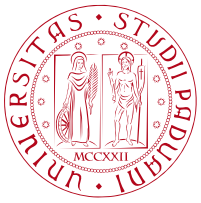
View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Padua@research



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Multivariate moment problems with applications to spectral estimation and physical layer security in wireless communications

Ph.D. Candidate
Chiara Masiero

Advisor
Prof. Augusto Ferrante

School Director
Prof. Matteo Bertocco

Coordinator
Prof. Carlo Ferrari

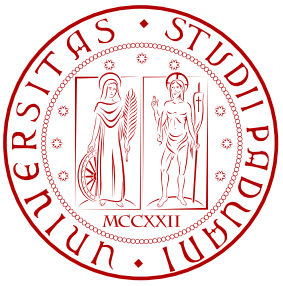
2014

Ph.D. School in
Information Engineering

Series XXVI

University of Padova

Dept. of Information Engineering



Sede amministrativa: Università degli Studi di Padova
Dipartimento: Ingegneria dell'Informazione

Scuola di dottorato di ricerca in: Ingegneria dell'Informazione
Indirizzo: Scienza e Tecnologia dell'Informazione
Ciclo: XXVI

Multivariate moment problems with applications to spectral estimation and physical layer security in wireless communications

Direttore della Scuola
Ch.mo Prof. Matteo Bertocco

Coordinatore
Ch.mo Prof. Carlo Ferrari

Supervisore
Ch.mo Prof. Augusto Ferrante

Dottoranda
Chiara Masiero

To my parents Anna and Paolo

"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."

Marie Curie

ABSTRACT

This thesis focuses on generalized moment problems and their applications in the framework of information engineering. Its contribution is twofold: The first part deals with multivariate spectral estimation, which is a key topic in system identification, whereas the second one is devoted to assessing performance of physical layer authentication techniques in modern wireless communication systems.

PART I: NEW CONVEX–OPTIMIZATION BASED TECHNIQUES FOR MULTIVARIATE SPECTRAL ESTIMATION

Multivariate spectral estimation amounts to the key task of describing the second order stochastic properties of a stationary process with many components. Modeling signals by means of stochastic processes is a well-established procedure in modern information engineering, with applications that spread from automatic control to telecommunications. Multivariate processes, in particular, play a major role when measurements provided by an array of sensors are simultaneously available and a joint stochastic model is sought. Under mild assumptions, estimating the spectral density of a process is equivalent to describing it by means of a finite memory linear model, which can be then used for smoothing and prediction, for instance. Of course, the interest in such models goes beyond engineering. Indeed, dealing with data in the form of multivariate signals is extremely common in a wide variety of disciplines, ranging from econometrics to medicine.

This thesis proposes two new techniques for tackling multivariate spectral estimation efficiently: *Relative entropy rate estimation* and *multivariate circulant rational covariance extension*.

Relative entropy rate estimation

This procedure provides a very natural multivariate extension of a state-of-the-art approach for scalar parametric spectral estimation with a complexity bound, known as THREE (Tunable High-Resolution Estimator). It allows to take into account available *a priori* information, which is modeled as a *prior* spectral density. In the framework of

THREE estimation, it is also assumed that the data samples feed a linear filter, whose output covariance is known. The filter can be designed arbitrarily and this possibility allows the user to impose interpolation constraints on the estimate. In particular, an accurate design implies high resolution of the estimate on prescribed frequency ranges, at the price of higher maximum complexity of the solution. Since the prior spectral density does not obey, in general, to the moment conditions imposed by the filter, a constrained approximation problem arises, whose solution is given by the spectral density which best approximates the *prior* while obeying to the interpolation constraints. It is worthwhile to notice that the possibility of achieving good results in the multivariate framework heavily relies on the choice of the metric which is used in order to evaluate the consistency of the candidate solution with the information encoded by the *prior*. The procedure that is described in the thesis is based on the choice of a pseudo-distance index which hinges on the notion of relative entropy rate for multivariate Gaussian processes. This choice makes the optimization problem strictly convex. Thus, a Newton-like algorithm can be used for computing the solution. Numerical examples show that this technique exhibits high-resolution features and that it is very efficient also in the case of short data records.

Multivariate circulant rational covariance extension

The problem of spectral estimation for periodic multivariate processes is studied. This framework resembles classic covariance extension, which is a tool for spectral estimation that hinges on the one-to-one correspondence between the infinite covariance lags sequence and the spectral density of a stationary process. However, periodicity may be viewed as the fact that the spectral density is defined on the discrete unit circle, and so a finite number of covariance lags provides a complete second order statistic description. Thus, a finite covariance extension problem arises. It is also shown that this issue is equivalent to a matrix completion problem for Hermitian, block-Toeplitz, block-circulant positive definite matrices. A convex optimization approach is proposed, in which the computation of the solution can be tackled efficiently by means of Fast Fourier Transform. Numerical examples show that this procedure provides an efficient tool for approximating regular covariance extension for multivariate processes. In addition, the interpretation of bilateral ARMA models paves the way to an insightful connection with the literature about reciprocal processes.

PART II: ASSESSMENT OF THE PERFORMANCE OF PHYSICAL LAYER AUTHENTICATION OVER RAYLEIGH FADING CHANNELS

The second part of the thesis deals with physical layer channel authentication. Physical layer security is complementary to higher layers security techniques and provides an effective defence mechanism for secure communication. Indeed, not only has it the capability of resisting the attacks based on massive computational capabilities which may be available in the near future; it also entails analytically predictable performance, because it is based on information theoretic arguments.

Authentication of the source of a message is a key task in secure communication: Indeed, every time a message is received, the receiver has to decide whether it was sent by the legitimate source or by an eavesdropper.

In the framework which is considered in this thesis, authentication is based on the channel estimate obtained by the receiver. Indeed, channel characteristics may allow to identify the link between a specific source and the receiver: This happens in many wide-band wireless systems, where even small changes in the position of the transmitter have a significant impact on the channel.

The authentication is performed in two steps. During a preliminary stage, the receiver can obtain an estimate of the legitimate channel recurring to higher layer security techniques. At the same time, also the eavesdropper performs channel estimation so, in the following phase, which corresponds to operating condition, it can preprocess its messages in order to deceive the receiver. In the second stage, the receiver performs channel estimation every time it is reached by a message, and it compares the resulting estimate with the legitimate pattern which was obtained in the first stage. Based on the comparison, the message is deemed as authentic or forged by the eavesdropper. Thus, a hypothesis testing problem arises. As a consequence, the worst case performance of the authentication scheme can be assessed by computing the tightest bound on the type I/II error probability region which corresponds to the optimal attacking strategy that can be carried out by the eavesdropper. Therefore, a non trivial constrained optimization problem arises, which may be recast as a moment problem where the joint probability density corresponding to the optimal attacking strategy has to be computed, and the constraints stem from the features of the communication setup and the fact that the eavesdropper has access to some side information about the legitimate channel. Once the optimization problem is modeled in terms of joint probability densities, its solution is approximated by means of a fixed point algorithm. Numerical examples suggest that this procedure is effective in assessing worst case channel authentication performance.

SOMMARIO

La tesi affronta il problema dei momenti generalizzato e le sue applicazioni nell'ambito dell'ingegneria dell'informazione. Il contributo proposto è duplice: la prima parte della tesi verte sul problema della stima della densità spettrale per processi stocastici multivariati, che sorge frequentemente nel contesto dell'identificazione dei sistemi dinamici, mentre la seconda parte è dedicata allo studio delle prestazioni delle tecniche di autenticazione a livello fisico nei moderni sistemi di comunicazione wireless.

I PARTE: NUOVE TECNICHE PER LA STIMA SPETTRALE MULTIVARIATA BASATE SULL'OTTIMIZZAZIONE CONVESSA

La stima spettrale multivariata ha il fondamentale obiettivo di descrivere le proprietà statistiche del secondo ordine di un processo stazionario a più componenti. Modellare i segnali di interesse come processi stocastici è una prassi consolidata nella moderna Ingegneria dell'Informazione, con ambiti d'applicazione che spaziano dal controllo automatico alle telecomunicazioni. I processi multivariati, in particolare, giocano un ruolo fondamentale nel caso in cui si vogliano modellare in maniera congiunta le diverse misure prodotte simultaneamente da una serie di sensori.

Le tecniche di stima spettrale offrono uno strumento utile per raggiungere questo obiettivo. Sotto certe ipotesi, infatti, stimare la densità spettrale di un processo è equivalente a un calcolare un modello lineare a memoria finita che ne descrive le proprietà statistiche del secondo ordine. La conoscenza di un modello del genere permette di implementare potenti tecniche di elaborazione di segnale, quali per esempio filtraggio e predizione.

Un altro aspetto motivante è che che il ricorso a segnali a più componenti per descrivere le misure a disposizione sta diventando sempre più comune nell'ambito di una grande varietà di discipline, che spaziano dall'econometria alla medicina, per cui l'interesse per modelli stocastici multivariati va di gran lunga oltre l'ambito ingegneristico.

Questa tesi propone due nuove tecniche per affrontare efficacemente il problema della stima spettrale multivariata: la stima basata sul tasso di entropia relativa tra processi e l'estensione di covarianza razionale per processi multivariati e periodici.

Stima spettrale basata sul tasso di entropia relativa

Questa procedura estende in modo molto naturale un approccio che rappresenta lo stato dell'arte per quanto riguarda la stima spettrale per processi scalari con un vincolo sulla massima complessità della soluzione. Questo approccio, noto come THREE - *Tunable High Resolution Estimator*, è scelto per la sua flessibilità: infatti, permette all'utente di tenere in considerazione le informazioni sulla densità spettrale vera del processo eventualmente disponibili, che sono codificate sotto forma di una densità spettrale *a priori*. Inoltre, all'interno di questo schema si fa anche l'assunzione che i dati vengano utilizzati per alimentare un filtro lineare di cui la covarianza di stato è nota: questo filtro può essere progettato liberamente e permette all'utente di imporre dei vincoli di interpolazione che possono essere utilizzati per ottenere una risoluzione elevata in corrispondenza ad alcuni intervalli di frequenza predeterminati, al prezzo di una maggiore complessità massima della soluzione. Dato che, in generale, la stima *a priori* della densità spettrale non soddisfa i vincoli imposti dal filtro, sorge un problema di ottimizzazione vincolata, la cui soluzione è data dalla densità spettrale che meglio approssima la densità *a priori* pur obbendendo alle condizioni di interpolazione. E' significativo sottolineare che la possibilità di ottenere buoni risultati nel caso multivariato dipende significativamente dalla scelta della metrica utilizzata per valutare quanto una possibile soluzione si discosta dalla densità *a priori*. In questa tesi, si è scelto di introdurre una pseudo-distanza che deriva dal concetto di tasso di entropia relativa per processi Gaussiani. Questa soluzione rende il problema di ottimizzazione strettamente convesso, per cui la densità spettrale che lo risolve può essere calcolata per mezzo di un algoritmo simile a quello di Newton. Infine, alcuni esempi numerici illustrano che questa tecnica permette di ottenere stime ad alta risoluzione in corrispondenza a particolari intervalli di frequenze prestabiliti. Inoltre, la procedura proposta si rivela molto efficace anche nel caso di scarsa numerosità campionaria dei campioni a disposizione.

Estensione circolare di covarianza per processi multivariati

Viene studiato il problema della stima spettrale per processi multivariati periodici, che si rifà al problema classico dell'estensione di covarianza. Questo è uno strumento fondamentale per la stima spettrale e si basa sulla corrispondenza biunivoca tra conoscenza dello spettro di un processo stazionario e conoscenza della sequenza infinita dei suoi coefficienti di covarianza. Tuttavia, a differenza del caso classico, nel caso di interesse, un numero finito di coefficienti di covarianza fornisce una statistica del secondo ordine sufficiente. Infatti ad un processo stazionario periodico corrisponde una

densità spettrale definita sul cerchio unitario discreto nel piano complesso ed essa è univocamente determinata risolvendo un problema di estensione finita della sequenza dei coefficienti di covarianza. Si dimostra anche che questo problema è equivalente ad un problema di completamento per matrici di Toeplitz a blocchi, definite positive e con struttura circolante a blocchi. La tesi propone un approccio di ottimizzazione convessa nel quale la soluzione può essere ricavata efficentemente ricorrendo agli algoritmi per il calcolo della trasformata di Fourier veloce (FFT - *Fast Fourier Transform*). Alcuni esempi numerici illustrano che questa procedura fornisce uno strumento efficace per approssimare l'estensione di covarianza per processi in generale non periodici. Inoltre, l'interpretazione dei modelli ARMA bilaterali corrispondenti stabilisce una connessione significativa con la letteratura inerente i modelli per processi reciproci.

