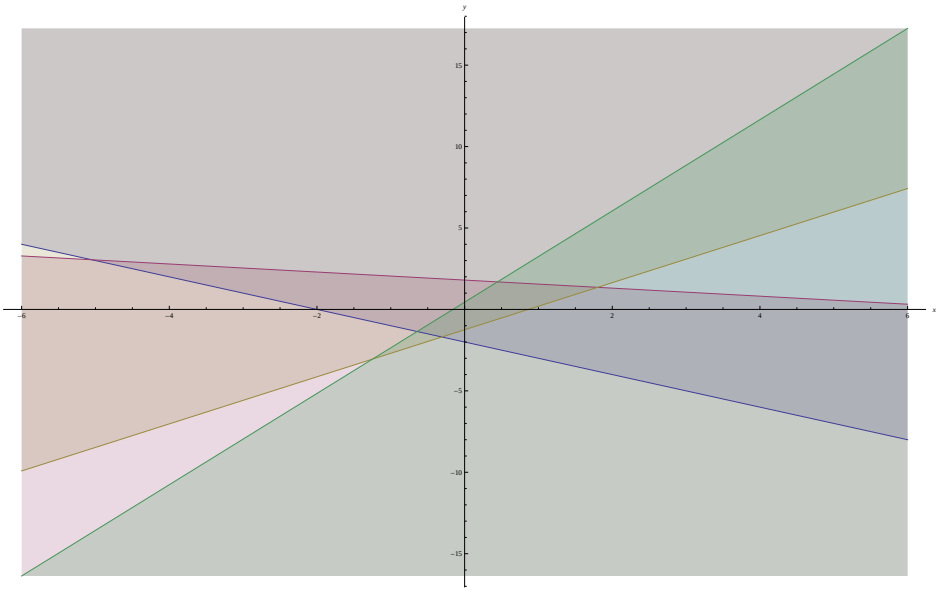
inverse, say $\mathbf{M}_N^o$, is

$$\mathbf{M}_N^o = \begin{bmatrix} 0.8264 & -0.3264 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3264 \\ -0.3264 & 0.8264 & -0.3264 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & -0.3264 & 0.8264 & -0.3264 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & -0.3264 & 0.8264 & -0.3264 & 0.0000 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & -0.3264 & 0.8264 & -0.3264 & 0.0000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3264 & 0.8264 & -0.3264 \\ -0.3264 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & -0.3264 & 0.8264 \end{bmatrix}$$

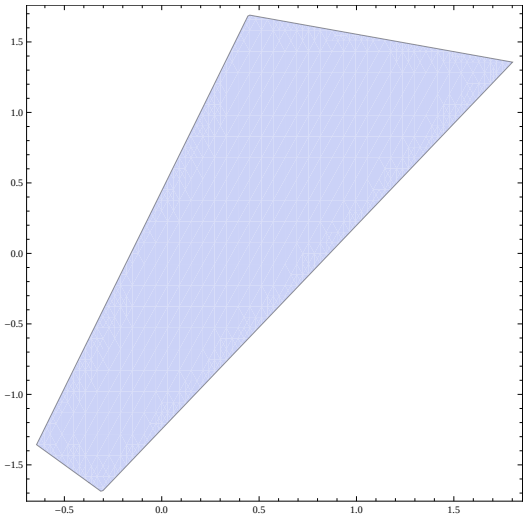
i.e. it has the claimed zero pattern (see 3.4.1).

Example 3.6.2. Let $N = 4$, $m = 2$. In red are the given data, specified in such a way to be consistent with a block-circulant structure. The third block has been left completely unspecified (x, y, z stands for unspecified entries).

$$\mathbf{\Sigma}_N = \left[\begin{array}{cc|cc|cc|cc} 2 & \frac{1}{2} & 1 & 1 & x & y & 1 & 0 \\ \frac{1}{2} & 2 & 0 & 1 & y & z & 1 & 1 \\ \hline 1 & 0 & 2 & \frac{1}{2} & 1 & 1 & x & y \\ 1 & 1 & \frac{1}{2} & 2 & 0 & 1 & y & z \\ \hline x & y & 1 & 0 & 2 & \frac{1}{2} & 1 & 1 \\ y & z & 1 & 1 & \frac{1}{2} & 2 & 0 & 1 \\ \hline 1 & 1 & x & y & 1 & 0 & 2 & \frac{1}{2} \\ 0 & 1 & y & z & 1 & 1 & \frac{1}{2} & 2 \end{array}\right]$$



(a)



(b)

Figure 3.4: Feasible polyhedral set as the intersection of half–planes for $\mathbf{\Sigma}_N = \text{Circ}\{2, 1, x, y, y, x, 1\}$.

The polynomial matrices $a(w^k)$, $k = 0, 1, 2, 3$ results

$$\Psi(w^0) = \begin{bmatrix} 4+x & \frac{3}{2}+y \\ \frac{3}{2}+y & 4+z \end{bmatrix},$$
$$\Psi(w^1) = \begin{bmatrix} 2-x & \left(\frac{1}{2}-i\right)-y \\ \left(\frac{1}{2}+i\right)-y & 2-z \end{bmatrix} = \Psi(w^3)^\top,$$
$$\Psi(w^2) = \begin{bmatrix} x & -\frac{1}{2}+y \\ -\frac{1}{2}+y & z \end{bmatrix}.$$

The respective eigenvalues are

$$\mathrm{Eig}\left\{\Psi(w^0)\right\} = \frac{1}{2}\left(8+x+z \pm \sqrt{9+x^2+12y+4y^2-2xz+z^2}\right)$$
$$\mathrm{Eig}\left\{\Psi(w^1)\right\} = \frac{1}{2}\left(4-x-z \pm \sqrt{5+x^2-4y+4y^2-2xz+z^2}\right)$$
$$\mathrm{Eig}\left\{\Psi(w^2)\right\} = \frac{1}{2}\left(x+z \pm \sqrt{1+x^2-4y+4y^2-2xz+z^2}\right).$$

and the set where they all are positive is shown in Figure 3.5. The maximum entropy completion $\mathbf{\Sigma}_N^o$ has $x = z = 0.4853$, $y = 0.4789$ and its inverse $\mathbf{M}_N^o$, has blocks

$$M_0 = \begin{bmatrix} 1.1707 & -0.0163 \\ -0.0163 & 1.1707 \end{bmatrix}, \qquad M_1 = \begin{bmatrix} -0.4469 & -0.4394 \\ 0.3335 & -0.4469 \end{bmatrix},$$
$$M_2 = \begin{bmatrix} 0.0000 & 0.0000 \\ 0.0000 & 0.0000 \end{bmatrix}, \qquad M_3 = M_1^\top,$$

where M_2 is the 2×2 zero matrix, as claimed. If we set $z = 1$ and leave the entries x and y in Σ_2 unspecified, the feasible region is the slice of the set shown in Figure 3.5 that corresponds to $z = 1$ and is bounded by parabolas. These are shown in Figure 3.6 along with the feasible set. The completion with maximal determinant corresponds to $x = 0.3548$ and $y = 0.4813$ and

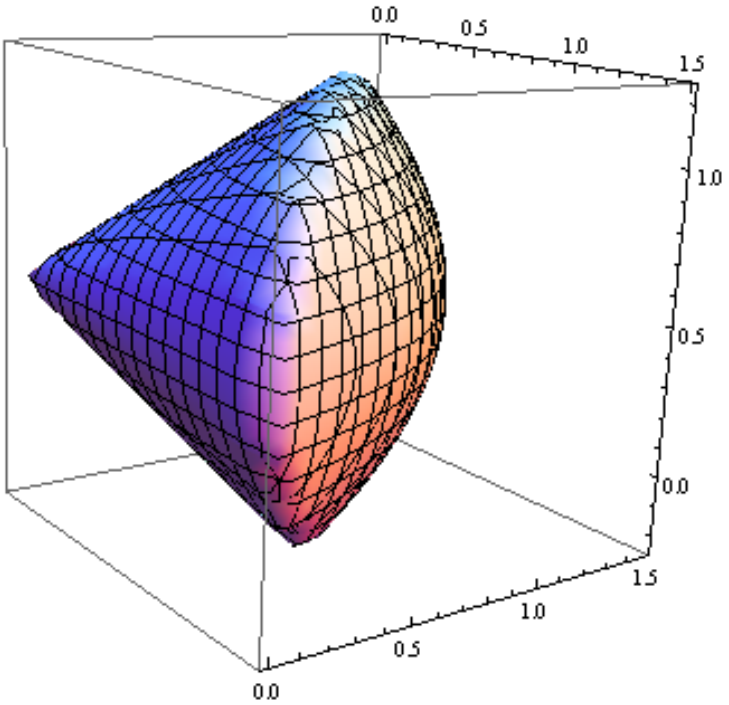
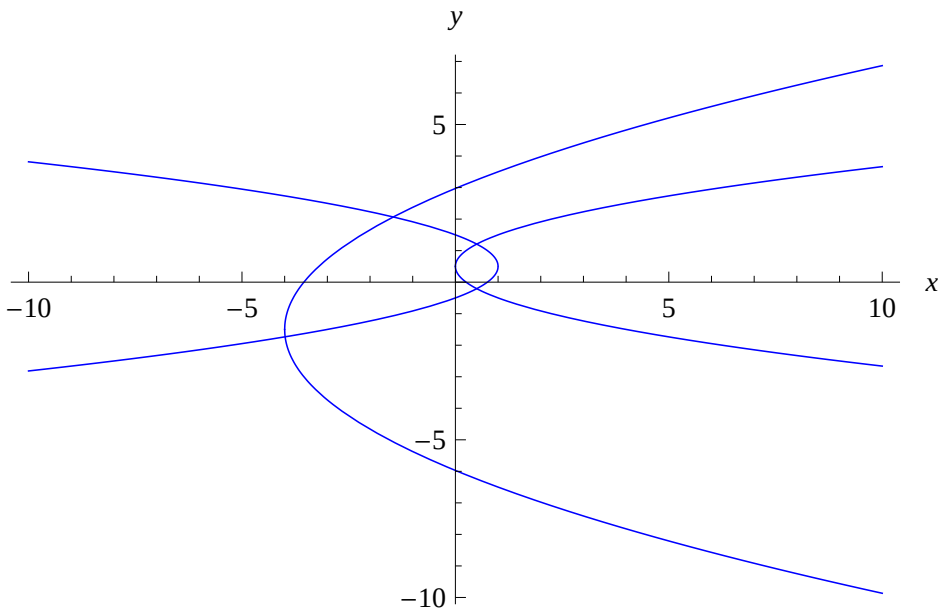
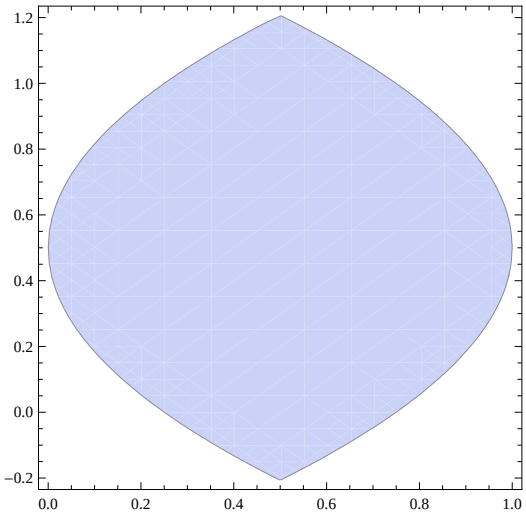


Figure 3.5: Feasible set $\{(x, y, z) \mid \mathbf{\Sigma}_N \geq 0\}$ for $\mathbf{\Sigma}_N = \text{Circ}\left\{\Sigma_0, \Sigma_1^\top, \Sigma_2, \Sigma_1\right\}$ with $m = 2$, Σ_2 unspecified.



(a)



(b)

Figure 3.6: Curves delineating the feasible set $\{(x, y) \mid \mathbf{\Sigma}_N \geq 0\}$ for $z = 1$ along with their intersection.

its inverse is $\mathbf{M}_N^o$ with

$$M_0 = \begin{bmatrix} 1.5507 & -0.0291 \\ -0.0291 & 1.5869 \end{bmatrix}, \qquad M_1 = \begin{bmatrix} -0.6353 & -0.8163 \\ 0.7344 & -0.1893 \end{bmatrix},$$
$$M_2 = \begin{bmatrix} 0.0000 & 0.0000 \\ 0.0000 & -0.9644 \end{bmatrix}, \qquad M_3 = M_1^\top.$$

once again with zeros in the complementary positions of those fixed in $\mathbf{\Sigma}_N$.