II PARTE: VALUTAZIONE DELLE PRESTAZIONI DELL'AUTENTICAZIONE A LIVELLO FISICO PER CANALI CON DISSOLVENZA DI RAYLEIGH

La seconda parte della tesi si occupa del problema dell'autenticazione della sorgente di un messaggio, attuata a livello fisico. Gli schemi implementati a livello fisico sono complementari ai meccanismi adottati nei livelli superiori ed offrono uno strumento di difesa efficace per garantire la sicurezza nelle comunicazioni. Non solo hanno la capacità di resistere ad attacchi basati su capacità di calcolo estremamente elevate, che potrebbero essere disponibili nel prossimo futuro, ma offrono anche garanzie analiticamente predicibili sulle prestazioni che possono essere ottenute, poichè si basano su argomenti derivanti dalla Teoria dell'Informazione.

L'autenticazione della sorgente di un mesaggio è un compito fondamentale per garantire la sicurezza della comunicazione. Infatti, ogni volta che un messaggio viene ricevuto, il destinatario deve decidere se esso sia stato inviato dalla sorgente legittima o da una fraudolenta.

Nel contesto che viene analizzato nella tesi, l'autenticazione di basa su un test di ipotesi inerente la stima del canale di comunicazione effettuata dal ricevitore. Infatti, le caratteristiche del canale possono permettere di identificare un collegamento tra una sorgente specifica ed il ricevitore: questo accade per essempio in molti sistemi *wireless* a banda larga, nei quali variazioni anche minime nella posizione del trasmettitore hanno un impatto significativo sul canale. Quindi, l'autenticazione è attuata in due fasi. Inizialmente, il ricevitore può ottenere una stima del canale che lo lega alla sorgente legittima, ricorrendo a tecniche di sicurezza più sofisticate ed onerose implementate

a livelli più elevati. Allo stesso tempo, anche la sorgente fraudolenta può stimare le caratteristiche dei canali che la collegano alla sorgente legittima ed al ricevitore, rispettivamente. Sulla base delle stime ottenute, successivamente potrà elaborare i messaggi inviati al ricevitore in modo che risultino simili a quelli inviati dalla sorgente legittima. A regime, il ricevitore effettua una stima di canale ogni qual volta riceve un messaggio e la confronta con le caratteristiche del canale legittimo che ha ottenuto durante la prima fase. Sulla base di questo confronto il messaggio può essere accettato come legittimo, oppure considerato contraffatto.

Quindi, l'analisi al caso pessimo delle prestazioni dello schema di autenticazione può essere ottenuta valutando la regione ammissibile per le probabilità di commettere errori di I e di II specie, rispettivamente, nel caso in cui la sorgente fraudolenta attui la strategia d'attacco ottima. Di conseguenza, sorge un problema di ottimizzazione vincolata tutt'altro che banale: si tratta di un problema dei momenti in cui deve essere valutata una probabilità congiunta che descrive la strategia d'attacco ottima per la sorgente fraudolenta, con vincoli che derivano dalle caratteristiche del sistema di comunicazione e dal fatto che la sorgente fraudolenta ha accesso ad alcune informazioni sul canale legittimo. Una volta che il problema è formalizzato in termini di probabilità congiunte, viene proposto un algoritmo che approssima la soluzione ottima. Infine, alcuni esempi numerici suggeriscono che questa procedura è efficace per l'analisi delle prestazioni al caso pessimo delle tecniche di autenticazione di canale.

CONTENTS

1 MAIN CONTRIBUTIONS

Part I: New convex-optimization based techniques for multivariate spectral estimation

The first part of this dissertation proposes two new approaches to multivariate spectral estimation. This issue is recast as a generalized moment problem that can be solved efficiently by means of convex optimization techniques.

Chapter 3 introduces relative entropy rate spectral estimation. The results were published in

- A. Ferrante, C. Masiero and M. Pavon, *A New Metric for Multivariate Spectral Estimation Leading to Lowest Complexity Spectra*, in Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), Orlando, FL, USA, December 12-15, 2011;
- A. Ferrante, C. Masiero and M. Pavon, *Time and Spectral Domain Relative Entropy: A New Approach to Multivariate Spectral Estimation*, in IEEE Transactions on Automatic Control, Vol. 57, N. 10, October 2012.

Chapter 4 deals with rational covariance extension for multivariate periodic processes, that provides an efficient approximating technique for the regular covariance extension problem. This topic is the subject of

- A. Lindquist, C. Masiero and G. Picci, *On the Multivariate Circulant Rational Covariance Extension Problem*, in Proceedings of the 52th IEEE Conference on Decision and Control (CDC), Florence, Italy, December 10-13, 2013;

Part II: Assessment of the performance of physical layer authentication over Rayleigh fading channels

Chapter 5 proposes a new technique for assessing the performance of physical layer authentication in wireless communication systems. This issue can be tackled in the framework of hypothesis testing. Then, the tightest bound of the region corresponding to the type I/II error probability is evaluated by solving a generalized moment problem. The results appear in

- A. Ferrante, N. Laurenti, C. Masiero, M. Pavon and S. Tomasin, *On the Achievable Error Region of Physical Layer Authentication Techniques over Rayleigh Fading Channels*, submitted to IEEE Transactions on Information Theory, http://arxiv.org/abs/1303.0707

Part I

New convex-optimization based techniques for multivariate spectral estimation

2 INTRODUCTION

2.1 MOTIVATIONS

Stochastic processes play a key role in modern information engineering. Indeed, they provide a very natural model for random signals, whose introduction outperforms the traditional deterministic viewpoint. For instance, in this framework mechanisms such as information transmission or the presence of disturbances in control systems can be dealt with naturally. Moreover, even in a non-probabilistic framework, a simple statistical model may be more useful than a deterministic one, especially when the latter turns out to have very high complexity. As a result, stochastic processes find applications in a wide variety of fields.

Multivariate stochastic processes, in particular, are getting more and more popular because they can model measurement processes where many sensors provide data simultaneously and a joint statistic description is sought. This situation is often encountered in practice. Just to mention a few important applications, we recall that in medicine different parameters are simultaneously evaluated in order to determine the state of a patient (Avventi, Lindquist, and Wahlberg [1]). Multivariate stochastic processes can also be successfully applied in monitoring air pollutants (see Dahlhaus [28]). Finally, they can be used in finance for analyzing stock markets (Songsiri and Vandenberghe [95]). Of course, the applications we can think of go far beyond. They embrace communications, control, econometrics, image processing, fault detection, surveillance, seismology, psychology, social sciences and so on.

Under stationarity assumptions (see Section A.5), the second order properties of a stochastic process are completely defined by its spectral density. In some real world situations, the hypothesis that the process of interest is stationary may appear too restrictive. However, even if it is not satisfied, if we focus our analysis on short time windows of the process this simplification roughly holds, so it still makes sense to perform spectral estimation. Thus, despite being a classic issue (see e.g. McClellan [77], Stoica and Moses [97]), spectrum estimation for multivariate stochastic processes keeps generating widespread interest in the natural and engineering sciences (Georgiou [48, 50], Ramponi, Ferrante, and Pavon [89], Rosen and Stoffer [91], Zorzi [105]).

2.2 SPECTRAL ESTIMATION AS A GENERALIZED MOMENT PROBLEM

If the process is purely non deterministic (see Section A.9) and its spectral density is rational, spectral estimation provides a finite memory state space linear model for the process which can be used for filtering and estimation, for instance (see Section A.10). Usually there is a trade-off between descriptive power and complexity. Models whose order is low require less computational and memory burden. Thus, the identified model is often required not to exceed a certain complexity, in practice.

In the case of interest complexity is evaluated in terms of the McMillan degree of the estimated rational spectral density. Notice that finding an input process that is consistent with the estimated covariance matrix and has *rational* spectrum of prescribed maximum degree turns into a Nevanlinna-Pick interpolation problem with bounded degree (Blomqvist, Lindquist, and Nagamune [5], Georgiou [52]). The latter can be viewed as a *generalized moment problem* which is advantageously recast in the frame of various convex optimization problems. An example is provided by the *covariance extension problem* and its generalization, see Byrnes, Georgiou, and Lindquist [11], Byrnes, Gusev, and Lindquist [13], Byrnes and Lindquist [20], Byrnes et al. [21], Georgiou [47, 50]. These problems pose a number of theoretical and computational challenges for which the reader is also referred to Byrnes and Linquist [17], Ferrante, Ramponi, and Ticozzi [42], Georgiou [49, 54], Georgiou and Lindquist [56], Pavon and Ferrante [85], Zorzi [106]. Besides signal processing, significant applications of this theory are found in modeling and identification (Byrnes, Enqvist, and Linquist [10], Enqvist and Karlsson [37], Georgiou and Lindquist [55]), H_∞ robust control (Byrnes et al. [19], Georgiou and Lindquist [57]), and biomedical engineering (Nasiri Amini, Ebbini, and Georgiou [80]).

Along the same line, the first part of this dissertation introduces two novel approaches to spectral estimation for multivariate processes. They both rely on recasting spectral estimation as a generalized moment problem, and can be implemented efficiently by means of convex optimization techniques.

In Chapter 3 we introduce relative entropy rate estimator, which allows the user to take into account available information on the spectrum, in the form of a prior spectral density. Moreover, interpolation conditions on the estimated spectral density can be imposed by properly designing a bank of filters, and this paves the way to high-resolution features.

Chapter 4 deals with the problem of multivariate circulant rational covariance extension, that is rational covariance extension for multivariate periodic processes. Not only

is this problem interesting in itself, in view of the important role played by periodic processes in many information engineering issues (for example, see Section 4.7 for an application to image processing). The approach we propose for solving multivariate circulant rational covariance extension also provides an efficient approximating technique for dealing with the regular multivariate rational covariance extension problem.

3 RELATIVE ENTROPY RATE ESTIMATION

3.1 INTRODUCTION TO RELATIVE ENTROPY RATE ESTIMATION

This chapter introduces a new procedure for multivariate spectral estimation, which draws inspiration from a scalar spectral estimation technique called THREE (see Byrnes, Georgiou, and Lindquist [12], Georgiou [51]). THREE-like paradigm aims at estimating rational spectral densities by resorting to convex optimization techniques. It entails extremely interesting features:

1. It is robust in case of short data records;
2. It exhibits high resolution in prescribed frequency ranges;
3. The maximum complexity of the estimate is tunable.

A peculiarity of this approach is that available data are processed by means of a bank of filters. Then, information on the input power spectrum is extracted based on the output covariance of the filter. Thus, spectral estimation is recast in the form of a generalized moment problem. Another feature of THREE-like estimation is that it is possible to take available information into account: Indeed, a *prior* spectral density can be provided as an input to the estimation procedure. Since the *prior* spectral density in general does not satisfy the moment conditions imposed by the filter, a constrained approximation problem arises. Therefore, one looks for a density which is consistent with the constraints and best approximates the *prior*. This implies that an adequate cost function has to be introduced in order to evaluate the quality of the approximation. One possibility is to consider Kullback-Leibler divergence, as proposed in Byrnes, Georgiou, and Lindquist [12]. This choice is also motivated by the connection with prediction error methods, see e.g. Lindquist [72], Stoorvogel and Van Schuppen [98].

In the multivariate framework, however, the selection of an adequate cost function is a challenging issue. Indeed, this step is crucial, because the chosen metric should guarantee that the corresponding optimization problem can be solved efficiently. Moreover, the distance index also affects the maximum complexity of the estimate, which is expressed in terms of its McMillan degree. Of course, one may think of generalizing scalar indexes to the multichannel scenario. However, this is not obvious at all.

In Georgiou [48] a multivariate version of Kullback-Leibler pseudo-distance inspired by the von Neumann-Umegaki's relative entropy of statistical quantum mechanics (see e.g. Vedral [100], Von Neumann [101]) is considered. However, the corresponding spectrum approximation problem leads to computable solutions of bounded McMillan degree only if the prior spectral density is the identity matrix multiplied by a scalar pseudo-polynomial. This limitation was first circumvented by Ferrante, Pavon, and Ramponi [40], who introduced a suitable extension of the scalar Hellinger distance. Indeed, no assumption on the *prior* spectral density is required (except for rationality). The McMillan degree of the resulting solution, however, is higher than in the original scalar THREE method.

The approach we propose is a THREE-like multivariate spectral estimation technique which allows to overcome these difficulties. It relies on the choice of a metric which stems from *relative entropy rate* for Gaussian processes. Remarkably this method features an upper bound on the complexity of the solution which is equal to the one featured by THREE in the scalar case. Like all previous THREE-like methods, relative entropy rate estimator - also referred to as *RER* in the following - exhibits high resolution features and works extremely well, outperforming classical identification methods, in the case of short observation records.

Furthermore, the choice of our distance measure between spectra is also motivated by a novel information-theoretic result. Indeed, after introducing the notion of *spectral entropy rate* for stationary Gaussian processes, we prove that the time and spectral domain relative entropy rates are equal.

This chapter is outlined as follows. Sections 3.2 collects some information-theoretic results for Gaussian vector processes. Section 3.3 gives a brief overview on spectral representation for stationary Gaussian processes. Section 3.4 provides more details about THREE-like spectral estimation methods and introduces a new metric leading to relative entropy rate multivariate spectral estimation. In Section 3.5, further motivation for the choice of the new metric is given, based on a profound connection between time and spectral domain relative entropy rates for Gaussian processes. In Section 3.6 the new approach is introduced and a non-trivial result on the existence of the optimum solution for the corresponding optimization problem is established. A globally convergent, matricial Newton-type algorithm for computing the solution is presented in Section 3.8 In Section 3.9, both scalar and multivariate examples are studied via simulation. Conclusions and future work are presented in Section 3.10

3.2 PRELIMINARY INFORMATION-THEORETIC RESULTS FOR GAUSSIAN PROCESSES

Here we extend the notions provided in Section A.2 to the framework of Gaussian processes. More details can be found e.g. in Cover and Thomas [27], Ihara [59], Pinsker [88].

Consider a multivariate discrete-time Gaussian process $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ taking values in $\mathbb{R}^m$. Let $Y_{[-n,n]}$ be the random vector obtained by considering the window $\mathbf{y}(-n), \mathbf{y}(-n+1), \cdots, \mathbf{y}(0), \cdots, \mathbf{y}(n-1), \mathbf{y}(n)$, and let $p_{Y_{[-n,n]}}$ denote the corresponding joint density.

Definition 3.2.1. The *(differential) entropy rate* of $\mathbf{y}(t)$ is defined by

$$h_r(y) := \lim_{n\to\infty} \frac{1}{2n+1} H(p_{Y_{[-n,n]}}), \tag{3.1}$$

if the limit exists.

The following fundamental result holds (see Kolmogorov [64]):

Theorem 3.2.1. *Let $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ be a $\mathbb{R}^m$ - valued, zero-mean, Gaussian, stationary, purely nondeterministic stochastic process with spectral density Φ_y. Then*

$$h_r(y) = \frac{m}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log\det \Phi_y(e^{j\vartheta})\, d\vartheta. \tag{3.2}$$

The multivariate Szegö-Kolmogorov formula establish a fundamental connection between the quantity appearing in (3.2) and the optimal one-step-ahead predictor. Indeed, it reads

$$\det R = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det \Phi_y(e^{j\vartheta})\, d\vartheta\right\}, \tag{3.3}$$

where R is the error covariance matrix corresponding to the optimal predictor.

Let $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$, $\mathbf{z}(t) = \{\mathbf{z}(k);\, k \in \mathbb{Z}\}$ be two zero-mean, jointly Gaussian, stationary, purely nondeterministic processes taking values in $\mathbb{R}^m$. Let $Y_{[-n,n]}$ and $Z_{[-n,n]}$ be defined as above.

Definition 3.2.2. The *relative entropy rate* between $\mathbf{y}(t)$ and $\mathbf{z}(t)$ is defined by

$$\mathcal{D}_r(y\|z) := \lim_{n\to\infty} \frac{1}{2n+1} \mathcal{D}(p_{Y_{[-n,n]}} \,\|\, p_{Z_{[-n,n]}}) \tag{3.4}$$

if the limit exists.

A key result is stated by the following theorem (see Ihara [59], Stoorvogel and Van Schuppen [98]):

Theorem 3.2.2. *Let* $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ *and* $\mathbf{z}(t) = \{\mathbf{z}(k);\, k \in \mathbb{Z}\}$ *be* $\mathbb{R}^m$ *- valued, zero-mean, Gaussian, stationary, purely nondeterministic processes with spectral density functions* Φ_y *and* Φ_z*, respectively. Assume, moreover, that at least one of the following conditions is satisfied:*

1. $\Phi_y \Phi_z^{-1}$ *is bounded;*
2. $\Phi_y \in L^2(-\pi, \pi)$ *and* Φ_z *is* coercive *(i.e.* $\exists\, \alpha > 0$ *s.t.* $\Phi_z(e^{j\vartheta}) - \alpha I_m > 0$ *a.e. on* $\mathbb{T}$*).*

Then

$$\begin{aligned} \mathcal{D}_r(y\|z) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \Big\{ &\log\det\left(\Phi_y^{-1}(e^{j\vartheta})\Phi_z(e^{j\vartheta})\right) \\ &+ \mathrm{Tr}\left[\Phi_z^{-1}(e^{j\vartheta})\left(\Phi_y(e^{j\vartheta}) - \Phi_z(e^{j\vartheta})\right)\right]\Big\}\, d\vartheta. \end{aligned} \tag{3.5}$$

3.3 ON THE SPECTRAL REPRESENTATION OF STATIONARY GAUSSIAN PROCESSES

We now state a few basic facts about the spectral representation of a stationary process that can be found, for instance, in Kramer and Leadbetter [65], Lindquist and Picci [73], Rozanov [92]. Let $y = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ be a $\mathbb{R}^m$ - valued, zero-mean, purely nondeterministic, Gaussian, stationary process and let $C_l := \mathbb{E}\left[\mathbf{y}(k+l)\mathbf{y}(k)^\top\right]$, $l \in \mathbb{Z}$, be its *covariance lags*. Then

$$C_l = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{jl\vartheta}\, dF(\vartheta), \tag{3.6}$$

where F is a bounded, non-negative, matrix-valued measure called *spectral measure*. The stationary process $\mathbf{y}(t)$ admits itself the spectral representation

$$\mathbf{y}(k) = \int_{-\pi}^{\pi} e^{jk\vartheta}\, d\hat{\mathbf{y}}(e^{j\vartheta}), \tag{3.7}$$

where $\hat{\mathbf{y}}$ is a m-dimensional stochastic orthogonal measure, see Rozanov [92]. It may be obtained by defining, as in Lindquist and Picci [73, pag. 44],

$$\chi_k(\vartheta_1, \vartheta_2) := \begin{cases} \frac{e^{-j\vartheta_2 k} - e^{-j\vartheta_1 k}}{-2\pi j k} & \text{if } k \neq 0 \\ \frac{\vartheta_2 - \vartheta_1}{2\pi} & \text{if } k = 0 \end{cases}, \tag{3.8}$$

and setting

$$\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2}) := \lim_{N\to+\infty} \sum_{k=-N}^{N} \chi_k(\vartheta_1, \vartheta_2)\mathbf{y}(k) \tag{3.9}$$

where the sequence converges in mean square. We use the notation $d\hat{\mathbf{y}}(e^{j\vartheta})$ as a shorthand for $\hat{\mathbf{y}}(e^{j\vartheta}, e^{j(\vartheta+d\vartheta)})$ (with $d\vartheta > 0$). It is well known that

$$\mathbb{E}\left[d\hat{\mathbf{y}}(e^{j\vartheta})d\hat{\mathbf{y}}(e^{j\vartheta})^*\right] = dF(\vartheta). \tag{3.10}$$

Since the process $\mathbf{y}(t)$ is assumed to be purely nondeterministic, then $dF(\vartheta) = \Phi_y(e^{j\vartheta})d\vartheta$, where Φ_y is the spectral density function.

Proposition 3.3.1. *Suppose $\vartheta_1, \vartheta_2 \in (-\pi, \pi]$, then $\hat{\mathbf{y}}(e^{-j\vartheta_2}, e^{-j\vartheta_1}) = \overline{\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})}$. If, moreover, ϑ_1, ϑ_2 have the same sign, then, $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ is a circularly symmetric, normally distributed random vector. Finally, let $\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4$ be such that $[\vartheta_1, \vartheta_2] \cap [\vartheta_3, \vartheta_4] = [\vartheta_1, \vartheta_2] \cap [-\vartheta_4, -\vartheta_3] = \emptyset$. Then, $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ and $\hat{\mathbf{y}}(e^{j\vartheta_3}, e^{j\vartheta_4})$ are independent random vectors.*

Proof. Observe that $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ is a complex-valued random vector that may be written as $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2}) = \hat{\mathbf{y}}_r(e^{j\vartheta_1}, e^{j\vartheta_2}) + j\hat{\mathbf{y}}_i(e^{j\vartheta_1}, e^{j\vartheta_2})$. In view of (3.9) the real part $\hat{\mathbf{y}}_r(e^{j\vartheta_1}, e^{j\vartheta_2})$ and the imaginary part $\hat{\mathbf{y}}_i(e^{j\vartheta_1}, e^{j\vartheta_2})$ are jointly Gaussian real random vectors, so $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ is a complex-valued Gaussian vector. Since $\mathbf{y}(t) =$ is a $\mathbb{R}^m$ - valued process, $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ (which may be thought of as an integrated version of a "Fourier transform") has the Hermitian symmetry or equivalently $\hat{\mathbf{y}}(e^{-j\vartheta_2}, e^{-j\vartheta_1}) = \overline{\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})}$. Moreover, for ϑ_1 and ϑ_2 with the same sign, $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ and $\hat{\mathbf{y}}(e^{-j\vartheta_2}, e^{-j\vartheta_1})$ are orthogonal. Thus, we get

$$\begin{aligned} 0 &= \mathbb{E}\left[\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})\hat{\mathbf{y}}(e^{-j\vartheta_2}, e^{-j\vartheta_1})^*\right] \\ &= \mathbb{E}\left[\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})^\top\right] \end{aligned} \tag{3.11}$$

or, equivalently, $\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})$ is circularly symmetric normally distributed. Finally, it is easy to see that two *complex* Gaussian random vectors $\mathbf{v}_1$, $\mathbf{v}_2$ are independent if and only if $\mathbb{E}\left[\mathbf{v}_1\mathbf{v}_2^\top\right] = \mathbb{E}\left[\mathbf{v}_1\mathbf{v}_2^*\right] = 0$. In our case, by the orthogonality property, we have $\mathbb{E}\left[\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})\hat{\mathbf{y}}(e^{j\vartheta_3}, e^{j\vartheta_4})^*\right] = 0$ and, by taking into account that $\hat{\mathbf{y}}(e^{j\vartheta_3}, e^{j\vartheta_4})^\top = \hat{\mathbf{y}}(e^{-j\vartheta_4}, e^{-j\vartheta_3})^*$, we also have $\mathbb{E}\left[\hat{\mathbf{y}}(e^{j\vartheta_1}, e^{j\vartheta_2})\hat{\mathbf{y}}(e^{j\vartheta_3}, e^{j\vartheta_4})^\top\right] = 0$. ∎

3.4 THREE-LIKE ESTIMATION AND GENERALIZED MOMENT PROBLEMS

Our purpose is solving the following

Problem 3.4.1 (Spectral estimation for multivariate processes)**.** Suppose that the available data $\{y_i\}_{i=1}^N$ are generated by an unknown, zero-mean, m-dimensional, $\mathbb{R}^m$ - valued, purely nondeterministic, stationary Gaussian process $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$.