Example 3.6.3. This example serves to stress the interconnection between feasibility and the size of the completion N which has been proved in Theorems 3.2.2, 3.2.4, exploting the informations on the shape of the feasible region gained with Theorem 3.2.1. Let

$$\mathbf{T}_2 = \begin{bmatrix} 1 & -0.91 \\ -0.91 & 1 \end{bmatrix}.$$

We want to investigate the feasibility of Problem 3.1.1 for $N = 7$ and $N = 9$. Since

$$\cos\left\{\frac{(N-1)}{N}\pi\right\} = \begin{cases} -0.9010 & \text{for } N = 7 \\ -0.9397 & \text{for } N = 9 \end{cases}$$

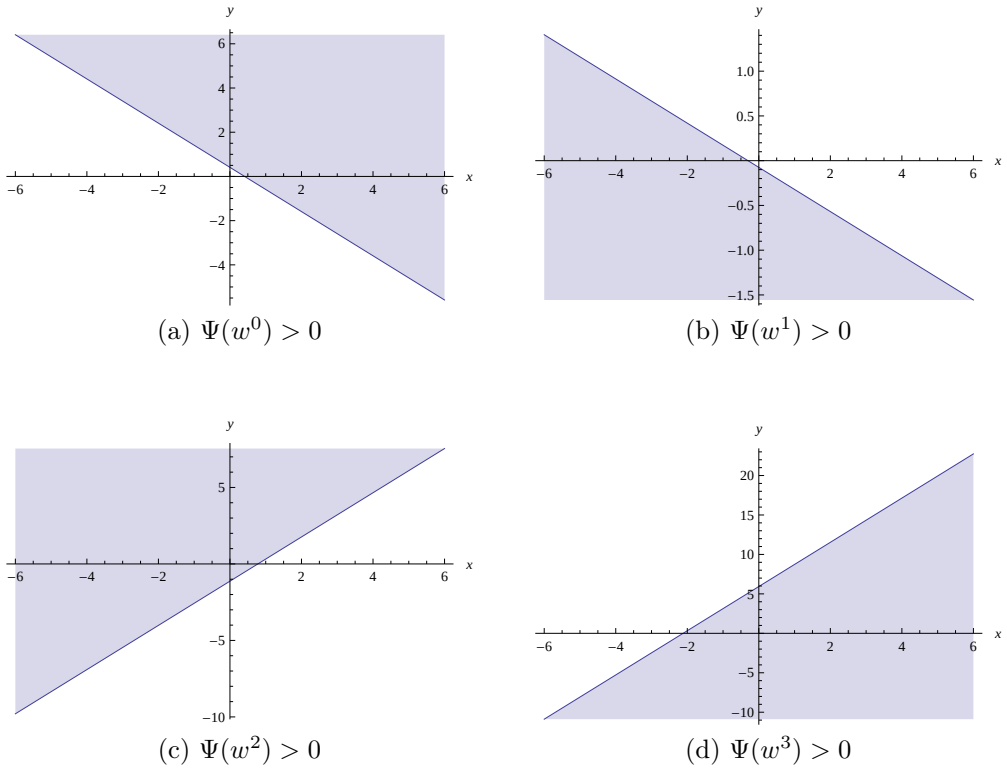
we expect that for $N = 7$ the problem is unfeasible while for $N \geq 9$ it is expected to become feasible. Indeed, this is the case, since for $N = 7$ the eigenvalues results

$$\Psi(w^0) = -0.82 + 2x + 2y$$
$$\Psi(w^1) = \Psi(w^6) = -0.134751 - 0.445042x - 1.80194y$$
$$\Psi(w^2) = \Psi(w^5) = 1.40499 - 1.80194x + 1.24698y$$
$$\Psi(w^3) = \Psi(w^4) = 2.63976 + 1.24698x - 0.445042y.$$

and the feasible set is empty, being the intersection of the four half–planes represented in Figure 3.7. On the other hand, if $N = 9$, the eigenvalues are

$$\begin{aligned}
\Psi(w^0) &= -0.82 + 2x + 2y + 2z \\
\Psi(w^1) = \Psi(w^8) &= -0.394201 + 0.347296x - y - 1.87939z \\
\Psi(w^2) = \Psi(w^7) &= 0.68396 - 1.87939x - y + 1.53209z \\
\Psi(w^3) = \Psi(w^6) &= 1.91 - x + 2y - z \\
\Psi(w^4) = \Psi(w^5) &= 2.71024 + 1.53209x - y + 0.347296z
\end{aligned}$$

and the feasible set is the nonempty region shown in Figure 3.8.



(a) $\Psi(w^0) > 0$ (b) $\Psi(w^1) > 0$

(c) $\Psi(w^2) > 0$ (d) $\Psi(w^3) > 0$

Figure 3.7: Half–planes representing the regions where the eigenvalues of the partially specified circulant matrix Circ $\{1, -0.91, x, y, y, x, -0.91\}$ are positive.

Example 3.6.4. This example highlights a new facet of the theory developed

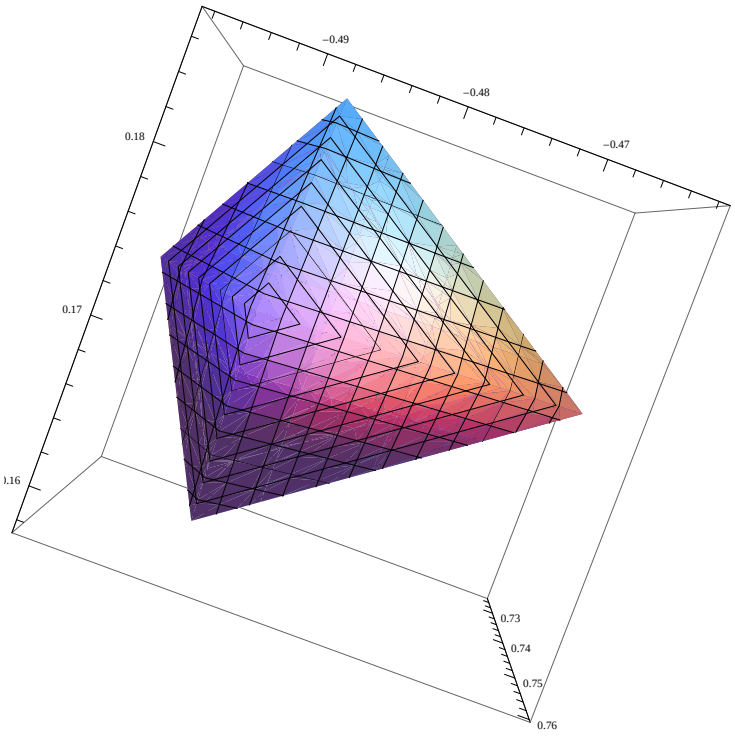


Figure 3.8: Feasible region $\{(x, y, z) \mid \mathbf{\Sigma}_N \geq 0\}$ for $\mathbf{\Sigma}_N = \text{Circ}\{1, -0.91, x, y, z, z, y, x, -0.91\}$.

so far and concerns the easiness by which polynomials in several variables that happen to coincide with the determinant of a partially specified circulant matrix can be factored via a Fourier transformation (an otherwise difficult task due to the irrationality of the factors in the absence of a suitable field extension). To this aim, let us come back to the partially specified matrix of Example 3.6.1. The determinant of $\mathbf{\Sigma}_N$ is a polynomial of degree 7,

$$\begin{aligned}\det(\mathbf{\Sigma}_N) &= 4 + 42x + 56x^2 - 294x^3 + 140x^4 + 84x^5 - 28x^6 \\ &+ 2x^7 - 14y - 28xy + 350x^2y - 196x^3y - 112x^4y - 84x^5y \\ &+ 14x^6y - 168xy^2 + 56x^2y^2 + 238x^3y^2 + 112x^4y^2 + 14x^5y^2 \\ &+ 28y^3 - 238x^2y^3 - 28x^3y^3 - 42x^4y^3 + 98xy^4 - 14y^5 \\ &+ 28x^2y^5 - 14xy^6 + 2y^7 \end{aligned} \tag{3.37}$$

in x and y. Over the ring of polynomials with rational coefficients it factors as (e.g., using Matlab or Mathematica)

$$\begin{aligned}\det(\mathbf{\Sigma}_N) = 2(2+x+y)\big(&1 + 5x - 8x^2 + x^3 - 2y + 5xy \\ &+3x^2y - y^2 - 4xy^2 + y^3\big)^2 .\end{aligned}$$

However, using (3.36a-3.36e), we already know that

$$\begin{aligned}\det(\mathbf{\Sigma}_N) = 2(2+x+y)&\times \\ &\left[2 - 2y\cos\left(\frac{\pi}{7}\right) - 2x\sin\left(\frac{\pi}{14}\right) + 2\sin\left(\frac{3\pi}{14}\right)\right]^2 \\ &\left[-2\left(-1 + x\cos\left(\frac{\pi}{7}\right) + \sin\left(\frac{\pi}{14}\right) - y\sin\left(\frac{3\pi}{14}\right)\right)\right]^2 \\ &\left[-2\left(-1 + \cos\left(\frac{\pi}{7}\right) + y\sin\left(\frac{\pi}{14}\right) - x\sin\left(\frac{3\pi}{14}\right)\right)\right]^2 .\end{aligned} \tag{3.38}$$

Provided we know a suitable field extension of $\mathbb{Q}$ which contains the coeffi-

cients of the factors, i.e., $\mathbb{Q}[\cos\left(\frac{\pi}{7}\right), \sin\left(\frac{\pi}{14}\right), \text{etc.}]$, the factorization can be carried out with standard methods [8]. Finding such an extension, in general, is a challenging problem. Of course, expressing a given rational polynomial as the determinant of a circulant matrix with rational coefficients may an equally challenging one, in general. Yet, we hope that the above observations may provide alternative ways to factor polynomials in certain suitable cases.

3.7 Final Remarks

Remark 3.7.1. Since the beginning of of this Chapter, we have been dealing only with Gaussian distributions in order to facilitate the comparison with Dempster's classical results. Now we show that the Gaussian assumption can be dispensed with, and our solution is indeed optimal in the larger family of (zero-mean) second-order distributions.

Theorem 3.7.1. *The Gaussian distribution with (zero mean and) covariance* $\mathbf{\Sigma}_N^o$ *defined by (3.30) maximizes the entropy functional* (2.37) *over the set of all (zero mean) probability densities whose covariance matrix satisfies the boundary conditions* (3.3b), (3.3c).

Proof. Let $\mathfrak{C}_N(\mathbf{T}_n)$ be the set of (block-circulant) covariance matrices satisfying the boundary conditions (3.3b), (3.3c) and let $p_{\mathbf{\Sigma}}$ be a probability density with zero mean and covariance $\mathbf{\Sigma}$. In particular, we shall denote by $g_{\mathbf{\Sigma}}$ the Gaussian density with zero mean and covariance $\mathbf{\Sigma}$. Now, by a famous theorem of Shannon [45], the probability distribution having maximum entropy in the class of all distribution with a fixed mean vector (which we take equal to zero) and variance matrix $\mathbf{\Sigma}$, is the Gaussian distribution $g_{\mathbf{\Sigma}}$. Hence:

$$\max_{\mathbf{\Sigma}\in\mathfrak{C}_N(\mathbf{T}_n)} \left\{ \max_{p_{\mathbf{\Sigma}}} \left[H(p_{\mathbf{\Sigma}}) \right] \right\} = \max_{\mathbf{\Sigma}\in\mathfrak{C}_N(\mathbf{T}_n)} \left\{ H(g_{\mathbf{\Sigma}}) \right\}$$

where the maximum in the right-hand side is attained by $g_{\mathbf{\Sigma}_N^o}$. □

Remark 3.7.2. For a moment, let us consider what would have happened if, instead of requiring the completed matrix to be circulant, we had added to the DMEP the constraint requiring the completed matrix to be Toeplitz (we will refer to this modified version of the problem as the TMEP). In this case, it would not have been true that in any case, i.e. for any number of missing elements in the Toeplitz symmetric structure, the completion would have maintained the inverses's zero–pattern property. In fact, this holds only if the given data lie on consecutive bands centred along the main diagonal.

Example 3.7.1 (TMEP)**.** In red: given pattern (oviously consistent with a Toeplitz symmetric structure). The completed matrix is forced to be Toeplitz. The given entries lie on consecutive bands. The maximum entropy completion results

$$\mathbf{\Sigma}_N^o = \begin{bmatrix} 1.0000 & 0.7000 & 0.4900 & 0.3430 & 0.2401 \\ 0.7000 & 1.0000 & 0.7000 & 0.4900 & 0.3430 \\ 0.4900 & 0.7000 & 1.0000 & 0.7000 & 0.4900 \\ 0.3430 & 0.4900 & 0.7000 & 1.0000 & 0.7000 \\ 0.2401 & 0.3430 & 0.4900 & 0.7000 & 1.0000 \end{bmatrix}$$

and its inverse still has zeros in the complementary positions of those of the data

$$(\mathbf{\Sigma}_N^o)^{-1} = \begin{bmatrix} 1.9608 & -1.3725 & 0.0000 & 0.0000 & 0.0000 \\ -1.3725 & 2.9216 & -1.3725 & 0.0000 & 0.0000 \\ 0.0000 & -1.3725 & 2.9216 & -1.3725 & 0.0000 \\ 0.0000 & 0.0000 & -1.3725 & 2.9216 & -1.3725 \\ 0.0000 & 0.0000 & 0.0000 & -1.3725 & 1.9608 \end{bmatrix}.$$

On the other end, if the given entries do not lie on consecutive bands, then

the maximum entropy completion is given by

$$\mathbf{\Sigma}_N^o = \begin{bmatrix} 1.0000 & 0.7000 & 0.3890 & 0.1000 & -0.0636 \\ 0.7000 & 1.0000 & 0.7000 & 0.3890 & 0.1000 \\ 0.3890 & 0.7000 & 1.0000 & 0.7000 & 0.3890 \\ 0.1000 & 0.3890 & 0.7000 & 1.0000 & 0.7000 \\ -0.0636 & 0.1000 & 0.3890 & 0.7000 & 1.0000 \end{bmatrix}$$

and its *inverse does not have zeros in the complementary positions of those assigned in* $\mathbf{\Sigma}_N$

$$(\mathbf{\Sigma}_N^o)^{-1} = \begin{bmatrix} 2.1082 & -1.6932 & 0.1010 & 0.3772 & 0.0000 \\ -1.6932 & 3.4681 & -1.7743 & -0.2020 & 0.3772 \\ 0.1010 & -1.7743 & 3.4055 & -1.7743 & 0.1010 \\ 0.3772 & -0.2020 & -1.7743 & 3.4681 & -1.6932 \\ 0.0000 & 0.3772 & 0.1010 & -1.6932 & 2.1082 \end{bmatrix}.$$

Moreover, if instead of adding some more constraints to the DMEP, we had simply provided it with data consistent with a Toeplitz structure, it would not have been true that the completion would have been automatically Toeplitz, this property once again being satisfied only if the data lie on a single band centred along the main diagonal (see [18]).

Example 3.7.2 (DMEP with Toeplitz data)**.** In red: given data consistent with a Toeplitz symmetric structure. If the specified elements lie *on consecutive bands*, the maximum entropy completion *is still Toeplitz*.

$$\mathbf{\Sigma}_N^o = \begin{bmatrix} 1.0000 & 0.7000 & 0.4900 & 0.3430 & 0.2401 \\ 0.7000 & 1.0000 & 0.7000 & 0.4900 & 0.3430 \\ 0.4900 & 0.7000 & 1.0000 & 0.7000 & 0.4900 \\ 0.3430 & 0.4900 & 0.7000 & 1.0000 & 0.7000 \\ 0.2401 & 0.3430 & 0.4900 & 0.7000 & 1.0000 \end{bmatrix}$$

while it fails to be Toeplitz if the given data lie on *non consecutive bands*

$$\mathbf{\Sigma}_N^o = \begin{bmatrix} 1.0000 & 0.7000 & 0.4124 & 0.1000 & -0.0447 \\ 0.7000 & 1.0000 & 0.7000 & 0.3580 & 0.1000 \\ 0.4124 & 0.7000 & 1.0000 & 0.7000 & 0.4124 \\ 0.1000 & 0.3580 & 0.7000 & 1.0000 & 0.7000 \\ -0.0447 & 0.1000 & 0.4124 & 0.7000 & 1.0000 \end{bmatrix}$$

These observations contributes to highlight the nontriviality of the results of this Chapter.

Remark 3.7.3. The theory developed so far can be interpreted as a particular covariance selection result in the vein of Dempster's paper; compare in particular [16, Proposition a]. In fact the results of this Section substantiate also the maximum entropy principle of Dempster (Proposition 2.2.1). It is however important to note that none of our results follows as a particular case from Dempster's results, since [16] deals with a very unstructured setting. In particular our main result (Theorem 3.4.1) that the solution, $\mathbf{\Sigma}_N^o$, to our primal problem (CMEP) has a block-circulant *banded* inverse, is completely original. Its proof uses in an essential way the characterization of the CMEP solution provided by our variational analysis and cleverly exploits the block-circulant structure.

Remark 3.7.4. Because of the equivalence of reciprocal AR modeling and the underlying process covariance having an inverse with a banded structure, explained in Section 2.1.1, we see that the maximum entropy principle leads in fact to (reciprocal) AR models. This makes contact with the ever-present problem in control an signal processing of (approximate) AR modeling from finite covariance data, whose solution dates back to the work of N. Levinson and P. Whittle. That AR modeling from finite covariance data is actually

equivalent to a positive band extension problems for infinite Toeplitz matrices has been realized and studied in the past decades by Dym, Gohberg and co-workers, see e.g. [18], [24] as representative references of a very large literature. We should stress here that band extension problems for infinite Toeplitz matrices are invariably attacked and solved by factorization techniques, but circulant matrices do not fit in the "banded algebra" framework used in the literature. Also, one should note that the maximum entropy property is usually presented in the literature as a final embellishment of a solution which was already obtained by factorization techniques. Here, for the circulant band extension problem, factorization techniques do not work and the maximum entropy principle turns out to be the key to the solution of the problem.

Finally, we anticipate that the results of this Chapter lead to an efficient iterative algorithm for the explicit solution of the CMEP which is guaranteed to converge to a unique minimum. This solves the variational problem and hence the circulant band extension problem which subsumes maximum likelihood identification of reciprocal processes. This algorithm is the object of the next Chapter.

Chapter 4

Algorithms for the Block-Circulant Band Extension Problem

This Chapter deals with efficient algorithms for solving the CMEP. Since it has been shown that the solutions of the CMEP and of the DMEP with circulant data coincide, all the methods available in the literature for the DMEP can be employed. If the graph associated with the specified entries is chordal, the solution of the DMEP can be expressed in closed form in terms of the principal minors of the sample covariance matrix. However, the sparsity pattern associated with the given entries in our problem is not chordal. For non-chordal graphs the maximum entropy completion has to be computed iteratively. A straightforward application of standard optimization algorithms is too expensive for large size problems, and several specialized algorithms have been proposed in the literature which deals with a general, very unstructured, setting. In this Chapter we propose a modified matricial gradient descent algorithm for the solution of the CMEP which naturally follows from the variational analysis of the previous Chapter and exploits in an essential way the circulant structure of our problem. This algorithm compares very well with the algorithms proposed in the literature for the solution of the DMEP. The outline of the Chapter is as follows. Section 4.1 briefly

reviews some of the most popular methods for the solution of the DMEP. In Section 4.2 the matricial gradient descent algorithm is introduced. An experimental comparison between the proposed algorithm and the methods available in the literature is presented in Section 4.3.

4.1 Algorithms for the covariance selection problem

Before discussing some of the algorithms proposed in the literature for the covariance selection problem, a brief digression into some basics of graph theory is needed. First of all, it is natural to describe the pattern of the specified entries of an $Nm \times Nm$ partial symmetric matrix $H = (h_{ij})$ by an undirected graph (because h_{ij} is specified when h_{ji} is) of Nm vertices which has an edge joining vertex i and vertex j if and only if the entry h_{ij} is specified. Because the diagonal entries are all assumed specified, we ignore loops at the vertices. The undirected graph will be denoted by $\mathcal{G} = (V, E)$ where V is the vertex set and E the edge set which consists of unordered pairs of distinct vertices. In any undirected graph we say that 2 vertices u, $v \in V$ are *adjacent* if $(u, v) \in E$ (i.e. if they are joined by an edge). For any vertex set $S \subseteq V$, consider the edge set $E(S) \subseteq E$ given by

$$E(S) := \{(u, v) \in E \mid u, v \in S\}$$

We say that $\mathcal{G}(S)$ is the *subgraph of* $\mathcal{G}$ *induced by* S if $\mathcal{G}(S) = (S, E(S))$. An induced subgraph $\mathcal{G}(S)$ is *complete* if the vertices in S are pairwise adjacent in $\mathcal{G}$. In this case we say that S *is complete in* $\mathcal{G}$.

Definition 4.1.1. *A* clique *is a complete subgraph that is not contained within another complete subgraph.*

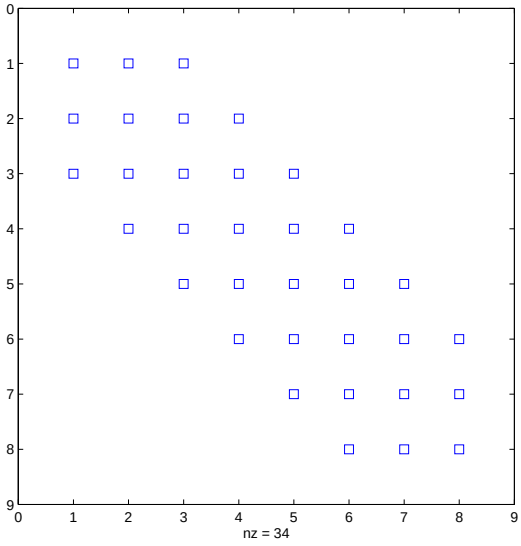
We let $[v_0, v_1, \ldots, v_k]$ denote a *simple path* of length k from v_0 to v_k in

$\mathcal{G}$, i.e. $v_i \neq v_j$ for $i \neq j$ and $(v_i, v_{i+1}) \in E$ for $0 \leq i \leq k-1$. Similarly $[v_0, v_1, \ldots, v_k, v_0]$ denotes a *simple cycle* of length $k+1$ in $\mathcal{G}$. Finally, a *chord* of a path (cycle) is any edge joining two nonconsecutive vertices of the path (cycle).

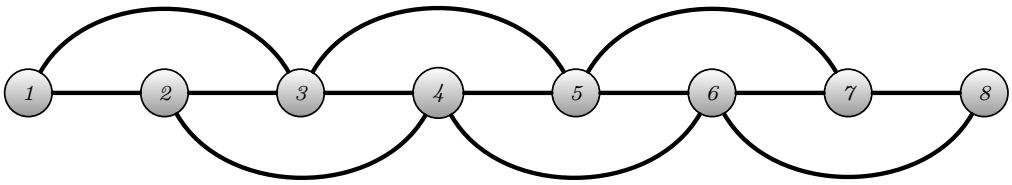
Definition 4.1.2. *An undirected graph is* chordal *(triangulated, rigid circuit) if every cycle of length greater than three has a chord.*

Example 4.1.1. A first example of chordal sparsity pattern is the banded sparsity pattern of Figure 4.1a. The associated graph is shown in Figure 4.1b which may be drawn also as shown in Figure 4.1c from which it is immediate to see that it is chordal. Another example of chordal sparsity pattern is the arrow pattern shown in Figure 4.2 along with its associated graph.

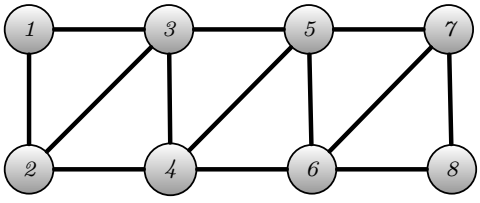
As anticipated in the introduction, if the graph of the specified entries is *chordal* ([4]), the maximum determinant matrix completion problem admits a closed form solution in terms of the principal minors of the sample covariance matrix (see [2], [35], [22], [40]). However, the graph associated with banded circulant sparsity patterns we are dealing with is not chordal, as it is apparent from the examples in Figures 4.3 and 4.4. Therefore we have to resort to *iterative algorithms*. For the applications we have in mind, we are dealing with vector–valued processes possibly defined on a quite large interval. A straightforward application of standard optimization algorithms is too expensive for problems of such a size, and several specialized algorithms have been proposed in the literature ([16, 46, 48, 34]) which deals with the general, very unstructured, setting of the DMEP. In his early work ([16]), Dempster proposed two iterative algorithms which however are very demanding from a computational point of view. Two popular methods are those proposed by T. P. Speed and H. T. Kiiveri in [46], that we now briefly discuss. Following the notation in [46], we will denote by $E_1, E_2, \ldots, E_{n_e}$ sets of unordered pairs of (not necessarily distinct) elements of V and by E their



(a)

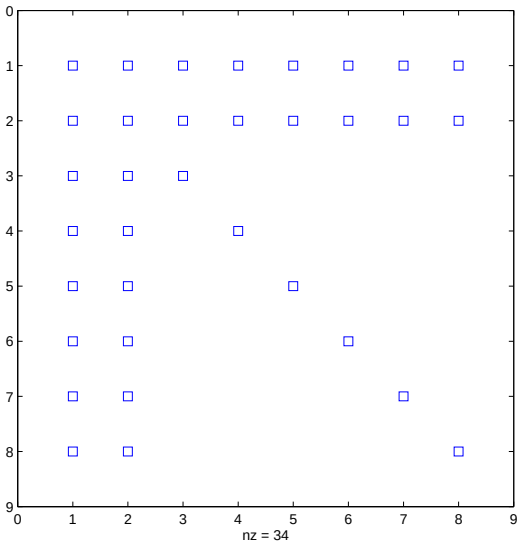


(b)

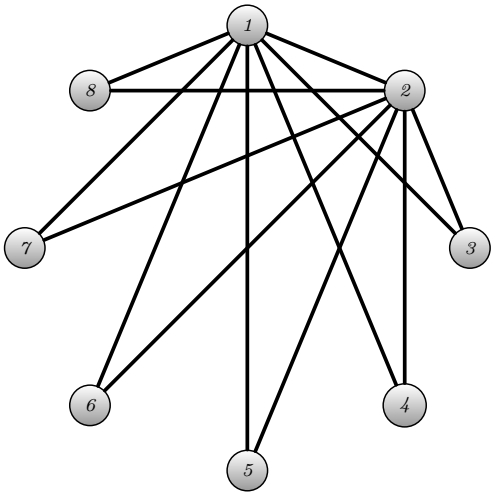


(c)

Figure 4.1: Banded Sparsity pattern for a 8×8 matrix (a) along with its associated graph (b), (c).



(a)



(b)

Figure 4.2: Banded Sparsity pattern for a 8×8 matrix (a) along with its associated graph (b).

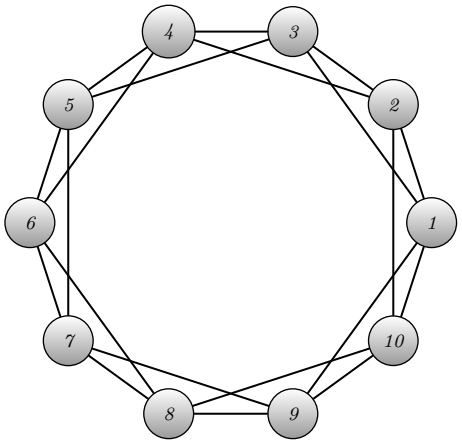


Figure 4.3: Graph associated to a banded circulant sparsity pattern for $N = 10$, $n = 2$. The graph is not chordal since, for example, the cycle $\{1, 3, 5, 7, 9\}$ does not have a chord.