Task: Based on the data sample, estimate the spectral density of the process $\mathbf{y}(t)$. Let Φ be such density. It is required that $\Phi \in \mathcal{S}_+^{m\times m}$, i.e. the family of bounded and coercive spectral densities defined on $\mathbb{T} := \{z \in : |z| = 1\}$ of $\mathbb{R}^m$ - valued processes.

As mentioned above, we resort to an approach that draws inspiration from scalar THREE spectrum estimation. This section provides further details about this paradigm. A general THREE-like approach is sketched in Fig. 1 and hinges on the following elements:

1. A rational filter to process the available data. The filter has transfer function
$$G(z) = (zI - A)^{-1}B, \tag{3.12}$$
where $A \in \mathbb{R}^{n\times n}$ has all its eigenvalues inside the unit circle, $B \in \mathbb{R}^{n\times m}$ is full rank, $n \geqslant m$, and (A, B) is a reachable pair;
2. An estimate of the steady-state covariance Σ of the output process $\mathbf{x}(t)$, defined by
$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{y}(t); \tag{3.13}$$
3. A *prior* spectral density $\Psi \in \mathcal{S}_+^{m\times m}$;
4. A metric that measures the distance between two spectral densities.

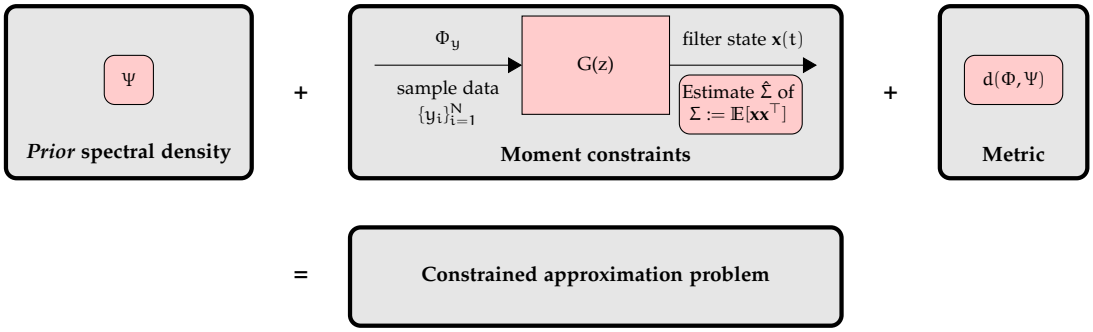


Figure 1: Scheme of THREE-like estimation paradigm. The red blocks denote the design parameters.

Rational filter $G(z)$

The filterbank (3.12), which can be arbitrarily designed by the user, imposes interpolation conditions. This occurs because Φ must satisfy

$$\int G\Phi G^* = \Sigma \tag{3.14}$$

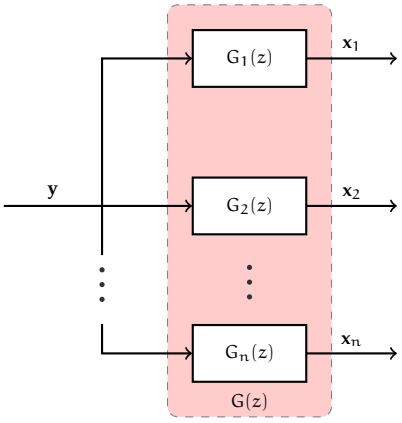


Figure 2: Block diagram of the bank of filters

(here and throughout the chapter, integration - when not otherwise specified - is on the unit circle with respect to normalized Lebesgue measure). The next example shows how the filter provides Nevanlinna-Pick interpolation data for the positive real part Φ_+ of Φ, see Byrnes, Georgiou, and Lindquist [12, Section II].

Example 3.4.1 (THREE: Connection with Nevanlinna-Pick interpolation)**.** It is convenient to write the transfer function $G(z)$ in the form

$$G(z) = \begin{bmatrix} G_1(z) \\ G_2(z) \\ \vdots \\ G_n(z) \end{bmatrix}. \tag{3.15}$$

whose corresponding block diagram representation is shown in figure 2.

For the sake of simplicity, assume now that the wide-sense stationary stochastic process $y(t)$ is scalar. Suppose that $G_k(z)$ is a first order stable filter, i.e.

$$G_k(z) = \frac{z}{z - p_k}, \quad |p_k| < 1. \tag{3.16}$$

Therefore

$$\mathbf{x}_k(t) = \mathbf{y}(t) + p_k \mathbf{x}_k(t-1), \tag{3.17}$$

and we can write $\mathbf{x}_k(t) = \sum_{h=0}^{\infty} {p_k}^h \mathbf{y}(t-h)$. Based on the results provided in Section A.10 we can conclude that the output process $\mathbf{x}_k(t)$ is stationary. If p_k is real, the output process is also real too and we get

$$\begin{aligned}
\mathbb{E}\left[\mathbf{x}_k^2(t)\right] &= \mathbb{E}\left[\left(\sum_{h=0}^{\infty} {p_k}^h \mathbf{y}(t-h)\right)^2\right] \\
&= \mathbb{E}\left[\left(\mathbf{y}(t) + p_k \mathbf{y}(t-1) + {p_k}^2 \mathbf{y}(t-2) + \dots\right)^2\right] \\
&= \mathbb{E}[y^2(t)]\left(1 + p_k^2 + p_k^4 + \dots\right) + 2p_k \mathbb{E}\left[\mathbf{y}(t)\mathbf{y}(t-1)\right]\left(1 + p_k^2 + p_k^4 + \dots\right) \\
&\quad + 2{p_k}^2 \mathbb{E}\left[\mathbf{y}(t)\mathbf{y}(t-2)\right]\left(1 + p_k^2 + p_k^4 + \dots\right) + \dots \\
&= r_0\left(1 + p_k^2 + p_k^4 + \dots\right) + 2p_k r_1\left(1 + p_k^2 + p_k^4 + \dots\right) \\
&\quad + 2{p_k}^2 r_2\left(1 + p_k^2 + p_k^4 + \dots\right) + \dots \\
&= \frac{2}{1 - {p_k}^2}\left(\frac{1}{2} r_0 + r_1 p_k + r_2 {p_k}^2 + \dots\right) \\
&= \frac{2}{1 - {p_k}^2} \Phi_+({p_k}^{-1}),
\end{aligned} \tag{3.18}$$

where the last equality can be written thanks to the additive decomposition of the spectrum (A.26). Thus, we conclude that

$$\Phi_+({p_k}^{-1}) = \frac{1}{2}\left(1 - {p_k}^2\right) \mathbb{E}\left[\mathbf{x}_k^2(t)\right], \tag{3.19}$$

which can be read as an interpolation condition on Φ_+, as soon as we estimate $\mathbb{E}\left[\mathbf{x}_k^2(t)\right]$. If p_k is a complex pole, instead, the covariance of the output process is given by

$$\mathbb{E}\left[|\mathbf{x}_k(t)|^2\right] = \frac{1}{1 - |p_k|^2}\left(\Phi_+({p_k}^{-1}) + \Phi_+(\overline{p_k}^{-1})\right), \tag{3.20}$$

where $\overline{p_k}$ is the complex conjugate of p_k.

Assume $G_k(z)$ is given by (3.16), for each $k = 1, \dots, n$. We can obtain it by means of the state space realization

$$\mathbf{x}(t+1) = A\mathbf{x}(t) + B\mathbf{y}(t), \quad \text{with} \quad A = \begin{bmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_n \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \tag{3.21}$$

Therefore, the covariance matrix of the n-dimensional output process $\mathbf{x}(t)$ reads

$$\begin{aligned}\Sigma &= \mathbb{E}\left[\mathbf{x}(t)\bar{\mathbf{x}}(t)\right] \\ &= \begin{bmatrix} \frac{w_1+\bar{w}_1}{1-p_1\bar{p}_1} & \frac{w_1+\bar{w}_2}{1-p_1\bar{p}_2} & \cdots & \frac{w_1+\bar{w}_n}{1-p_1\bar{p}_n} \\ \frac{w_2+\bar{w}_1}{1-p_2\bar{p}_1} & \frac{w_2+\bar{w}_2}{1-p_2\bar{p}_2} & \cdots & \frac{w_2+\bar{w}_n}{1-p_2\bar{p}_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n+\bar{w}_1}{1-p_n\bar{p}_1} & \frac{w_n+\bar{w}_2}{1-p_n\bar{p}_2} & \cdots & \frac{w_n+\bar{w}_n}{1-p_n\bar{p}_n} \end{bmatrix},\end{aligned} \tag{3.22}$$

where, by means of (A.25),

$$w_k = \Phi_+(p_k{}^{-1}) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{e^{-j\theta}+p_k}{e^{-j\theta}-p_k}\Phi(e^{j\theta})d\theta, \quad k = 1,\dots,n. \tag{3.23}$$

Thus, the covariance has the form of a Pick matrix. The problem of parameterizing the set of spectral density functions that satisfy (3.14) can be recast in the form of a classical *Nevanlinna-Pick interpolation problem*. Recall that, since $\Sigma \geqslant 0$, such a problem is feasible. If the covariance is a positive definite matrix, then there are infinitely many solutions.

Next, we show that THREE-like estimation allows to deal with rational covariance extension, too. The latter is the problem of computing a rational spectral density which is consistent with the available covariance lags. This is a classical issue in partial stochastic realization, system identification and control theory (see e.g. Byrnes, Gusev, and Lindquist [14], Byrnes et al. [21], Enqvist [34, 35], Enqvist and Karlsson [37]). Recall that the covariance lags up to order n are the first $n+1$ coefficients of the series expansion of $\Phi_+(z)$ near infinity. Thus, we can interpret the information they convey as interpolation constraints on the input spectral density.

Example 3.4.2 (THREE: Connection with covariance extension problem)**.** Again, for the sake of simplicity, let $\mathbf{y}(t)$ be scalar. Assume that the filter realizes the transfer function

$$G(z) = \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-n+2} \\ z^{-n+1} \end{bmatrix}, \tag{3.24}$$

whose state-space realization can be obtained by means of the matrices

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{3.25}$$

The k-th component of output of the filter, fed by $\mathbf{y}(t)$, is given by its time-delayed version $\mathbf{y}(t-k+1)$. Therefore, the covariance of the output process is equal to

$$\Sigma = \begin{bmatrix} r_0 & r_1 & \dots & r_{n-1} \\ \bar{r}_1 & r_0 & \dots & r_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{r}_{n-1} & \bar{r}_{n-2} & \dots & r_0 \end{bmatrix}. \tag{3.26}$$

Thus, we find that the class of functions that satisfy (3.14) is given by all the possible solutions of the covariance extension problem of completing the sequence $r_0, \dots, r_{n-1}$ up to a positive, infinite sequence. As in the previous case, this set is nonempty since $\Sigma \geqslant 0$ and has infinitely many elements if $\Sigma > 0$.

Thus, by means of a proper choice of $G(z)$ we can recast traditional issues in system identification and control in the framework of THREE paradigm. In addition, we can also model the effect of linear measurement tools used for processing the available data. There is even more: Indeed, a clever choice of the location of the filterbank poles can improve the resolution of the estimate in the corresponding frequency range (see Section 3.9 and [12, 62]). As one may expect, there is also a trade-off: High order filters lead to solutions of higher complexity, see Section 3.6. Therefore, the choice of the filter can be also interpreted as a mechanism that permits tuning.