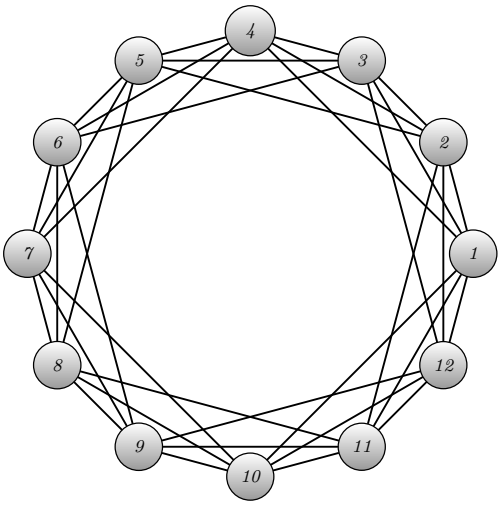


Figure 4.4: Graph associated to a banded circulant sparsity pattern for $N = 12$, $n = 3$. The graph is not chordal since, for example, the cycle $\{1, 4, 7, 10\}$ does not have a chord.

union. In its generality, the problem we want to solve is: given two positive definite matrices G and H, find a third matrix F with the property that

$$F(\alpha,\beta) = G(\alpha,\beta) \qquad \text{if } (\alpha,\beta) \in E \tag{4.1a}$$

$$F^{-1}(\alpha,\beta) = H(\alpha,\beta) \qquad \text{if } (\alpha,\beta) \notin E \tag{4.1b}$$

Both the algorithms in [46] are special cases of the following general cyclic algorithm.

Algorithm 4.1 General cyclic algorithm (Speed and Kiiveri [46])

Generate a sequence $\{F_n\}$ of positive definite matrices satisfying $F_0 = H^{-1}$ and, for $n \geq 1$,

$$F_n(\alpha,\beta) = G(\alpha,\beta) \qquad \text{if } (\alpha,\beta) \in E_{n'} \tag{4.2a}$$

$$F_n^{-1}(\alpha,\beta) = F_{n-1}^{-1}(\alpha,\beta) \qquad \text{if } (\alpha,\beta) \notin E_{n'} \tag{4.2b}$$

where $n' = n_{\bmod n_e}$ (the reminder when n is divided by n_e).

The idea is to maintain 4.1b while cycling through the E_i (at each step only the elements of F_n^{-1} corresponding to the indices in $E_{n'}$ are affected, see 4.2b) and forcing 4.1a ($|E_{n'}|$ elements at time, see 4.2a). A convergence proof can be found in [46, Proposition 3] and relies on the notion of $\mathcal{I}$–divergence (or discrimination information) [11]. The crucial step in the algorithm involves going from F_{n-1} to F_n and relies on the following Lemma ([46, Lemma 2, (i)]).

Lemma 4.1.1 (Speed and Kiiveri)**.** *Let Q, R and B be positive definite matrices. Then for $a \subseteq V$ the matrix*

$$Q^{-1} = R^{-1} + \left[\begin{array}{c|c} (B_a)^{-1} - (R_a)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] \tag{4.3}$$

is positive definite and satisfies

$$Q(\alpha,\beta) = B(\alpha,\beta) \qquad \text{if } \alpha \in a \text{ and } \beta \in a \tag{4.4a}$$

$$Q^{-1}(\alpha,\beta) = R^{-1}(\alpha,\beta) \qquad \text{if } \alpha \notin a \text{ and } \beta \notin a \tag{4.4b}$$

Proof. See ([46, proof of Lemma 2]). □

First algorithm Letting $\mathbf{R}_N$ denote the sample covariance matrix, the first algorithm is:

Algorithm 4.2 First algorithm (Speed and Kiiveri [46])

Compute all the cliques $\tilde{c}_t$ in the complementary graph $\tilde{\mathcal{G}}$, say $\{\tilde{c}_t, t = 1, \ldots, n_{\tilde{c}_t}\}$;
Initialize $\mathbf{\Sigma}_0 = \mathbf{R}_N$;
while some stopping criterion is satisfyied **do**
for all the cliques $\tilde{c}_t$ in the complementary graph **do**

$$\mathbf{\Sigma}_N^{(t)} = \mathbf{\Sigma}_N^{(t-1)} + \left[\begin{array}{c|c} \left\{\operatorname{diag}\left[\left((\mathbf{\Sigma}_N^{(t-1)})^{-1}\right)_{\tilde{c}_t}\right]^{-1}\right\}^{-1} - \left[\left((\mathbf{\Sigma}_N^{(t-1)})^{-1}\right)_{\tilde{c}_t}\right]^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] \tag{4.5}$$

end for
end while

which is the general algorithm 4.1 with $F = \mathbf{\Sigma}_N^{-1}$, $E_{n'} = \tilde{c}_t$, $G = I_{Nm}$ and $H = \mathbf{R}_N$ ($E = \tilde{c}_1 \cup \cdots \cup \tilde{c}_{n_{\tilde{c}_t}}$). In our setting, it reads as follows. Every cycle consists of as many steps as the cliques in the complementary graph $\tilde{\mathcal{G}}$ (the graph associated to the elements in $\mathcal{I}_b^c$). At each step only the elements in $\mathbf{\Sigma}_N$ corresponding to the current clique $\tilde{c}_t$ (i.e. only a subset of the entries in $\mathcal{I}_b^c$) are modified in such a way to set the elements of $\mathbf{\Sigma}_N^{-1}$ in the corresponding positions to the desired zero–value. Through the iterations the elements in $\mathbf{\Sigma}_N^{(t)}$ are fixed over $\mathcal{I}_b$ while the elements of $\left(\mathbf{\Sigma}_N^{(t)}\right)^{-1}$ vary over $\mathcal{I}_b^c$.

This first algorithm can be seen as a generalization of an algorithm by

Wermuth and Scheidt [48], which, instead of cycling through the cliques $\tilde{c}_t$, iterates through the edges in E.

The Bron–Kerbosch algorithm [6] has been employed for finding the cliques in the graph.

Example 4.1.2. Let $\sigma_0 = 9$, $\sigma_1 = 3$ and suppose we want to compute the maximum entropy completion for $N = 6$ by means of Algorithm 4.2. The complementary graph $\tilde{\mathcal{G}}$ for the given pattern is shown in Figure 4.5. The cliques are

$$\{1,3,5\}\,,\ \{2,4,6\}\,,\ \{1,4\}\,,\ \{2,5\}\,,\ \{3,6\}\,.$$

Below we show the first cycle of the algorithm.

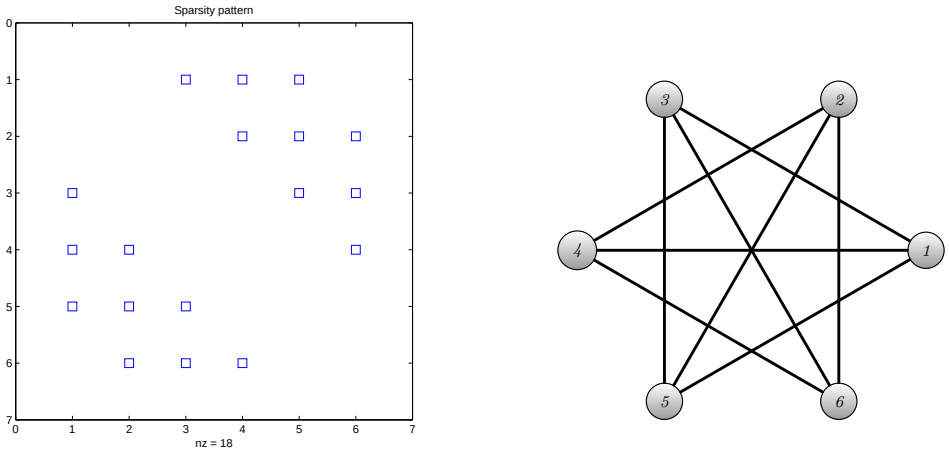


Figure 4.5: Complementary graph $\tilde{\mathcal{G}}$ (on the right) and corresponding sparsity pattern (on the left). The blue squares represent the unspecified entries.

********** *Initialization* **********

$$\boldsymbol{\Sigma}_N^{(0)} = \begin{bmatrix} 9.0000 & 3.0000 & 0.5000 & 0.2000 & 0.5000 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.5000 & 0.2000 & 0.5000 \\ 0.5000 & 3.0000 & 9.0000 & 3.0000 & 0.5000 & 0.2000 \\ 0.2000 & 0.5000 & 3.0000 & 9.0000 & 3.0000 & 0.5000 \\ 0.5000 & 0.2000 & 0.5000 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.5000 & 0.2000 & 0.5000 & 3.0000 & 9.0000 \end{bmatrix}$$

********** *First Iteration* **********

********* *Step 1* *********

$$\boldsymbol{\Sigma}_N^{(1)} = \begin{bmatrix} 9.0000 & 3.0000 & 0.9739 & 0.2000 & 0.9739 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.5000 & 0.2000 & 0.5000 \\ 0.9739 & 3.0000 & 9.0000 & 3.0000 & 0.9739 & 0.2000 \\ 0.2000 & 0.5000 & 3.0000 & 9.0000 & 3.0000 & 0.5000 \\ 0.9739 & 0.2000 & 0.9739 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.5000 & 0.2000 & 0.5000 & 3.0000 & 9.0000 \end{bmatrix}$$

$$\left(\boldsymbol{\Sigma}_N^{(1)}\right)^{-1} = \begin{bmatrix} 0.1408 & -0.0445 & 0.0000 & 0.0018 & -0.0000 & -0.0445 \\ -0.0445 & 0.1400 & -0.0445 & 0.0071 & 0.0018 & 0.0071 \\ 0.0000 & -0.0445 & 0.1408 & -0.0445 & -0.0000 & 0.0018 \\ 0.0018 & 0.0071 & -0.0445 & 0.1400 & -0.0445 & 0.0071 \\ -0.0000 & 0.0018 & -0.0000 & -0.0445 & 0.1408 & -0.0445 \\ -0.0445 & 0.0071 & 0.0018 & 0.0071 & -0.0445 & 0.1400 \end{bmatrix}$$

********* *Step 2* *********

$$\boldsymbol{\Sigma}_N^{(2)} = \begin{bmatrix} 9.0000 & 3.0000 & 0.9739 & 0.2925 & 0.9739 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.5000 & 0.2000 & 0.5000 \\ 0.9739 & 3.0000 & 9.0000 & 3.0000 & 0.9739 & 0.2000 \\ 0.2925 & 0.5000 & 3.0000 & 9.0000 & 3.0000 & 0.5000 \\ 0.9739 & 0.2000 & 0.9739 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.5000 & 0.2000 & 0.5000 & 3.0000 & 9.0000 \end{bmatrix}$$

$$\left(\boldsymbol{\Sigma}_N^{(2)}\right)^{-1} = \begin{bmatrix} 0.1407 & -0.0446 & 0.0006 & -0.0000 & 0.0006 & -0.0446 \\ -0.0446 & 0.1400 & -0.0447 & 0.0076 & 0.0016 & 0.0071 \\ 0.0006 & -0.0447 & 0.1408 & -0.0445 & 0.0000 & 0.0016 \\ -0.0000 & 0.0076 & -0.0445 & 0.1400 & -0.0445 & 0.0076 \\ 0.0006 & 0.0016 & 0.0000 & -0.0445 & 0.1408 & -0.0447 \\ -0.0446 & 0.0071 & 0.0016 & 0.0076 & -0.0447 & 0.1400 \end{bmatrix}$$

********* *Step 3* *********

$$\boldsymbol{\Sigma}_N^{(3)} = \begin{array}{cccccc} 9.0000 & 3.0000 & 0.9739 & 0.2925 & 0.9739 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.8730 & 0.2000 & 0.8448 \\ 0.9739 & 3.0000 & 9.0000 & 3.0000 & 0.9739 & 0.2000 \\ 0.2925 & 0.8730 & 3.0000 & 9.0000 & 3.0000 & 0.8730 \\ 0.9739 & 0.2000 & 0.9739 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.8448 & 0.2000 & 0.8730 & 3.0000 & 9.0000 \end{array}$$

$$\left(\boldsymbol{\Sigma}_N^{(3)}\right)^{-1} = \begin{array}{cccccc} 0.1396 & -0.0425 & -0.0014 & 0.0046 & -0.0014 & -0.0425 \\ -0.0425 & 0.1393 & -0.0425 & 0.0000 & 0.0061 & 0.0000 \\ -0.0014 & -0.0425 & 0.1396 & -0.0423 & -0.0019 & 0.0061 \\ 0.0046 & 0.0000 & -0.0423 & 0.1392 & -0.0423 & -0.0000 \\ -0.0014 & 0.0061 & -0.0019 & -0.0423 & 0.1396 & -0.0425 \\ -0.0425 & 0.0000 & 0.0061 & -0.0000 & -0.0425 & 0.1393 \end{array}$$

********* *Step 4* *********

$$\boldsymbol{\Sigma}_N^{(4)} = \begin{array}{cccccc} 9.0000 & 3.0000 & 0.9739 & 0.2925 & 0.9739 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.8730 & 0.5143 & 0.8448 \\ 0.9739 & 3.0000 & 9.0000 & 3.0000 & 0.9739 & 0.2000 \\ 0.2925 & 0.8730 & 3.0000 & 9.0000 & 3.0000 & 0.8730 \\ 0.9739 & 0.5143 & 0.9739 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.8448 & 0.2000 & 0.8730 & 3.0000 & 9.0000 \end{array}$$

$$\left(\boldsymbol{\Sigma}_N^{(4)}\right)^{-1} = \begin{array}{cccccc} 0.1396 & -0.0424 & -0.0014 & 0.0041 & 0.0005 & -0.0431 \\ -0.0424 & 0.1390 & -0.0424 & 0.0018 & 0.0000 & 0.0019 \\ -0.0014 & -0.0424 & 0.1396 & -0.0429 & -0.0001 & 0.0055 \\ 0.0041 & 0.0018 & -0.0429 & 0.1392 & -0.0423 & 0.0000 \\ 0.0005 & 0.0000 & -0.0001 & -0.0423 & 0.1393 & -0.0425 \\ -0.0431 & 0.0019 & 0.0055 & 0.0000 & -0.0425 & 0.1393 \end{array}$$