Output covariance estimate

In order to initialize our estimation procedure, we need an adequate estimate $\hat{\Sigma}$ of the output covariance Σ. This step is crucial in guaranteeing the feasibility of the corresponding optimization problem. Indeed, the output covariance of a linear filter $G(z)$ has to obey to some structure constraints (see e.g. Burg, Luenberger, and Wenger [8], Georgiou [54]). In the following, $\mathcal{Q}_m \subset \mathbb{R}^{m \times m}$ denotes the $m(m+1)/2$-dimensional, real vector space of m-dimensional symmetric matrices. Let $\mathcal{C}_{R+}^{m \times m}$ be the set of continuous spectral densities of m-dimensional $\mathbb{R}^m$ - valued processes defined on the unit

circle $\mathbb{T}$. We indicate by $\mathcal{V}\left(\mathcal{C}_{R+}^{m\times m}\right)$ the linear space generated by $\mathcal{C}_{R+}^{m\times m}$. Consider the linear operator defined by

$$\begin{aligned}\Gamma &: \mathcal{V}\left(\mathcal{C}_{R+}^{m\times m}\right) \to \mathcal{Q}_n \\ \Gamma(\Phi) &:= \int G\Phi G^*.\end{aligned} \tag{3.27}$$

Then, the output covariance belongs to Range Γ. However, the sample covariance

$$\Sigma_s := \frac{1}{N}\sum_{k=1}^{N} x_k x_k^\top \tag{3.28}$$

usually does not. Thus, the following problem needs to be solved:

Problem 3.4.2 (Computation of $\hat{\Sigma}$)**.** Given $G(z)$ and $\{y_k\}_{k=1}^N$, find a positive definite matrix $\hat{\Sigma}$ such that

1. it is compatible with the filter, i.e. $\hat{\Sigma} \in \text{Range}\,\Gamma$;
2. it is "close" to the sample covariance Σ_s.

In Georgiou [54] a method for computing a basis of Range Γ was first proposed. Then, problem 3.4.2 was solved in Ramponi, Ferrante, and Pavon [89] by means of a projection-based technique. In the following, we will resort to a more recent technique, that was proposed in Ferrante, Pavon, and Zorzi [41]. It is based on the solution of an ancillary optimization problem, so that the "best" approximant $\hat{\Sigma}$ is chosen as

$$\hat{\Sigma} = \underset{\Sigma\,\in\,\text{Range}\,\Gamma,\Sigma>0}{\arg\min} \quad \frac{1}{2}\left[\log\det\Sigma_s{}^{-1}\Sigma + \text{Tr}\Sigma^{-1}\Sigma_s - n\right], \tag{3.29}$$

where Σ_s is the sample covariance (3.28). Numerical results in Ferrante, Pavon, and Zorzi [41] suggest that this technique can be implemented efficiently and that it outperforms the previously available ones.

Prior spectral density

In practice, it is often the case that some information on the input spectral density is available. Our estimation procedure allows to take such *a priori* information into account. In particular, it is encoded in the form of a spectral density Ψ, which is used to initialize the estimation algorithm. For example, Ψ may simply be a coarse estimate of the true spectrum. When no prior information on Φ is available, the *prior* Ψ is set equal either to the identity matrix or to the sample covariance of the available data

$\{y_i\}_{i=1}^N$. Section 3.6 – see inequality (3.69) – shows that the prior Ψ yields a smooth parametrization of solutions with bounded degree. Thus, the choice of Ψ, as well as the design of the filter $G(z)$, allows the user to define the trade-off between the complexity of the solution and its capability of achieving high resolution or capturing rich dynamics. Thus, it can be also interpreted as a tuning parameter.

Metric

In general, the *prior* spectral density Ψ does not obey to the interpolation conditions (3.14). Thus, we have to compute a spectral density satisfying (3.14) which is as "close" as possible to Ψ. Therefore, it is necessary to introduce an adequate distance index. This is a frequently encountered problem in spectral estimation. Similarly, one may also want to compare all the densities in a given family in an informative, quantitative manner. Thus, it is very important to develop problem-specific metrics (see e.g. Georgiou [44, 45, 53], Jiang, Ning, and Georgiou [60]). Moreover, distances between power spectra can be used quite effectively in identifying transitions, changes, and affinity between time series or even spatial series. Applications include automated phoneme recognition, for instance. In that case, suitable metrics allow to identify natural transition time markers in speech. In a similar fashion, this idea can be fruitfully applied to image segmentation. Indeed, two-dimensional distributions can be identified on the inside and outside of a curve. Then, the curve is evolved using geometric active contours to ensure maximal separation of the spectral content of two regions. This idea has been recently applied to visual tracking (see e.g. Sandhu, Georgiou, and Tannenbaum [94]).

As briefly explained in Section 3.1, the crucial choice of the metric used for comparing spectral densities in the framework of multivariate THREE-like estimation is dictated by the following essential requirements:

1. The variational analysis should lead to a computable solution;

2. The solution should be rational of low McMillan degree at least when the prior Ψ is such.

In the scalar case described in Byrnes, Georgiou, and Lindquist [12], Georgiou and Lindquist [56], the authors chose to minimize the following Kullback-Leibler type criterion, which features both of the above specifications:

$$d_{KL}(\Psi, \Phi) = \int \Psi \log \frac{\Psi}{\Phi}. \tag{3.30}$$

In the multivariable case, however, it is not obvious at all how to choose a proper metric. In Georgiou [48], a multivariate generalization of Kullback-Leibler divergence was introduced. This choice drew inspiration from *Von Neumann-Umegaki's relative entropy* (see e.g. Nielsen and Chuang [82]), frequently employed in statistical quantum mechanics. The resulting spectrum approximation problem, however, leads to computable solutions of bounded McMillan degree only in the case when the prior spectral density has the form $\Psi(z) = \psi(z)I$, where $\psi(z)$ is a scalar spectral density. When $\Psi = I$, in particular, this apporach yields the *maximum entropy solution*, as described in Blomqvist, Lindquist, and Nagamune [5], Georgiou [48, 50]. This limitation was first overcome in Ferrante, Pavon, and Ramponi [40], where the following extension of Hellinger distance was suggested:

$$\begin{aligned} d_H(\Psi,\Phi)^2 := \inf_{W_\Psi, W_\Phi} \operatorname{Tr} \int (W_\Psi - W_\Phi)(W_\Psi - W_\Phi)^*, \\ \text{such that} \quad W_\Psi W_\Psi^* = \Psi \quad \text{and} \quad W_\Phi W_\Phi^* = \Phi. \end{aligned} \tag{3.31}$$

The cost function (3.31) is just the L^2-distance between the sets of *square spectral factors* of the two spectra. Thus, in contrast with Kullback-Leibler divergence, this is a *bona fide* distance. Moreover, the variational analysis can be carried out explicitely also in the case of a general rational *prior* density, leading to a computable rational solution. The complexity of the solution, however, is usually noticeably higher than in the original scalar THREE approach.

Now we introduce a new metric based on relative entropy rate for Gaussian processes. Based on this cost function, in the next section we propose a new multivariate THREE-like spectral estimation technique, which turns out to be rather effective and achieves the performance of scalar THREE spectral estimator in terms of maximum complexity of the solution. Motivated by relation (3.5), we define a new pseudo-distance among spectra in $\mathcal{S}_+^{m\times m}$:

$$\begin{aligned} d_{RER}(\Phi,\Psi) := \frac{1}{4\pi}\int_{-\pi}^{\pi} \Big\{ \log\det\left(\Phi^{-1}(e^{j\vartheta})\Psi(e^{j\vartheta})\right) \\ + \operatorname{Tr}\left[\Psi^{-1}(e^{j\vartheta})\left(\Phi(e^{j\vartheta}) - \Psi(e^{j\vartheta})\right)\right] \Big\}\, d\vartheta. \end{aligned} \tag{3.32}$$

Notice that in the case of scalar spectra, $d_{RER}(\Phi,\Psi) = 1/2\, d_{IS}(\Phi,\Psi)$, where

$$d_{IS}(\Phi,\Psi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\{ \frac{\Phi(e^{j\vartheta})}{\Psi(e^{j\vartheta})} - \log\frac{\Phi(e^{j\vartheta})}{\Psi(e^{j\vartheta})} - 1 \right\} d\vartheta \tag{3.33}$$

is the classical Itakura-Saito distance, which is related to maximum likelihood estimation and is frequently encountered in speech processing (Basseville [4], Gray et al. [58]). Further motivation for this distance choice is provided by a novel information-theoretic result relating time and spectral domain relative entropy rates, stated in Section 3.5.

3.5 SPECTRAL RELATIVE ENTROPY RATE

Consider two zero-mean, jointly Gaussian, stationary, purely nondeterministic stochastic processes $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ and $\mathbf{z}(t) = \{\mathbf{z}(k);\, k \in \mathbb{Z}\}$ taking values in $\mathbb{R}^m$ with spectral representation

$$\mathbf{y}(k) = \int_{-\pi}^{\pi} e^{jk\vartheta}\, d\hat{\mathbf{y}}(e^{j\vartheta}), \quad \mathbb{E}\left\{ d\hat{\mathbf{y}}(e^{j\vartheta}) d\hat{\mathbf{y}}(e^{j\vartheta})^* \right\} = \Phi_y(e^{j\vartheta}) d\vartheta, \tag{3.34}$$

$$\mathbf{z}(k) = \int_{-\pi}^{\pi} e^{jk\vartheta}\, d\hat{\mathbf{z}}(e^{j\vartheta}), \quad \mathbb{E}\left\{ d\hat{\mathbf{z}}(e^{j\vartheta}) d\hat{\mathbf{z}}(e^{j\vartheta})^* \right\} = \Phi_z(e^{j\vartheta}) d\vartheta. \tag{3.35}$$

Let $\vartheta_k = \frac{\pi k}{n}$, and consider the complex Gaussian random vectors $\delta\hat{y}_k := \hat{\mathbf{y}}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$ and $\delta\hat{z}_k := \hat{\mathbf{z}}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$, with $k = 0, 1 \ldots, 2n$. Define now the random vectors

$$\hat{Y}_k := \begin{bmatrix} \delta\hat{y}_0 \\ \vdots \\ \delta\hat{y}_{k-1} \end{bmatrix}, \quad \hat{Z}_k := \begin{bmatrix} \delta\hat{z}_0 \\ \vdots \\ \delta\hat{z}_{k-1} \end{bmatrix}, \quad k = 1, \ldots, 2n, \tag{3.36}$$

and denote their joint probability density by $p(\hat{Y}_k)$ and $p(\hat{Z}_k)$, respectively.

Definition 3.5.1. The *spectral relative entropy rate* between between y and z is defined by the following limit, provided it exists:

$$\mathcal{D}_r(d\hat{y} \| d\hat{z}) := \lim_{n\to\infty} \frac{1}{2n} \mathcal{D}(p\left(\hat{Y}_{2n}\right) \,\|\, p\left(\hat{Z}_{2n}\right)). \tag{3.37}$$

We now establish a remarkable connection between time-domain and spectral-domain relative entropy rates. First we need the following lemma:

Lemma 3.5.1 (Dai Pra, [29])**.** *Let* $\mathbf{u}, \mathbf{v}$ *be* k*-dimensional, real random vectors with probability distributions* P, Q*, respectively. Let* $f : \mathbb{R}^k \to \mathbb{R}^h$ *be measurable and* P_a, Q_a *be the probability distributions of the augmented vectors* $[\mathbf{u}^\top\ f(\mathbf{u})^\top]^\top$ *and* $[\mathbf{v}^\top\ f(\mathbf{v})^\top]^\top$*, respectively. Then*

$$\mathcal{D}(P_a \,\|\, Q_a) = \mathcal{D}(P \,\|\, Q). \tag{3.38}$$

Proof. Recall the variational formula for relative entropy Dembo and Stroock [31]:

$$\mathcal{D}(P \,\|\, Q) = \sup_{\varphi \in \phi} \left\{ \mathbb{E}[\varphi(\mathbf{v})] - \log \mathbb{E}\left[e^{\varphi(\mathbf{u})} \right] \right\}, \tag{3.39}$$

where ϕ is the set of all measurable and bounded functions $\varphi : \mathbb{R}^k \to \mathbb{R}$. Consider a measurable and bounded function $\varphi : \mathbb{R}^k \to \mathbb{R}$. Define $\varphi_a : \mathbb{R}^{k+h} \to \mathbb{R}$ by

$$\varphi_a([x^\top\ x'^\top]^\top) := \varphi(x), \tag{3.40}$$

where $x' \in \mathbb{R}^h$. Obviously, φ_a is bounded and measurable, and

$$\begin{aligned}\mathbb{E}[\varphi(\mathbf{v})] - \log \mathbb{E}\left[e^{\varphi(\mathbf{u})}\right] &= \mathbb{E}[\varphi_a(\mathbf{v_a})] - \log \mathbb{E}\left[e^{\varphi_a(\mathbf{u_a})}\right] \\ &\leqslant \mathcal{D}(P_a \,\|\, Q_a).\end{aligned} \tag{3.41}$$

By taking the supremum, we get that $\mathcal{D}(P \,\|\, Q) \leqslant \mathcal{D}(P_a \,\|\, Q_a)$. The opposite inequality can be proven along the same lines. Indeed, let $\psi_a : \mathbb{R}^{h+k} \to \mathbb{R}$ be a measurable and bounded function. Define $\psi : \mathbb{R}^k \to \mathbb{R}$ by $\psi(x) := \psi_a(x, f(x))$. Then, ψ is measurable and bounded too, so that

$$\begin{aligned}\mathbb{E}[\psi_a(\mathbf{v_a})] - \log \mathbb{E}\left[e^{\psi_a(\mathbf{u_a})}\right] &= \mathbb{E}[\psi(\mathbf{v})] - \log \mathbb{E}\left[e^{\psi(\mathbf{u})}\right] \\ &\leqslant \mathbb{D}(P\|Q).\end{aligned} \tag{3.42}$$

In view of (3.39), we now get $\mathcal{D}(P_a \,\|\, Q_a) \leqslant \mathcal{D}(P \,\|\, Q)$. ∎

Now we can prove the main result:

Theorem 3.5.1. *Let* $\mathbf{y}$ *and* $\mathbf{z}$ *be as above. Assume that both* Φ_y *and* Φ_z *are piecewise continuous, coercive spectral densities. The following equality holds:*

$$\mathcal{D}_r(y\|z) = \mathcal{D}_r(d\hat{y}\|d\hat{z}). \tag{3.43}$$

Proof. In view of proposition 3.3.1, the last n components of $\hat{Y}_{2n}$ are functions (the complex conjugate) of the first n and the same holds for $\hat{Z}_{2n}$. Hence, in view of lemma 3.5.1, we have $\mathcal{D}(p\left(\hat{Y}_{2n}\right) \,\|\, p\left(\hat{Z}_{2n}\right)) = \mathcal{D}(p\left(\hat{Y}_n\right) \,\|\, p\left(\hat{Z}_n\right))$. Using again proposition 3.3.1, we have that the elements of $\hat{Y}_n$ are independent random vectors and the same holds for the elements of $\hat{Z}_n$. Hence, we have the following additive decomposition:

$$\mathcal{D}(p\left(\hat{Y}_{2n}\right) \,\|\, p\left(\hat{Z}_{2n}\right)) = \mathcal{D}(p\left(\hat{Y}_n\right) \,\|\, p\left(\hat{Z}_n\right)) = \sum_{k=0}^{n-1} \mathcal{D}\left(p(\delta\hat{y}_k)\|p(\delta\hat{z}_k)\right), \tag{3.44}$$

with $p(\delta\hat{y}_k)$ and $p(\delta\hat{z}_k)$ being the probability densities of the random vector $\delta\hat{y}_k = \hat{\mathbf{y}}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$ and $\delta\hat{z}_k = \hat{\mathbf{z}}(e^{j\vartheta_k}, e^{j\vartheta_{k+1}})$, respectively. Since $\delta\hat{y}_k$ and $\delta\hat{z}_k$ are jointly Gaussian and circularly symmetric, by (A.16) and (3.34)-(3.35), we get,

$$\begin{aligned}\mathcal{D}(p(\delta\hat{y}_k) \,\|\, p(\delta\hat{z}_k)) = \log\det\left[Q_y^{-1}(\vartheta_k, \vartheta_{k+1})Q_z(\vartheta_k, \vartheta_{k+1})\right] \\ + \operatorname{Tr}\left[Q_z^{-1}(\vartheta_k, \vartheta_{k+1})Q_y(\vartheta_k, \vartheta_{k+1})\right] - m,\end{aligned} \tag{3.45}$$

where, by virtue of the orthogonal increments property,

$$Q_y(\vartheta_k, \vartheta_{k+1}) := \int_{\vartheta_k}^{\vartheta_{k+1}} \Phi_y(e^{j\xi})d\xi, \tag{3.46}$$

$$Q_z(\vartheta_k, \vartheta_{k+1}) := \int_{\vartheta_k}^{\vartheta_{k+1}} \Phi_z(e^{j\xi})d\xi. \tag{3.47}$$

By piecewise continuity and the mean value theorem, we have that, except for a finite number of k's,

$$\begin{aligned}\mathcal{D}(p(\delta\hat{\mathbf{y}}_k)\,\|\,p(\delta\hat{z}_k)) &= \log\det\left[\left(\Phi_y(e^{j\vartheta_k})\frac{\pi}{n}\right)^{-1}\Phi_z(e^{j\bar{\vartheta}_k})\frac{\pi}{n}\right]\\ &\quad+\operatorname{Tr}\left[\left(\Phi_z(e^{j\bar{\vartheta}_k})\frac{\pi}{n}\right)^{-1}\Phi_y(e^{j\bar{\vartheta}_k})\frac{\pi}{n}\right]-m\\ &= \log\det[\Phi_y(e^{j\bar{\vartheta}_k})^{-1}\Phi_z(e^{j\bar{\vartheta}_k})]\\ &\quad+\operatorname{Tr}\left[\Phi_z(e^{j\bar{\vartheta}_k})^{-1}\Phi_y(e^{j\bar{\vartheta}_k})\right]-m,\end{aligned} \tag{3.48}$$

where $\vartheta_k \leqslant \bar{\vartheta}_k < \vartheta_{k+1}$. By employing the latter expression together with (3.44) and (3.37), we get

$$\begin{aligned}\mathcal{D}_r(d\hat{y}\|d\hat{z}) &= \lim_{n\to\infty}\frac{1}{2n}\mathcal{D}\left(\hat{Y}_n\|\hat{Z}_n\right)\\ &= \lim_{n\to\infty}\frac{1}{2n}\sum_{k=0}^{n-1}\mathbb{D}\left(p(\delta\hat{y}_k)\|p(\delta\hat{z}_k)\right)\\ &= \lim_{n\to\infty}\frac{1}{2n}\sum_{k=0}^{n-1}\log\det\Phi_y(e^{j\bar{\vartheta}_k})^{-1}\Phi_z(e^{j\bar{\vartheta}_k})\\ &\quad+\operatorname{Tr}\left[\Phi_z^{-1}(e^{j\bar{\vartheta}_k})\left(\Phi_y(e^{j\bar{\vartheta}_k})-\Phi_z(e^{j\bar{\vartheta}_k})\right)\right]\\ &= \lim_{n\to\infty}\frac{1}{2\pi}\sum_{k=0}^{n-1}\Big\{\log\det\Phi_y(e^{j\bar{\vartheta}_k})^{-1}\Phi_z(e^{j\bar{\vartheta}_k})\\ &\quad+\operatorname{Tr}\left[\Phi_z^{-1}(e^{j\bar{\vartheta}_k})\left(\Phi_y(e^{j\bar{\vartheta}_k})-\Phi_z(e^{j\bar{\vartheta}_k})\right)\right]\Big\}\frac{\pi}{n}\\ &= \frac{1}{2\pi}\int_0^{\pi}\left\{\log\det\left(\Phi_y^{-1}(e^{j\vartheta})\Phi_z(e^{j\vartheta})\right)\right.\\ &\quad\left.+\operatorname{Tr}\left[\Phi_z^{-1}(e^{j\vartheta})\left(\Phi_y(e^{j\vartheta})-\Phi_z(e^{j\vartheta})\right)\right]\right\}d\vartheta\\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log\det\left(\Phi_y^{-1}(e^{j\vartheta})\Phi_z(e^{j\vartheta})\right)\right.\\ &\quad\left.+\operatorname{Tr}\left[\Phi_z^{-1}(e^{j\vartheta})\left(\Phi_y(e^{j\vartheta})-\Phi_z(e^{j\vartheta})\right)\right]\right\}d\vartheta,\end{aligned} \tag{3.49}$$

which, by (3.5), is (3.43). ■

Remark 3.5.1. As is well known, the fundamental property of the Fourier transform is that it is isometric. The above result may be interpreted as a further invariance principle of the Fourier transform: the relative entropy rate is the same in the time and spectral domain.

3.6 RER ESTIMATION: PRELIMINARY RESULTS

Based on the choice of (3.32) as a metric for evaluating the "distance" between multivariate spectral densities, we propose a new THREE-like spectral estimator for multivariate processes, the so-called RER (Relative Entropy Rate) estimator. It solves the following

Problem 3.6.1 (RER constrained spectrum approximation problem)**.** Let $\Psi \in \mathcal{S}_+^{m\times m}$, $G(z)$ as in (3.12) and $\Sigma = \Sigma^\top > 0$. Find Φ° that solves

$$\text{minimize } d_{RER}(\Phi,\Psi) \text{ over } \left\{\Phi \in \mathcal{S}_+^{m\times m} \mid \int G\Phi G^* = \Sigma\right\}. \tag{3.50}$$

Remark 3.6.1. Alternatively, we could also minimize the distance index (3.32) with respect to the second argument. This would be consistent with the usual choice in dealing with some minimum prediction error and model reduction problems, see Lindquist and Picci [73]. However, this approach is not suitable for our purposes. Indeed, it may lead to a non rational solution, even if the prior Ψ is rational. Consider

$$d_{RER}(\Psi,\Phi) = \int \frac{1}{2}\left[\log\frac{\det\Phi}{\det\Psi} + \mathrm{Tr}(\Phi^{-1}\Psi) - m\right]. \tag{3.51}$$

The corresponding Lagrangian function is defined by

$$\tilde{L}_\Psi(\Phi,\Lambda) = \int \left[\log\frac{\det(\Phi)}{\det(\Psi)} + \mathrm{Tr}(\Phi^{-1}\Psi) + \mathrm{Tr}(\Lambda G\Phi G^*)\right] - \mathrm{Tr}\Lambda. \tag{3.52}$$

and the first directional derivative can be written as

$$\delta\tilde{L}_\Psi(\Phi,\Lambda;\delta\Phi) = \int \mathrm{Tr}\left\{\left[\Phi^{-1} - \Phi^{-1}\Psi\Phi^{-1} + G^*\Lambda G\right]\delta\Phi\right\}. \tag{3.53}$$

The condition that has to be satisfied by stationary points, i.e. $\delta\tilde{L}_\Psi(\Phi,\Lambda;\delta\Phi) = 0$, for all $\delta\Phi \in C(\mathbb{T})$, is equivalent to ask that

$$\Phi^{-1} - \Phi^{-1}\Psi\Phi^{-1} + G^*\Lambda G = 0. \tag{3.54}$$

This equality allows to obtain an expression for stationary points in terms of Λ. Indeed,

$$\Phi - \Psi + \Phi G^*\Lambda G\Phi = 0 \tag{3.55}$$

Assume W_Λ such that $G^*\Lambda G = {W_\Lambda}^* W_\Lambda$. Then

$$\begin{aligned} 0 &= \Phi - \Psi + \Phi G^*\Lambda G\Phi \\ &= \underbrace{W_\Lambda\Phi{W_\Lambda}^*}_{\Phi_\Lambda} - \underbrace{W_\Lambda\Psi{W_\Lambda}^*}_{\Psi_\Lambda} + W_\Lambda\Phi{W_\Lambda}^*W_\Lambda\Phi{W_\Lambda}^* \\ &= \Phi_\Lambda - \Psi_\Lambda + {\Phi_\Lambda}^2. \end{aligned} \tag{3.56}$$

Equation (3.56) is solved by choosing

$$\Phi_\Lambda = -\frac{1}{2}I + \left(\frac{1}{4}I + \Psi_\Lambda\right)^{\frac{1}{2}}, \tag{3.57}$$

as can be proven by substitution:

$$-\frac{1}{2}I + \left(\frac{1}{4}I + \Psi_\Lambda\right)^{\frac{1}{2}} - \Psi_\Lambda + \frac{1}{4}I - \left(\frac{1}{4}I + \Psi_\Lambda\right)^{\frac{1}{2}} + \frac{1}{4}I + \Psi_\Lambda = 0 \tag{3.58}$$

Therefore, we can obtain the following expression for the optimal solution Φ°

$$\Phi^\circ = -\frac{1}{2}(G^*\Lambda G)^{-1} + {W_\Lambda}^{-1}\left(\frac{1}{4}I + W_\Lambda \Psi {W_\Lambda}^*\right)^{\frac{1}{2}} {W_\Lambda}^{-*} \tag{3.59}$$

which may be not rational, even for rational Ψ.