********* *Step 5* *********

$$\boldsymbol{\Sigma}_N^{(5)} = \begin{array}{cccccc} 9.0000 & 3.0000 & 0.9739 & 0.2925 & 0.9739 & 3.0000 \\ 3.0000 & 9.0000 & 3.0000 & 0.8730 & 0.5143 & 0.8448 \\ 0.9739 & 3.0000 & 9.0000 & 3.0000 & 0.9739 & 0.4851 \\ 0.2925 & 0.8730 & 3.0000 & 9.0000 & 3.0000 & 0.8730 \\ 0.9739 & 0.5143 & 0.9739 & 3.0000 & 9.0000 & 3.0000 \\ 3.0000 & 0.8448 & 0.4851 & 0.8730 & 3.0000 & 9.0000 \end{array}$$

$$\left(\boldsymbol{\Sigma}_N^{(5)}\right)^{-1} = \begin{array}{cccccc} 0.1396 & -0.0430 & 0.0003 & 0.0035 & 0.0005 & -0.0430 \\ -0.0430 & 0.1391 & -0.0425 & 0.0019 & -0.0005 & 0.0035 \\ 0.0003 & -0.0425 & 0.1394 & -0.0429 & 0.0016 & 0.0000 \\ 0.0035 & 0.0019 & -0.0429 & 0.1392 & -0.0428 & 0.0017 \\ 0.0005 & -0.0005 & 0.0016 & -0.0428 & 0.1394 & -0.0425 \\ -0.0430 & 0.0035 & 0.0000 & 0.0017 & -0.0425 & 0.1391 \end{array}$$

Second algorithm The second algorithm is:

Algorithm 4.3 Second algorithm – IPS (Speed and Kiiveri [46])

Compute all the cliques c_t in the graph $\mathcal{G}$, say $\{c_t, t = 1, \ldots, n_{c_t}\}$;
Initialize $\mathbf{\Sigma}_0 = I_{Nm}$;
while some stopping criterion is satisfyied **do**
for all the cliques $\tilde{c}_t$ in the graph $\mathcal{G}$ **do**

$$\left(\mathbf{\Sigma}_N^{(t)}\right)^{-1} = \left(\mathbf{\Sigma}_N^{(t-1)}\right)^{-1} + \left[\begin{array}{c|c} ((\mathbf{R}_N)_{c_t})^{-1} - \left(\left(\mathbf{\Sigma}_N^{(t-1)}\right)_{c_t}\right)^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] \tag{4.6}$$

end for
end while

which is the general algorithm 4.1 with $F = \mathbf{\Sigma}_N$, $E_{n'} = c_t$, $G = \mathbf{R}_N$ and $H = I_{Nm}$ ($E = c_1 \cup \cdots \cup c_{n_{c_t}}$). In our setting, it reads as follows. Every cycle consists of as many steps as the cliques in the graph of the specified entries $\mathcal{G}$. At each step only the elements in $\mathbf{\Sigma}_N^{-1}$ corresponding to the current clique c_t (i.e. only a subset of the entries in $\mathcal{I}_b$) are modified in such a way to set the elements of $\mathbf{\Sigma}_N$ in the corresponding positions to the desired value, namely equal to the sample covariance $\mathbf{R}_n$ (see (4.4a)). Through the iterations the elements in $\left(\mathbf{\Sigma}_N^{(t)}\right)^{-1}$ are fixed over $\mathcal{I}_b^c$ while the elements of $\mathbf{\Sigma}_N^{(t)}$ vary over $\mathcal{I}_b$. This algorithm is the analogous of iterative proportional scaling (IPS) for contingency tables [26] and can be also seen as a Gaussian version of the general procedure by Kullback in [34].

Example 4.1.3. Let us apply Algorithm 4.3 to the data of Example 4.1.2. Algorithm 4.3 works on the cliques of the graph associated to the given entries (Figure 4.6) which are: $\{1, 2\}$, $\{2, 3\}$, $\{3, 4\}$, $\{4, 5\}$, $\{5, 6\}$, $\{6, 7\}$. The first iteration is shown below.

********** *First Iteration* **********

********* *Step 1* *********

$$K_1^{-1} = \begin{bmatrix} 0.1250 & -0.0417 & 0 & 0 & 0 & 0 \\ -0.0417 & 0.1250 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$K_1 = \begin{bmatrix} 9 & 3 & 0 & 0 & 0 & 0 \\ 3 & 9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

********* *Step 2* *********

$$K_2^{-1} = \begin{bmatrix} 0.1250 & -0.0417 & 0 & 0 & 0 & 0 \\ -0.0417 & 0.1389 & -0.0417 & 0 & 0 & 0 \\ 0 & -0.0417 & 0.1250 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 9 & 3 & 1 & 0 & 0 & 0 \\ 3 & 9 & 3 & 0 & 0 & 0 \\ 1 & 3 & 9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

********* *Step 3* *********

$$K_3^{-1} = \begin{bmatrix} 0.1250 & -0.0417 & 0 & 0 & 0 & 0 \\ -0.0417 & 0.1389 & -0.0417 & 0 & 0 & 0 \\ 0 & -0.0417 & 0.1389 & -0.0417 & 0 & 0 \\ 0 & 0 & -0.0417 & 0.1250 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$K_3 = \begin{bmatrix} 9.0000 & 3.0000 & 1.0000 & 0.3333 & 0 & 0 \\ 3.0000 & 9.0000 & 3.0000 & 1.0000 & 0 & 0 \\ 1.0000 & 3.0000 & 9.0000 & 3.0000 & 0 & 0 \\ 0.3333 & 1.0000 & 3.0000 & 9.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

********* *Step 4* *********

$$K_4^{-1} = \begin{bmatrix} 0.1250 & -0.0417 & 0 & 0 & 0 & 0 \\ -0.0417 & 0.1389 & -0.0417 & 0 & 0 & 0 \\ 0 & -0.0417 & 0.1389 & -0.0417 & 0 & 0 \\ 0 & 0 & -0.0417 & 0.1389 & -0.0417 & 0 \\ 0 & 0 & 0 & -0.0417 & 0.1250 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

$$K_4 = \begin{bmatrix} 9.0000 & 3.0000 & 1.0000 & 0.3333 & 0.1111 & 0 \\ 3.0000 & 9.0000 & 3.0000 & 1.0000 & 0.3333 & 0 \\ 1.0000 & 3.0000 & 9.0000 & 3.0000 & 1.0000 & 0 \\ 0.3333 & 1.0000 & 3.0000 & 9.0000 & 3.0000 & 0 \\ 0.1111 & 0.3333 & 1.0000 & 3.0000 & 9.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

********* *Step 5* *********

$$K_5^{-1} = \begin{bmatrix} 0.1250 & -0.0417 & 0 & 0 & 0 & 0 \\ -0.0417 & 0.1389 & -0.0417 & 0 & 0 & 0 \\ -0.0000 & -0.0417 & 0.1389 & -0.0417 & 0.0000 & 0 \\ -0.0000 & 0.000 & -0.0417 & 0.1389 & -0.0417 & 0 \\ 0 & 0 & 0.0000 & -0.0417 & 0.1389 & -0.0417 \\ 0 & 0 & 0 & 0 & -0.0417 & 0.1250 \end{bmatrix}$$

$$K_5 = \begin{bmatrix} 9.0000 & 3.0000 & 1.0000 & 0.3333 & 0.1111 & 0.0370 \\ 3.0000 & 9.0000 & 3.0000 & 1.0000 & 0.3333 & 0.1111 \\ 1.0000 & 3.0000 & 9.0000 & 3.0000 & 1.0000 & 0.3333 \\ 0.3333 & 1.0000 & 3.0000 & 9.0000 & 3.0000 & 1.0000 \\ 0.1111 & 0.3333 & 1.0000 & 3.0000 & 9.0000 & 3.0000 \\ 0.0370 & 0.1111 & 0.3333 & 1.0000 & 3.0000 & 9.0000 \end{bmatrix}$$

********* *Step 6* *********

$$K_6^{-1} = \begin{bmatrix} 0.1389 & -0.0417 & 0 & 0 & 0 & -0.0412 \\ -0.0417 & 0.1389 & -0.0417 & 0 & 0 & 0 \\ -0.0000 & -0.0417 & 0.1389 & -0.0417 & 0.0000 & -0.0000 \\ -0.0000 & 0.0000 & -0.0417 & 0.1389 & -0.0417 & 0.0000 \\ 0.0000 & -0.0000 & 0.0000 & -0.0417 & 0.1389 & -0.0417 \\ -0.0412 & 0 & -0.0000 & 0.0000 & -0.0417 & 0.1389 \end{bmatrix}$$

$$K_6 = \begin{bmatrix} 9.0000 & 3.0325 & 1.1084 & 0.6621 & 1.0986 & 3.0000 \\ 3.0325 & 9.0217 & 3.0397 & 1.1108 & 0.6628 & 1.0986 \\ 1.1084 & 3.0397 & 9.0241 & 3.0404 & 1.1108 & 0.6621 \\ 0.6621 & 1.1108 & 3.0404 & 9.0241 & 3.0397 & 1.1084 \\ 1.0986 & 0.6628 & 1.1108 & 3.0397 & 9.0217 & 3.0325 \\ 3.0000 & 1.0986 & 0.6621 & 1.1084 & 3.0325 & 9.0000 \end{bmatrix}$$

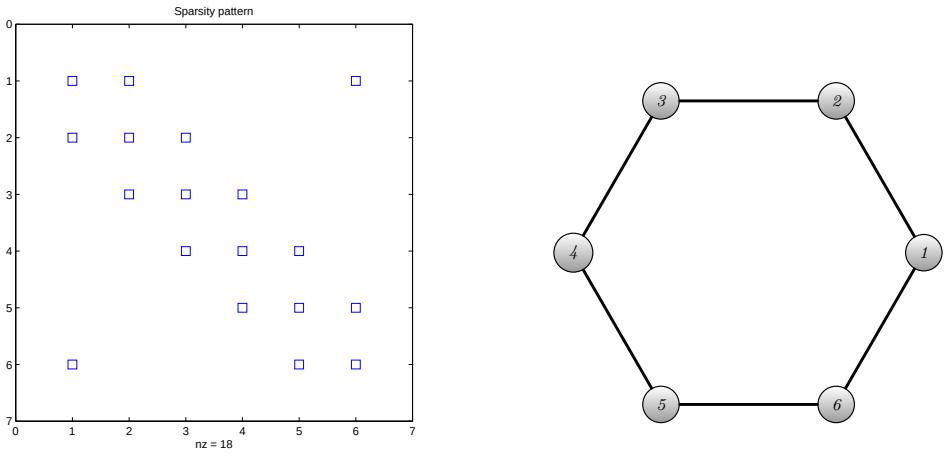


Figure 4.6: Graph $\mathcal{G}$ associated with the given data (on the right) and corresponding sparsity pattern (on the left). The blue squares represent the given entries.

Comparison between the two algorithms The choice of which algorithm is to be preferred in any given situation is very much dependent on the number and size of the cliques in $\mathcal{G}$ and $\tilde{\mathcal{G}}$. In our setting the complexity of the graph associated with the given entries depends on the bandwidth n. In particular, for bandwidth not too large with respect to the completion size (which is the case we are interested in, see Section 2.2), the complexity of the graph associated with the given data $\mathcal{G}$ is far lower than the complexity of its complementary (which, for small n's, is almost complete), see Figures 4.8 – 4.9. The execution time of the two algorithms has been compared for a completion size $N = 30$ and a bandwidth n varying between 2 and 8. The results are shown in Figure 4.7 and Table 4.1. It turns out that for n small the second algorithm (which, from now on, will be referred to as IPS) runs faster than the first, and thus has to be preferred.

Covariance selection via chordal embedding Dahl, Vanderberghe and Roychowdhury in [12] propose a new technique to improve the efficiency of Newton's method for covariance selection problems based on chordal embed-

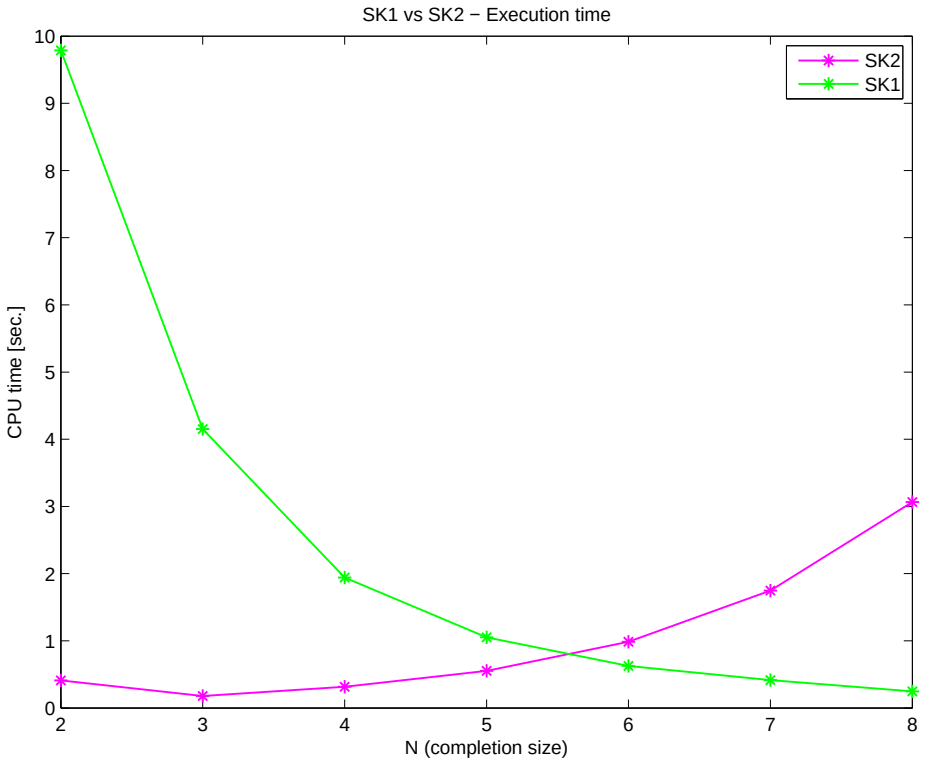
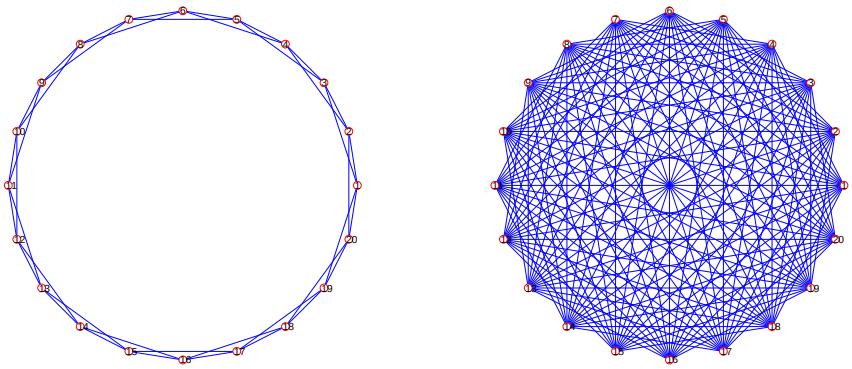


Figure 4.7: Comparison between the execution time of the first and second algorithm for $N = 30$, $m = 1$, $n = \{1, \ldots, 8\}$.