3.6.1 Feasibility

First we deal with the issue of feasibility of problem 3.6.1: Does exist $\Phi \in \mathcal{S}_+^{m\times m}(\mathbb{T})$ satisfying (3.14), where G is the transfer function of the bank of filters (3.12) and Σ is the steady-state covariance of the output process?

A major role is played by the operator (3.4). The following result can be obtained along the same lines of Georgiou [54] (see also Ramponi, Ferrante, and Pavon [89]).[1]

Theorem 3.6.1. *Consider $\Sigma = \Sigma^\top \in \mathbb{R}^{n\times n}$ and a filter defined as in (3.12). Then:*

1. *Σ is in* Range(Γ) *if and only if there exists $H \in \mathbb{R}^{m\times n}$ such that*

$$\Sigma - A\Sigma A^\top = BH + H^\top B^\top. \tag{3.60}$$

2. *Let the $\Sigma \in \mathbb{R}^{n\times n}$ be positive definite. Then, there exists $H \in \mathbb{R}^{m\times n}$ that solves (3.60) if and only if there exists $\Phi \in \mathcal{C}_{R+}^{m\times m}$ such that $\Gamma(\Phi) = \Sigma$.*

From now on we assume feasibility of problem 3.6.1. In view of the previous result, this is equivalent to the fact that Equation (3.60) admits a solution $\bar{H}$. Moreover, to simplify the exposition, we assume that $\Sigma = I$. This can be done without loss of generality. In fact, if $\Sigma \neq I$, it suffices to replace G with $G' := \Sigma^{-1/2}G$ and (A, B) with $(A' = \Sigma^{-1/2}A\Sigma^{1/2}, B' = \Sigma^{-1/2}B)$ to obtain an equivalent problem where $\Sigma = I$.

1 In Georgiou [54] the general case was considered when $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$ and the process $\mathbf{y}(t)$ is complex-valued, too. In that case, it was proven that the Hermitian matrix $\Sigma \in \mathbb{C}^{n\times n}$ belongs to Range(Γ) if and only if there exists $H \in \mathbb{C}^{m\times n}$ solving the feasibility equation $\Sigma - A\Sigma A^* = BH + H^*B^*$.

3.6.2 Form of the optimum solution

For the sake of simplicity, notice that problem 3.6.1 is equivalent to minimizing, over $\mathcal{S}_+^{m\times m}$,

$$2d_{RER}(\Phi,\Psi)+m=\int\left\{\log\det\left(\Phi^{-1}\Psi\right)+\mathrm{Tr}\left(\Psi^{-1}\Phi\right)\right\}, \tag{3.61}$$

subject to (3.14). Recall that the inner product in $\mathcal{Q}_n$ is defined by $\langle M,N\rangle=\mathrm{Tr}[MN]$. Thus, the Lagrangian reads

$$\begin{aligned}L_\Psi(\Phi,\Lambda)&=2d_{RER}(\Phi,\Psi)+m+\langle\Lambda,\int G\Phi G^*-\Sigma\rangle\\&=\int\left[\log\frac{\det(\Psi)}{\det(\Phi)}+\mathrm{Tr}(\Psi^{-1}\Phi)+\mathrm{Tr}(\Lambda G\Phi G^*)\right]-\mathrm{Tr}\Lambda,\end{aligned}$$

where it is assumed that $\Sigma=I$ and the Lagrange parameter $\Lambda\in\mathcal{Q}_n$. Notice that each $\Lambda\in\mathcal{Q}_n$ can be uniquely decomposed as $\Lambda=\Lambda_\Gamma+\Lambda_\perp$, where $\Lambda_\Gamma\in\mathrm{Range}\,(\Gamma)$ and $\Lambda_\perp\in(\mathrm{Range}\,(\Gamma))^\perp$. Each $\Lambda_\perp\in(\mathrm{Range}\,(\Gamma))^\perp$ is such that $G^*(e^{j\vartheta})\Lambda_\perp G(e^{j\vartheta})\equiv 0$ (see Ramponi, Ferrante, and Pavon [89, Section III]). Moreover, $\mathrm{Tr}[\Lambda_\perp]=\langle\Lambda_\perp,I\rangle=0$, because $I\in\mathrm{Range}\,(\Gamma)$ in view of the feasibility assumption. Hence, a term $\Lambda_\perp\in(\mathrm{Range}\,(\Gamma))^\perp$ gives no contribution to the Lagrangian (3.62). Therefore, from now on, we will assume that the Lagrange parameter Λ belongs to $\mathrm{Range}(\Gamma)$.

Now we aim at computing the form of the optimum solution. Notice that $L_\Psi(\cdot,\Lambda)$ in (3.62) is strictly convex on $\mathcal{S}_+^{m\times m}$. thus, in order to find the minimizer, we impose the first variation to be zero in each direction $\delta\Phi\in L_2^{m\times m}$. Recall that, for a positive definite matrix X, the directional derivative of $\log\det(X)$ in direction δX is given by

$$\delta\log\det(X;\delta X)=\mathrm{Tr}(X^{-1}\delta X). \tag{3.62}$$

Thus, we find

$$\begin{aligned}\delta L(\Phi,\Lambda;\delta\Phi)&=\int\left[-\mathrm{Tr}(\Phi^{-1}\delta\Phi)+\mathrm{Tr}(\Psi^{-1}\delta\Phi)+\mathrm{Tr}(G^*\Lambda G\delta\Phi)\right]\\&=\int\langle-\Phi^{-1}+\Psi^{-1}+G^*\Lambda G,\delta\Phi\rangle.\end{aligned} \tag{3.63}$$

Since $\left[-\Phi^{-1}+\Psi^{-1}+G^*\Lambda G\right]\in L_2^{m\times m}$, (3.63) vanishes $\forall\,\delta\Phi\in L_2^{m\times m}$ if and only if

$$\Phi=\Phi^\circ(\Lambda):=\left[\Psi^{-1}+G^*\Lambda G\right]^{-1}. \tag{3.64}$$

Let W_Ψ be the stable and minimum phase spectral factor of Ψ,[2] and $G_1(e^{j\vartheta})$ be defined by

$$G_1(e^{j\vartheta}):=G(e^{j\vartheta})W_\Psi(e^{j\vartheta}). \tag{3.65}$$

2 Since $\Psi\in\mathcal{S}_+^{m\times m}$, W_Ψ exists. It is unique up to multiplication on the right by a constant orthogonal matrix.

We can also compute an alternative form of (3.64)

$$\Phi^\circ(\Lambda) = W_\Psi(I + G_1^* \Lambda G_1)^{-1} W_\Psi^*. \tag{3.66}$$

Since Φ° is required to be a bounded spectral density, we need, as indicated by (3.66), to restrict the Lagrange multiplier Λ to the subset $\mathcal{L}_+$, where

$$\mathcal{L}_+ := \{\Lambda \in \mathcal{Q}_n \,|\, I + G_1^* \Lambda G_1 > 0 \ \text{ a.e. on } \mathbb{T}\}. \tag{3.67}$$

In conclusion, the natural set for the Lagrangian multiplier Λ is

$$\mathcal{L}_+^\Gamma := \mathcal{L}_+ \cap \text{Range}\,(\Gamma). \tag{3.68}$$

3.6.3 Complexity of the solution

Notice that (3.64) yields an upper bound on the McMillan degree $\deg[\Phi^\circ]$ of the optimum approximant Φ°. Indeed, it follows from (3.64) that

$$\deg[\Phi^\circ] \leqslant \deg[\Psi] + 2n, \tag{3.69}$$

where n is the McMillan degree of $G(z)$. The McMillan degree can be regarded as a measure of complexity. It is worthwhile that RER estimation improves on the best so far available upper bound on the complexity of the solution in the framework of THREE-like multivariate spectral estimation. Indeed, the method proposed in Ferrante, Pavon, and Ramponi [40], which hinges on a multivariate extension of Hellinger distance (see also Section 3.4), entails a complexity upper bound which is equal to $\deg[\Psi] + 4n$.

In light of (3.69), it emerges that better resolution – which requires $G(z)$ to have poles in the prescribed frequency range, and thus higher degree n, see Section 3.9 – can be attained at a price of higher complexity of the solution. At the same time, the *prior* acts as a tuning parameter, too, since it affects the McMillan degree of Φ°.

3.7 THE DUAL PROBLEM

To sum up, the main result of the previous section is that for each $\Lambda \in \mathcal{L}_+^\Gamma$ there exists a unique $\Phi^\circ \in \mathcal{S}_+^{m\times m}$, whose form is given by (3.64), that minimizes the Lagrangian functional. The solution has maximum McMillan degree equal to $2n + \deg \Psi$. If we can find a Λ° s.t. $\Phi^\circ(\Lambda^\circ)$ satisfies the integral constraints (3.14), such a $\Phi^\circ(\Lambda^\circ)$ is the solution of problem 3.6.1. Thus, we resort to duality theory: Indeed, the dual problem

is finite dimensional, in contrast with the primal one. Moreover, duality allows us to prove uniqueness and existence of the solution to Problem 3.6.1.

Instead of maximizing

$$\inf_{\Phi} L(\Phi, \Lambda) = L(\Phi^\circ, \Lambda) = \int \log\det(I + G_1^* \Lambda G_1) + n - \mathrm{Tr}\Lambda, \tag{3.70}$$

we will equivalently minimize the *dual functional*:

$$\begin{aligned} J_\Psi(\Lambda) &:= -L(\Phi^\circ(\Lambda), \Lambda) + n \\ &= \int \left[\mathrm{Tr}\Lambda - \log\det(I + G_1^* \Lambda G_1)\right]. \end{aligned} \tag{3.71}$$

Recall that, given a matrix $A = A^* > 0$, we have $\int \log\det A = \int \mathrm{Tr}\log A$. Hence, we can express the dual functional also as

$$J_\Psi(\Lambda) = \int \mathrm{Tr}\left[\Lambda - \log(I + G_1^* \Lambda G_1)\right]. \tag{3.72}$$

Given $\delta\Lambda \in \mathcal{Q}_n$, by means of (3.62) we can evaluate its first variation:

$$\begin{aligned} \delta J_\Psi(\Lambda; \delta\Lambda) &= \nabla J_{\Psi,\Lambda}(\delta\Lambda) \\ &= \int \left\{ \mathrm{Tr}\left[\delta\Lambda\right] - \mathrm{Tr}\left[(I + G_1^* \Lambda G_1)^{-1} G_1^* \delta\Lambda G_1 \right] \right\}. \end{aligned} \tag{3.73}$$

The results of this section show that there exists a unique $\Lambda^\circ \in \mathcal{L}_+^\Gamma$ minimizing $J_\Psi(\Lambda)$ in (5.44). Such a Λ° annihilates the directional derivative (3.73) in any direction $\delta\Lambda \in \mathcal{Q}_n$, namely

$$\langle I - \int G_1 (I + G_1^* \Lambda^\circ G_1)^{-1} G_1^*, \delta\Lambda \rangle = 0 \quad \forall \delta\Lambda \in \mathcal{Q}_n, \tag{3.74}$$

or, equivalently,

$$I = \int G_1 (I + G_1^* \Lambda^\circ G_1)^{-1} G_1^* = \int G \Phi^\circ(\Lambda^\circ) G^*. \tag{3.75}$$

This means that the corresponding spectral density $\Phi^\circ := \Phi(\Lambda^\circ) = \left[\Psi^{-1} + G^* \Lambda^\circ G\right]^{-1}$, satisfies constraint (3.14) and is therefore the unique solution of problem 3.6.1.