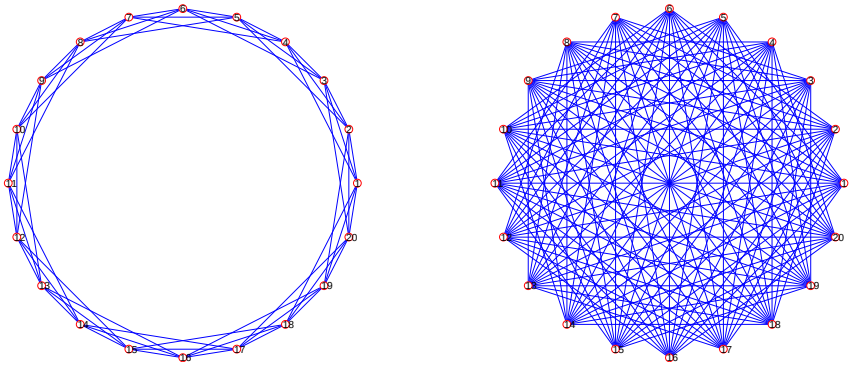
	CD		IPS	
n	cl. (max. cl. size)	CPU time [s]	cl. (max. cl. size)	CPU time [s]
2	4608(10)	9.7877	30(3)	0.4109
3	2406(7)	4.1515	30(4)	0.1783
4	1241(6)	1.9419	30(5)	0.3153
5	706(5)	1.0525	30(6)	0.5535
6	445(4)	0.6258	30(7)	0.9854
7	295(3)	0.4145	30(8)	1.7477
8	175(3)	0.2480	30(9)	3.0665

Table 4.1: Execution time of the first and second algorithm for $N = 30$, $m = 1$, bandwidth $n = \{2, \ldots, 8\}$.



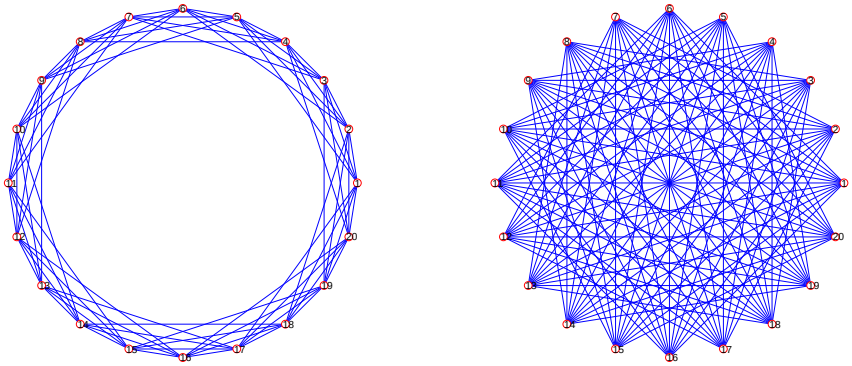
(a) $\mathcal{G}$ for $n = 2$

(b) $\tilde{\mathcal{G}}$ for $n = 2$



(c) $\mathcal{G}$ for $n = 3$

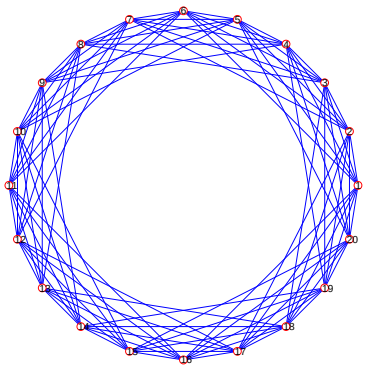
(d) $\tilde{\mathcal{G}}$ for $n = 3$



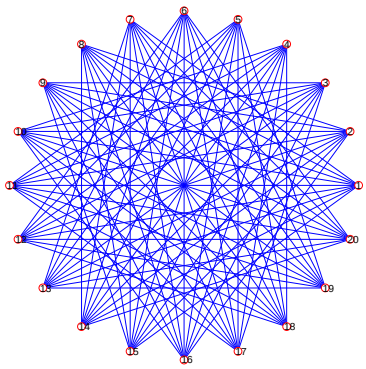
(e) $\mathcal{G}$ for $n = 4$

(f) $\tilde{\mathcal{G}}$ for $n = 4$

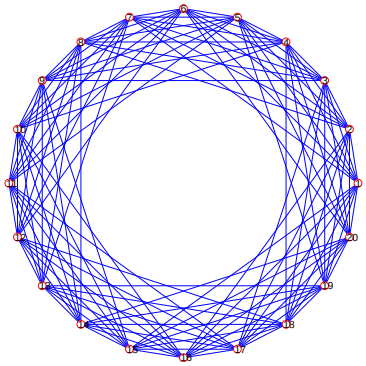
Figure 4.8: Graph $\mathcal{G}$ associated with the given data (on the right) and its complementary $\tilde{\mathcal{G}}$ (on the left) for $N = 20$ and bandwidth $n = 2, 3, 4$.



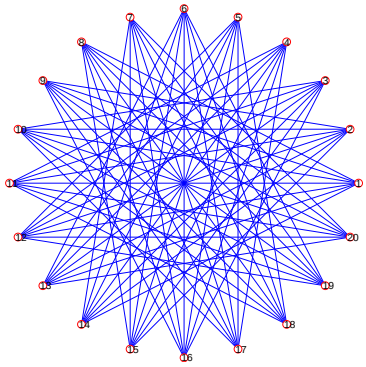
(a) $\mathcal{G}$ for $n = 5$



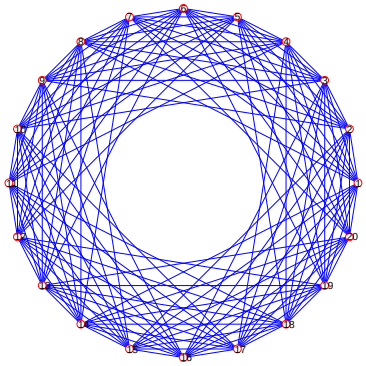
(b) $\tilde{\mathcal{G}}$ for $n = 5$



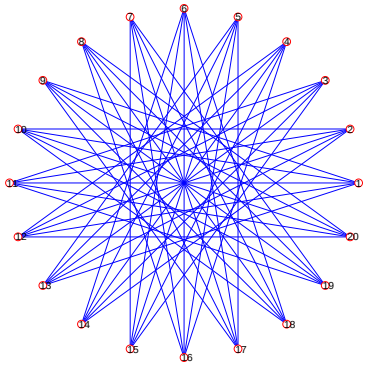
(c) $\mathcal{G}$ for $n = 6$



(d) $\tilde{\mathcal{G}}$ for $n = 6$



(e) $\mathcal{G}$ for $n = 7$



(f) $\tilde{\mathcal{G}}$ for $n = 7$

Figure 4.9: Graph $\mathcal{G}$ associated with the given data (on the right) and its complementary $\tilde{\mathcal{G}}$ (on the left) for $N = 20$ and bandwidth $n = 5, 6, 7$.

ding: the given sparsity pattern is embedded in a chordal one for which they provide efficient techiniques for computing the gradient and the Hessian. The complexity of the method is dominated by the cost of forming and solving a system of linear equations in which the number of unknowns depends on the number of nonzeros added in the chordal embedding. For circulant sparsity pattern it is easy to check that the number of nonzeros added in the chordal embedding is quite large, so that the method does not seem to be practicable.

4.2 Matricial Gradient Descent Algorithm

In this section we propose a *modified gradient descent algorithm with backtracking line search* (see, e.g., [5, Ch. 9]) for the numerical solution of the dual problem (3.26). This task requires some care because we are working in a *matricial space*. The algorithm is as follows.

Algorithm 4.4 Matricial gradient descent algorithm

Given a starting point $\Lambda \in \operatorname{dom} \bar{J}$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$
while $\left\|\nabla_\Lambda \bar{J}(\Lambda)\right\|_2 > \eta$ **do**
  $\Delta\Lambda := -\nabla_\Lambda \bar{J}(\Lambda)$
  while $\bar{J}(\Lambda + t\Delta\Lambda) > \bar{J}(\Lambda) + \alpha t \operatorname{tr}\left\{\nabla \bar{J}(\Lambda)^\top \Delta\Lambda\right\}$ **do**
    $t := \beta t$
  end while
  $\Lambda := \Lambda + t\Delta\Lambda$
end while

where $\bar{J}$ denote the functional

$$\bar{J}(\Lambda) := \operatorname{tr}(\Lambda T_n) - \operatorname{tr}\log\left\{\Pi_{\mathfrak{C}_{N,s}}\left(E_n \Lambda E_n^\top\right)\right\}. \tag{4.7}$$

Proposition 4.2.1. *The proposed Algorithm 4.4 is a gradient descent algorithm restricted to the subspace*

$$\left\{(\Lambda, \Theta) \mid \pi_{\mathfrak{C}_{N,s}^\perp}\left(E_n \Lambda E_n^\top\right) = -\left(U_N \Theta U_N^\top - \Theta\right)\right\}. \tag{4.8}$$

Proof. Once again, let $(\bar{\Lambda}, \bar{\Theta})$ be the unique minimum point of the functional J on $\mathcal{L}_+$. We know that $(\bar{\Lambda}, \bar{\Theta})$ are such that $\Sigma^o = E_n \bar{\Lambda} E_n^\top + U_N \bar{\Theta} U_N^\top - \bar{\Theta}$ is circulant. Thus one can think to restrict to look for the solution of the optimization problem on the set

$$\left\{(\Lambda, \Theta) \mid \left(E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta\right) \text{ is circulant}\right\}$$

i.e. on the set

$$\left\{(\Lambda, \Theta) \mid \pi_{\mathfrak{C}_{N,s}^\perp}\left(E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta\right) = 0\right\}$$

which, taking into account that $\left(U_N \Theta U_N^\top - \Theta\right) \in \mathfrak{C}_{N,s}^\perp$, can be written as

$$\left\{(\Lambda, \Theta) \mid \pi_{\mathfrak{C}_{N,s}^\perp}\left(E_n \Lambda E_n^\top\right) = -\left(U_N \Theta U_N^\top - \Theta\right)\right\}.$$

If we compute the dual function J on the set (4.8) we obtain

$$\begin{aligned} & J(\Lambda, \Theta) \,|_{\left\{(\Lambda,\Theta) \,|\, \pi_{\mathfrak{C}_{N,s}^\perp}\left(E_n \Lambda E_n^\top\right) = -\left(U_N \Theta U_N^\top - \Theta\right)\right\}} && (4.9) \\ & = \operatorname{tr}(\Lambda T_n) - \operatorname{tr} \log\left(E_n \Lambda E_n^\top + U_N \Theta U_N^\top - \Theta^\top\right) \\ & = \operatorname{tr}(\Lambda T_n) - \operatorname{tr} \log\left(E_n \Lambda E_n^\top - \pi_{\mathfrak{C}_{N,s}^\perp}\left(E_n \Lambda E_n^\top\right)\right) \\ & = \operatorname{tr}(\Lambda T_n) - \operatorname{tr} \log\left(\pi_{\mathfrak{C}_{N,s}}\left(E_n \Lambda E_n^\top\right)\right) && (4.10) \end{aligned}$$

which is the modified functional defined above. Thus the proposed algorithm is nothing but a gradient descent algorithm in which the search of the minimum point has been restricted to the subspace where the optimal solution is known to be, i.e. to the subspace (4.8). □

An explicit formula for $\pi_{\mathfrak{C}_{N,s}}\left(E_n \Lambda E_n^\top\right)$ has been computed in Section 3.4.

The gradient $\nabla_\Lambda \bar{J}(\Lambda)$ is given by

$$\nabla_\Lambda \bar{J}(\Lambda) = -E_n^\top \left[\pi_{\mathfrak{C}_{N,s}}\left(E_n \Lambda E_n^\top\right)\right]^{-1} E_n + \mathbf{T}_n \,.$$

4.2.1 Numerical experiments

The matricial gradient descent algorithm has been implemented in Matlab. The results are shown in Figures 4.10 and 4.11 along with Tables 4.2 and 4.3. The implementation exploits the block–circulant symmetric structure (recall, in particular, that for block-circulant matrices the inverse can be computed efficiently by means of a Fourier transform). At each iteration the algorithm requires the inversion of $\left\lceil \frac{N+1}{2} \right\rceil$ matrices of order m. It follows that the execution time increases as the completion size N and the block size m increase (see Figure 4.10 and Table 4.2). Finally, it also increases, even to a lesser amount, for increasing bandwidth n (see Figure 4.11 and Table 4.3).

4.3 Comparison between matricial gradient descent and iterative proportional scaling

In this section we compare the iterative proportional scaling (IPS) and gradient descent (GD) algorithms. Both the algorithms are implemented in Matlab. The execution times for different completion size N and block size m are plotted in Figures 4.12 and 4.13 along with Tables 4.4 and 4.5.

It can be seen that the gradient descent algorithm runs faster than the iterative proportional scaling and the gap between the two increases as N increases. Moreover, the gap becomes much more evident as m grows, making the gradient descent algorithm more attractive for applications where the process under observation is vector–valued.

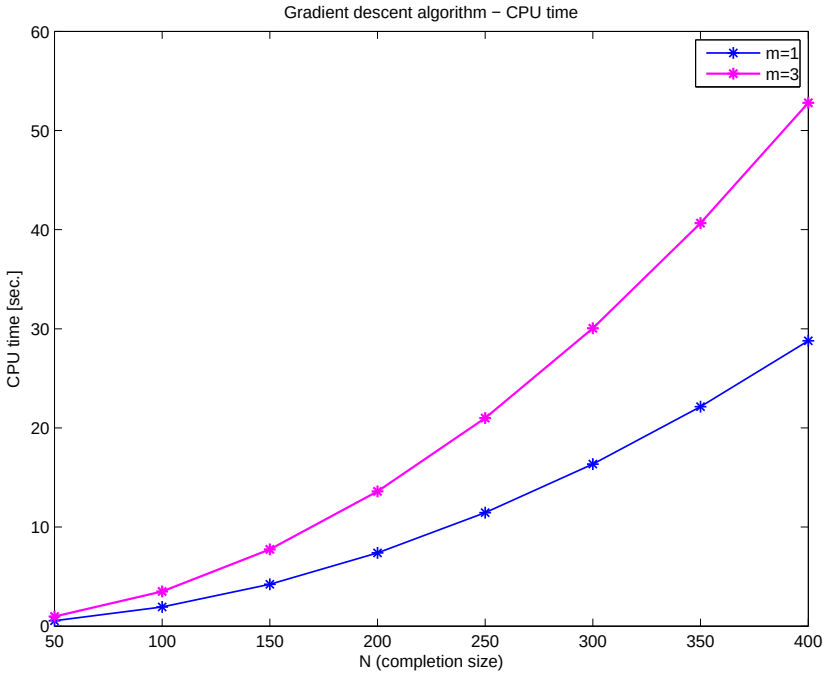


Figure 4.10: Matricial gradient descent algorithm: CPU time [sec.] for bandwidth $n = 1$, $m = \{1, 3\}$, and completion size N varying from 50 to 400.

		$m = 1$	$m = 3$
N	50	0.5535	0.9749
	100	1.9376	3.4989
	150	4.2258	7.7427
	200	7.3857	13.5903
	250	11.4440	20.9953
	300	16.3449	30.0519
	350	22.1412	40.6536
	400	28.7854	52.7949

Table 4.2: Matricial gradient descent algorithm: CPU time [sec.] plotted in Figure 4.10 for bandwidth $n = 1$, $m = \{1, 3\}$, and completion size N varying from 50 to 400.

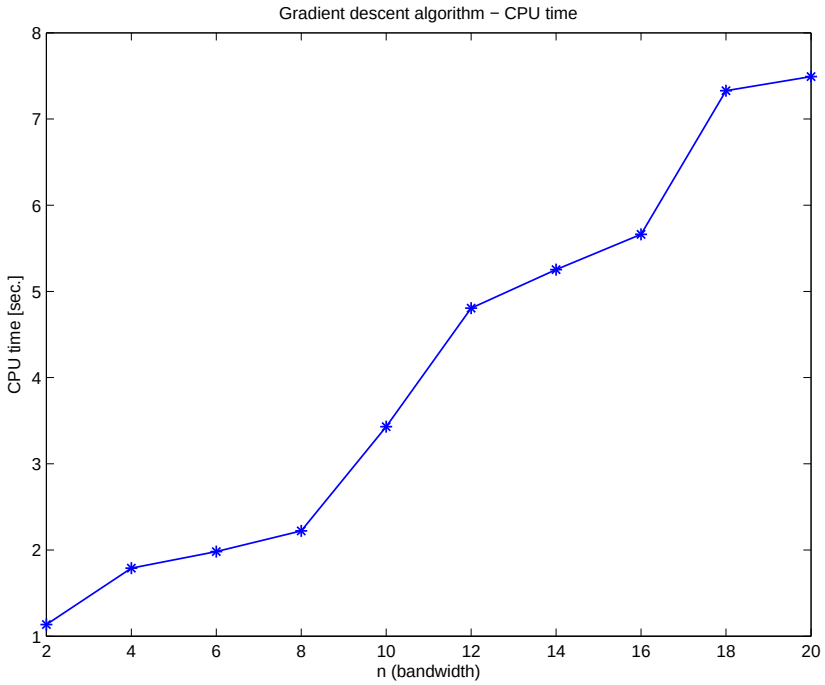


Figure 4.11: Matricial gradient descent algorithm: CPU time [in sec.] for $N = 50$, $m = 1$, n varying between 2 and 20.

n	CPU time [sec.]
2	1.1351
4	1.7895
6	1.9818
8	2.2215
10	3.4312
12	4.8058
14	5.2528
16	5.6626
18	7.3284
20	7.4922

Table 4.3: Matricial gradient descent algorithm: CPU time [in sec.] plotted in Figure 4.11 for $N = 50$, $m = 1$, n varying between 2 and 20.

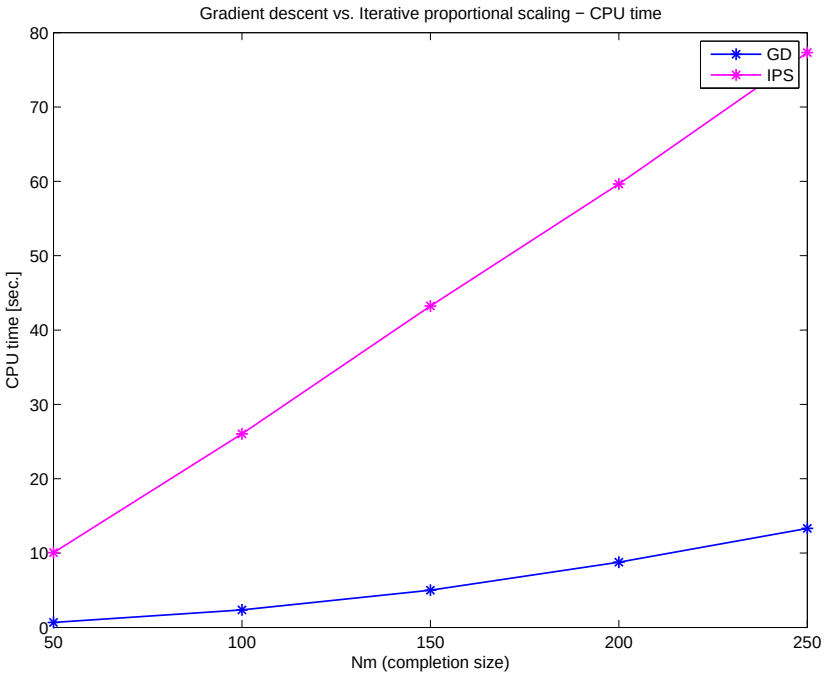


Figure 4.12: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [25, 50, 75, 100, 125]$, $m = 2$, bandwidth $n = 5$.

N	m	IPS	GD
25	2	10.0707	0.6689
50	2	26.0420	2.3574
75	2	43.2215	5.0059
100	2	59.6334	8.7657
125	2	77.3164	13.3175

Table 4.4: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [25, 50, 75, 100, 125]$, $m = 2$, bandwidth $n = 5$.

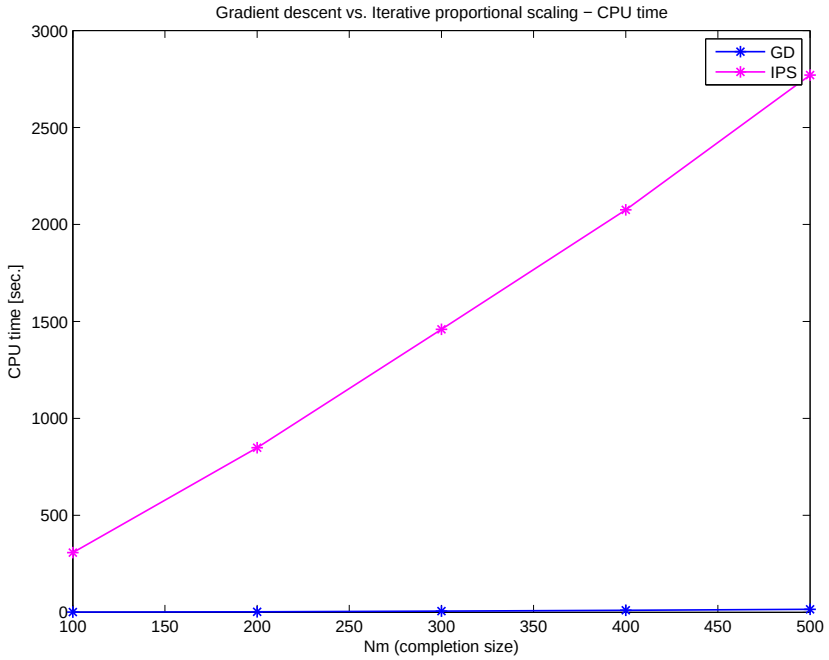


Figure 4.13: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [25, 50, 75, 100, 125]$, $m = 4$, bandwidth $n = 5$.

N	m	IPS	GD
25	4	307.4842	0.7801
50	4	848.3512	2.7421
75	4	1459.3912	5.7583
100	4	2075.0478	10.1143
125	4	2770.5883	15.1963

Table 4.5: Matricial gradient descent algorithm vs. iterative proportional scaling: CPU time [in sec.] for $N = [25, 50, 75, 100, 125]$, $m = 4$, bandwidth $n = 5$.

Chapter 5

Conclusions

In this dissertation a class of stationary reciprocal processes on a finite interval has been introduced which are the acausal analog of autoregressive (AR) processes on the integer line. These processes seem to be useful to describe signals which naturally live in a finite region of time or space. Maximum likelihood identification of these AR-type reciprocal models has been discussed. In particular it has been shown that the maximum likelihood identification problem leads to a matrix completion problem for block– circulant matrices. While circulant covariance matrices have been widely studied in the signal processing literature, the completion problem for such matrices does not seem to have been addressed before. In the present work, it has been shown that this problem can be solved by maximizing an entropy functional. Moreover, the interconnection between the block–circulant matrix completion problem and the covariance selection problem has been highlighted.

Finally, a new algorithm, which originates from the variational analysis of Chapter 3 and heavily exploits the block–circulant structure, has been proposed . This algorithm compares very favourably with the algorithms in the literature for the covariance selection problem.

Appendix A

Circulant Matrices

In this Appendix we generalize some relevant results about circulants in [13] for block–matrices.

Definition A.0.1. *A* block-circulant *matrix with N, $m \times m$ blocks, is a block-Toeplitz matrix whose block-rows (or equivalently, block-columns) are shifted cyclically, i.e.*

$$\mathbf{C}_N = \begin{bmatrix} C_0 & C_1 & \dots & \dots & C_{N-1} \\ C_{N-1} & C_0 & C_1 & \dots & \dots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & C_1 \\ C_1 & C_2 & \dots & C_{N-1} & C_0 \end{bmatrix}.$$

where $C_k \in \mathbb{R}^{m\times m}$, $k = 0, \dots, N-1$.

From the definition it is apparent that a block–circulant matrix is completely defined by its first block–row, so that it can be denoted by

$$\mathbf{C}_N = \mathrm{Circ}\{C_0, C_1, \dots, C_{N-1}\}. \tag{A.1}$$

The most simple example of circulant matrix is perhaps the $N \times N$ cir-

culant shift, namely the matrix

$$S = \mathrm{Circ}\left\{0, 1, 0, 0, \ldots, 0\right\}.$$

Clearly $S^N = I_N$, and, as is well–known (and easy to check), S^k has the (eigenvalue–eigenvector) decomposition

$$S^k F = F W^k \tag{A.2}$$

where

$$W := \mathrm{diag}\left\{1, w, w^2, \ldots, w^{N-1}\right\}$$

with $w := e^{-j\frac{2\pi}{N}}$, j denoting the imaginary unit $\sqrt{-1}$ and F is the Fourier matrix of order N, i.e. the matrix whose (k,l)-entry is

$$f_{k,l} = \frac{1}{\sqrt{N}} w^{(k-1)(l-1)}. \tag{A.3}$$

The circulant shift S plays a fundamental role in the theory of circulants. In fact, it turns out that every block–circulant matrix $\mathrm{Circ}\left\{C_0, C_1, \ldots, C_{N-1}\right\}$ can be represented as

$$\mathrm{Circ}\left\{C_0, C_1, \ldots, C_{N-1}\right\} = \sum_{k=0}^{N-1} S^k \otimes C_k \tag{A.4}$$

where $\otimes$ denotes the Kronecker product. Moreover, the following characterization of circulants holds.

Proposition A.0.1. *Let $\mathbf{C}_N \in \mathbb{R}^{Nm \times Nm}$. $\mathbf{C}_N$ is block–circulant if and only if*

$$\left(S \otimes I_m\right)^{-1} \mathbf{C}_N \left(S \otimes I_m\right) = \mathbf{C}_N \tag{A.5}$$

namely, if and only if it commutes with $(S \otimes I_m)$.

Sketch of the proof. It suffices to expand (A.5) and note that this constrains

the $(m \times m)$-block entries of $\mathbf{C}_N$ to have the circulant symmetry. $\square$

Proposition A.0.2. *Every block-circulant matrix with $N \times N$ blocks each of size $m \times m$, say $\mathbf{C}_N = Circ\{C_0, C_1, \ldots, C_{N-1}\}$, can be diagonalized as*

$$(F^* \otimes I_m)\, \mathbf{C}_N\, (F \otimes I_m) = \mathrm{diag}\{\Psi(w^0), \Psi(w^1), \Psi(w^2), \ldots, \Psi(w^{N-1})\}$$

where the $\Psi(w^\ell)$'s are the polynomial matrices

$$\Psi(x) = \sum_{k=0}^{N-1} x^k C_k$$

computed for $x = w^\ell$, $\ell = 0, \ldots, N-1$.

Proof. By the representation (A.4) and the decomposition (A.2), recalling the properties of the Kronecker product, we get

$$\begin{aligned}
(F^* \otimes I_m)\, \mathbf{C}_N\, (F \otimes I_m) &= (F^* \otimes I_m) \left(\sum_{k=0}^{N-1} S^k \otimes C_k \right) (F \otimes I_m) \\
&= \sum_{k=0}^{N-1} \left(F^* S^k F \right) \otimes C_k \\
&= \sum_{k=0}^{N-1} W^k \otimes C_k \\
&= \sum_{k=0}^{N-1} \mathrm{diag}\{w^0 C_k, w^k C_k, w^{2k} C_k, \ldots, w^{(n-1)k} C_k\} \\
&= \mathrm{diag}\{\Psi(w^0), \Psi(w^1), \Psi(w^2), \ldots, \Psi(w^{n-1})\}. \quad \text{(A.6)}
\end{aligned}$$

$\square$

Proposition A.0.3. *The inverse of a block–circulant matrix is block–circulant.*

Proof. In case the inverse exists, then, by (A.6), it is

$$\mathbf{C}_N^{-1} = (F \otimes I_m)\big(\mathrm{diag}\{\Psi(w^0)^{-1},\, \Psi(w^1)^{-1},\, \Psi(w^2)^{-1},\, \ldots,\, \Psi(w^{N-1})^{-1}\}\big)(N^* \otimes I_m).$$

If we denote with E_k the diagonal matrix with a 1 at the k–th diagonal entry, i.e.

$$E_k = \mathrm{diag}\{0, \ldots, 0, 1, 0, \ldots, 0\}$$

since $(S \otimes I_m)(F \otimes I_m) = (F \otimes I_m)(W \otimes I_m)$ while $W^{-1} = W^*$, we get

$$\begin{aligned}(S \otimes I_m)\mathbf{C}_N^{-1}(S^* \otimes I_m) &= (F \otimes I_m)(W \otimes I_m) \\ &\quad \times \left(\sum_{k=0}^{n-1} E_k \otimes \Psi(w^k)^{-1}\right)(W^{-1} \otimes I_m)(F^* \otimes I_m) \\ &= (F \otimes I_m)\left(\sum_{k=0}^{N-1} E_k \otimes \Psi(w^k)^{-1}\right)(F^* \otimes I_m) \\ &= \mathbf{C}_N^{-1}.\end{aligned}$$

which, by Theorem A.0.1, concludes the proof. □

List of Publications

International Journals

[J2] F. Carli, T. T. Georgiou, "On the Covariance Completion Problem under a Circulant Structure", *IEEE Transactions on Automatic Control* (accepted for publication)

[J1] F. Carli, A. Ferrante, M. Pavon and G. Picci "A Maximum Entropy Solution of the Covariance Extension Problem for Reciprocal Processes" *IEEE Transactions on Automatic Control* (accepted for publication)

International Conference Proceedings

[C6] F. Carli, A. Ferrante, M. Pavon, G. Picci "A Maximum Entropy approach to the Covariance Extension Problem for Reciprocal Processes" *Proc. of Int. Symp. Mathematical Theory of Network and Systems*, Budapest, Hungary (5–9 July, 2010).

[C5] F. Carli, G. Picci "On the Factorization Approach to Band Extension of Block-Circulant Matrices" *Proc. of Int. Symp. Mathematical Theory of Network and Systems*, Budapest, Hungary (5–9 July, 2010).

[C4] F. Carli, T. T. Georgiou "On the Maximum Entropy Completion of Circulant Covariance Matrices" *Proc. of Int. Symp. Mathematical Theory of Network and Systems*, Budapest, Hungary (5–9 July, 2010).

[C3] F. Carli, A. Ferrante, M. Pavon and G. Picci "A Maximum Entropy solution of the Covariance Selection Problem for Reciprocal Processes" *A Celebration of the Field of Systems and Control: An international symposium on the occasion of two milestones in the careers of Chris Byrnes and Anders Lindquist*, Stockholm, Sweden, (September 9–11, 2009).

[C2] G. Picci, F. Carli "Modeling and Identification of Reciprocal Processes" *In Proc. of the 48th IEEE Conference on Decision and Control*, Shanghai, China (December 16–18, 2009).

[C1] G. Picci, F. Carli "Modelling and Simulation of Images by Reciprocal Processes" *Proc. of EUROSIM/UKSIM08*, Cambridge, England (April 1–3, 2008).

Bibliography

[1] O.E. Barndorff-Nilsen. *Information and Exponential families in Statistica Theory.* Wiley, New York, 1978.

[2] W.W. Barrett, C. R. Johnson, and M. Lundquist. Determinantal formulation for matrix completions associated with chordal graphs. *Linear Algebra and Applications*, 121:265–289, 1989.

[3] S. Bernstein. Sur les liaisons entre le grandeurs aleatoires. In *Proc. Intern. Congr. Math*, pages 288–309, Zurich, Switzerland, 1932.

[4] J. R. S. Blair and B. Peyton. An introduction to chordal graphs and clique trees. In A. George, J. R. Gilbert, and J. W. H. Liu, editors, *Graph theory and sparse matrix computation.* Springer-Verlag, 1993.

[5] S. Boyd and L. Vanderberghe. *Convex Optimization.* Cambridge University Press, 2004.

[6] C. Bron and J. Kerbosch. Algorithm 475: finding all cliques of an undirected graph. *Commun. ACM*, 16(9):575–577, 1973.

[7] C. Byrnes and A. Lindquist. Interior point solutions of variational problems and global inverse function theorems. *International Journal of Robust and Nonlinear Control (special issue in honor of V.A. Yakubovich on the occation of his 80th birthday)*, 17:463–481, 2007.

[8] D.G. Cantor and H. Zassenhaus. A new algorithm for factoring polynomials over finite fields. *Mathematics of Computation*, 36:587–592, 1981.

[9] J. P. Carmichael, J. C. Massé, and R. Theodorescu. Processus gaussiens stationnaires réciproques sur un intervalle. *C. R. Acad. Sci. Paris Sér. I Math.*, 295:291–293, 1982.

[10] S. C. Chay. On quasi-markov random fields. *J. Multivariate Anal.*, 2:14–76, 1972.

[11] I. Csiszar. I–divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[12] J. Dahl, L. Vanderberghe, and V. Roychowdhury. Covariance selection for non–chordal graphs via chordal embedding. *Optimization Methods and Software*, 23:501–520, 2008.

[13] P. Davis. *Circulant Matrices*. John Wiley & Sons, 1979.

[14] P. Delsarte, Y. V. Genin, and Y. G. Kamp. Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits Systems*, CAS 25:149–160, 1978.

[15] A. Dembo, C. Mallows, and L. Shepp. Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation. *IEEE Trans. Information Theory*, IT-35:1206–1212, 1989.

[16] A.P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

[17] J. L. Doob. *Stochastic processes*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1990. Reprint of the 1953 original, A Wiley-Interscience Publication.

[18] H. Dym and I. Gohberg. Extension of band matrices with band inverses. *Linear Algebra and Applications*, 36:1–24, 1981.

[19] P. Faure. Stochastic realization algorithms. In R. Mehra and D. Lainiotis, editors, *System Identification: advances and case studies*, pages 1–25. Academic Press, 1976.

[20] A. Ferrante, M. Pavon, and F. Ramponi. Further results on the Byrnes-Georgiou-Lindquist generalized moment problem. In A. Ferrante A. Chiuso and S. Pinzoni, editors, *Modeling, Estimation and Control: Festschrift in honor of Giorgio Picci on the occasion of his sixty-fifth Birthday*, pages 73–83. Springer-Verlag, 2007.

[21] R. Frezza. *Models of Higher-order and Mixed-order Gaussian Reciprocal Processes with Application to the Smoothing Problem.* PhD thesis, Applied Mathematics Program, U.C.Davis, 1990.

[22] M. Fukuda, M. Kojima, K. Murota, and K. Nakata. Exploiting sparsity in semidefinite programming via matrix completion i: general framework. *SIAM Journal in Optimization*, 11:647–674, 2000.

[23] T. T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Information Theory*, 49:2910–2917, 2003.

[24] I. Gohberg, Goldberg, and M. Kaashoek. *Classes of Linear Operators vol II.* Birkhauser, Boston, 1994.

[25] R. Grone, C.R. Johnson, E. M. Sa, and H. Wolkowicz. Positive Definite Completions of Partial Hermitian Matrices. *Linear Algebra and Its Applications*, 58:109–124, 1984.

[26] S. J. Haberman. *The Analysis of frequancy data.* Univ. Chicago Press, 1974.

[27] E.J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. Wiley, 1988.

[28] B. Jamison. Reciprocal processes: The stationary gaussian case. *Ann. Math. Stat.*, 41:1624–1630, 1970.

[29] B. Jamison. Reciprocal processes. *Zeitschrift. Wahrsch. Verw. Gebiete*, 30:65–86, 1974.

[30] B. Jamison. The Markov process of Schrodinger. *Zeitschrift. Wahrsch. Verw. Gebiete*, 32:323–331, 1975.

[31] A. J. Krener. Reciprocal diffusions and stochastic differential equations of second order. *Stochastics*, 24:393–422, 1988.

[32] A.J. Krener. Realization of reciprocal processes. In C.I. Byrnes, G.B. Di Masi, and A. Kurshanskij, editors, *Proc IIASA Conf. on Modeling and Adaptive Control*. Springer, 1986.

[33] A.J. Krener. Reciprocal processes and the stochastic realization problem for acausal systems. In C.I. Byrnes and A. Lindquist, editors, *Modeling Identification and Robust Control*, pages 197–211. North Holland, 1986.

[34] S. Kullback. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39(4):1236–1243, 1968.

[35] S. L. Lauritzen. *Graphical models*. Oxford University Press, 1996.

[36] B. C. Levy and A. Ferrante. Characterization of stationary discrete-time Gaussian reciprocal processes over a finite interval. *SIAM J. Matrix Anal. Appl.*, 24:334–355, 2002.

[37] B. C. Levy, R. Frezza, and A.J. Krener. Modeling and estimation of discrete-time Gaussian reciprocal processes. *IEEE Trans. Automatic Control*, AC-35(9):1013–1023, 1990.

[38] Anders Lindquist and Giorgio Picci. Realization theory for multivariate stationary Gaussian processes. *SIAM J. Control Optim.*, 23(6):809–857, 1985.

[39] P. Masani. The prediction theory of multivariate stochastic proceses, iii. *Acta Mathematica*, 104:141–162, 1960.

[40] K. Nakata, K. Fujitsawa, M. Fukuda, M. Kojima, and K. Murota. Exploiting sparsity in semidefinite programming via matrix completion ii: implementation and numerical details. *Mathematical Programming Series B*, 95:303–327, 2003.

[41] J. A. Sand. *Four papers in Stochastic Realization Theory.* PhD thesis, Dept. of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 1994.

[42] J. A. Sand. Reciprocal realizations on the circle. *SIAM J. Control and Optimization*, 34:507–520, 1996.

[43] E. Schrödinger. Uber die umkehrung der naturgesetze. *Sitzungsber. Preuss. Akad. Wiss. Berlin Phys. Math. Kl. 8*, 9:144–153, 1931.

[44] E. Schrödinger. Sur la theorie relativiste de l'electron et l'interpretation de la mecanique quantique. *Ann. Inst. H. Poincare*, 2:269–310, 1932.

[45] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. Journal*, 27:379–423, 623–656, 1948. Reprinted in: C.E. Shannon , W. Weaver , *The Mathematical Theory of Communication*, Univ. of Illinois, Press 1949.

[46] T. P. Speed and H. T. Kiiveri. Gaussian markov distribution over finite graphs. *The Annals of Statistics*, 14(1):138–150, 1986.

[47] G. J. Tee. Eigenvectors of block-circulant and alternating circulant matrices. *Re. Lett. Inf. Math. Sci.*, 8:123–142, 2005.

[48] N. Wermut and E. Scheidt. Fitting a covariance selection model to a matrix. algorithm as105. *Appl. Statist.*, 26:88–92, 1977.

[49] P. Whittle. On the fitting of multivariate autoregressions and the approximate spectral factorization of a spectral density matrix. *Biometrica*, 50:129–134, 1963.

[50] D.C. Youla and N.N. Kazanjian. Bauer-type factorization of positive matrices and the theory of matrix polynomials orthogonal on the unit circle. *IEEE Trans. Circuits and Systems*, CAS-25:57–65, 1978.

[51] P.W. Zehna. Invariance of maximum likelihood estimators. *Annals of Mathematical Statistics*, 37:744, 1966.