Uniqueness of the minimizing $\Lambda_0 \in \mathcal{L}_+^\Gamma$ is an obvious consequence of the following result.

Theorem 3.7.1. *The dual functional* $J_\Psi(\Lambda)$ *belongs to* $\mathcal{C}^2(\mathcal{L}_+^\Gamma)$ *and is* strictly *convex on* $\mathcal{L}_+^\Gamma$.

Proof. Consider a sequence $M_n \in \mathrm{Range}\,(\Gamma)$, such that $M_n \to 0$, and define, for $N \in \mathcal{Q}_n$, $Q_N(z) = I + G_1^*(z) N G(z)$. By Lemma 5.2 in Ramponi, Ferrante, and Pavon [89],

$Q_{\Lambda+M_n}^{-1}$ converges uniformly to Q_Λ^{-1}, so that it is bounded above. Hence, applying the bounded convergence theorem, we get

$$\lim_{n\to\infty} \int \mathrm{Tr}\left[Q_{\Lambda+M_n}^{-1} G_1^* \delta\Lambda G_1\right] = \int \mathrm{Tr}\left[Q_\Lambda^{-1} G_1^* \delta\Lambda G_1\right], \tag{3.76}$$

so that $J_\Psi(\Lambda)$ belongs to $\mathcal{C}^1(\mathcal{L}_+^\Gamma)$. Consider now the second variation. Let us denote the matrix inversion operator by $R : M \mapsto M^{-1}$ and recall that its first derivative in direction δM is given by $\delta R(M, \delta M) = -M^{-1}\delta M M^{-1}$. Then, for $\delta\Lambda_1$ and $\delta\Lambda_2$ in $\mathcal{Q}_n$, we have

$$\delta^2 J_\Psi(\Lambda; \delta\Lambda_1, \delta\Lambda_2) = \int \mathrm{Tr}\left[(I + G_1^*\Lambda G_1)^{-1} G_1^* \delta\Lambda_2 G_1 (I + G_1^*\Lambda G_1)^{-1} G_1^* \delta\Lambda_1 G_1\right], \tag{3.77}$$

so that $J_\Psi(\Lambda)$ is $\mathcal{C}^2(\mathcal{L}_+^\Gamma)$. The bilinear form $H_\Lambda(\cdot,\cdot) := \delta^2 J_\Psi(\Lambda; \cdot, \cdot)$ is the *Hessian* of J_Ψ at Λ. For $\delta\Lambda \in \mathrm{Range}(\Gamma)$, which implies that $(\Lambda + \varepsilon\delta\Lambda) \in \mathcal{L}_+^\Gamma$ for sufficiently small ε, consider $H_\Lambda(\delta\Lambda, \delta\Lambda) = \delta^2 J_\Psi(\Lambda; \delta\Lambda, \delta\Lambda)$. We get

$$\begin{aligned} H_\Lambda(\delta\Lambda, \delta\Lambda) &= \int \mathrm{Tr}\left[(I + G_1^*\Lambda G_1)^{-1} G_1^* \delta\Lambda G_1 (I + G_1^*\Lambda G_1)^{-1} G_1^* \delta\Lambda G_1\right] \\ &= \int \mathrm{Tr}\left[Q_\Lambda^{-\frac{1}{2}} G_1^* \delta\Lambda G_1 Q_\Lambda^{-1} G_1^* \delta\Lambda G_1 Q_\Lambda^{-\frac{1}{2}}\right] \end{aligned} \tag{3.78}$$

which vanishes if and only if the integrand is identically zero. Moreover $G_1^*\delta\Lambda G_1 = W_\Psi^* G^* \delta\Lambda G W_\Psi$ is identically zero on $\mathbb{T}$ if and only if $\delta\Lambda \in \mathrm{Range}(\Gamma)^\perp$. On the other hand we have assumed $\delta\Lambda \in \mathrm{Range}(\Gamma)$, so that the integrand is identically zero if and only if $\delta\Lambda = 0$. In conclusion, the Hessian is positive-definite and the dual functional is strictly convex on $\mathcal{L}_+^\Gamma$. ■

The next and most delicate step is to prove that, although the set $\mathcal{L}_+^\Gamma$ is open and unbounded, a Λ° minimizing J_Ψ over $\mathcal{L}_+^\Gamma$ *does* exist. To this aim, first we prove that the function $J_\Psi(\Lambda)$ is inf-compact, i.e. $\forall\, \alpha \in \mathbb{R}$, the set $\left\{\Lambda \in \mathcal{L}_+^\Gamma \,|\, J_\Psi(\Lambda) \leqslant \alpha\right\}$ is compact. To establish this fact, define $\overline{\mathcal{L}_+^\Gamma}$ to be the closure of $\mathcal{L}_+^\Gamma$, i.e. the set

$$\overline{\mathcal{L}_+^\Gamma} = \left\{\Lambda = \Lambda^\top \in \mathbb{R}^{n\times n} \,|\, \Lambda \in \mathrm{Range}(\Gamma),\, I + G_1^*\Lambda G_1 \geqslant 0,\, \forall e^{j\vartheta} \in \mathbb{T}\right\}. \tag{3.79}$$

Given that, for Λ belonging to the boundary $\partial\mathcal{L}_+^\Gamma$, the Hermitian matrix $I + G_1^*\Lambda G_1$ is singular, in at least one point of $\mathbb{T}$, it is useful to introduce the following sequence of functions on $\overline{\mathcal{L}_+^\Gamma}$:

$$J_\Psi^n(\Lambda) = \int \mathrm{Tr}\left[\Lambda - \log\left(I + G_1^*\Lambda G_1 + \frac{1}{n} I\right)\right], \quad n \geqslant 1. \tag{3.80}$$

Recall that a real-valued function f is said to be lower semicontinuous at x_0 if, $\forall\, \varepsilon > 0$, there exists a neighborhood U of x_0 such that, $\forall\, x \in U$, $f(x) \geqslant f(x_0) - \varepsilon$. Recall also that, given $f : \mathbb{R}^{n\times n} \to \mathbb{R}$, its *epigraph* epi (f) is defined by

$$\text{epi}\,(f) := \left\{(x, a) \in \mathbb{R}^{n\times n} \times \mathbb{R} \,|\, a \geqslant f(x)\right\}. \tag{3.81}$$

Moreover, f is a lower semicontinuous (convex) function if and only if its epigraph is closed (convex), see e.g. Rockafellar [90]. The following lemmata allow to conclude that $J_\Psi(\Lambda)$ is inf-compact over $\overline{\mathcal{L}_+^\Gamma}$.

Lemma 3.7.1. *The pointwise limit $J_\Psi^\infty(\Lambda)$, defined as $J_\Psi^\infty(\Lambda) := \lim_{n\to\infty} J_\Psi^n(\Lambda)$, exists and is a lower semicontinuous and convex function defined over $\overline{\mathcal{L}_+^\Gamma}$, with values in the extended reals.*

Proof. The additive term $\frac{1}{n}I$ ensures that, for each n, $J_\Psi^n(\Lambda)$ is a continuous and convex function of Λ on the closed set $\overline{\mathcal{L}_+^\Gamma}$. From the properties of $J_\Psi^n(\Lambda)$, it follows that epi $(J_\Psi^n(\Lambda))$ is a closed and convex subset of $\mathbb{R}^{n\times n} \times \mathbb{R}$. In addition, the pointwise sequence is monotonically increasing, since $J_\Psi^n(\Lambda) < J_\Psi^{n+1}(\Lambda)$. Therefore, it converges to $J_\Psi^\infty(\Lambda) := \sup_n J_\Psi^n(\Lambda)$. Since the intersection of closed sets is closed and the intersection of convex sets is convex, epi $J_\Psi^\infty(\Lambda) = \cap_n$epi $J_\Psi^n(\Lambda)$ is closed and convex. As a consequence, $J_\Psi^\infty(\Lambda)$ is lower semicontinuous and convex. ∎

Lemma 3.7.2. *Assume that the feasibility condition (3.60) holds. Given $\Lambda \in \mathcal{L}_+^\Gamma$, there exist two real constants $\mu > 0$ and α such that:*

$$\text{Tr}\,[\Lambda] \geqslant \mu \text{Tr}\left[\int (G_1^* \Lambda G_1 + I)\right] + \alpha. \tag{3.82}$$

Proof. Since $\Sigma = I$, by feasibility, there exists $\Phi_I \in \mathcal{S}_+^{m\times m}$ such that $\int G\Phi_I G^* = I$. Thus,

$$\begin{aligned}
\text{Tr}\,[\Lambda] &= \text{Tr}\left[\int G\Phi_I G^* \Lambda\right] \\
&= \text{Tr}\left[\int G^* \Lambda G \Phi_I\right] \\
&= \text{Tr}\left[\int W_\Psi^* G^* \Lambda G W_\Psi W_\Psi^{-1} \Phi_I W_\Psi^{-*}\right] \\
&= \text{Tr}\left[\int G_1^* \Lambda G_1 \Xi\right],
\end{aligned} \tag{3.83}$$

where the cyclic property of the trace was employed and the auxiliary spectral density $\Xi := W_\Psi^{-1} \Phi_I W_\Psi^{-*}$ has been defined. By defining $\alpha := -\text{Tr}\left[\int \Xi\right]$, it follows that

$$\begin{aligned}
\text{Tr}\,[\Lambda] &= \text{Tr}\left[\int (G_1^* \Lambda G_1 + I)\Xi\right] - \text{Tr}\left[\int \Xi\right] \\
&= \text{Tr}\left[\int (G_1^* \Lambda G_1 + I)\Xi\right] + \alpha.
\end{aligned} \tag{3.84}$$

Let Δ be such that $(G_1^*\Lambda G_1 + I) = \Delta^*\Delta$ (recall that we are assuming $\Lambda \in \mathcal{L}_+^\Gamma$ so that $G_1^*\Lambda G_1 + I$ is positive definite on $\mathbb{T}$ and admits a right spectral factor Δ) so that $\mathrm{Tr}\left[(G_1^*\Lambda G_1 + I)\Xi\right] = \mathrm{Tr}\left[\Delta\Xi\Delta^*\right]$. Given that $\Xi = W_\Psi^{-1}\Phi_I W_\Psi^{-*}$ is a coercive spectrum, because both Φ_I and Ψ belong to $\mathcal{S}_+^{m\times m}$, there exists $\mu > 0$ s.t. $\Xi(e^{j\vartheta}) \geqslant \mu I$, $\forall\, e^{j\vartheta} \in \mathbb{T}$. Recalling that the trace and the integral are monotonic functionals, it is possible to conclude that

$$\begin{aligned} \mathrm{Tr}\left[\Lambda\right] &= \mathrm{Tr}\left[\int (G_1^*\Lambda G_1 + I)\Xi\right] + \alpha \\ &\geqslant \mu \mathrm{Tr}\left[\int (G_1^*\Lambda G_1 + I)\right] + \alpha. \end{aligned} \tag{3.85}$$

■

Lemma 3.7.3. *Let* $\mathcal{B} := \left\{\Lambda \in \partial\mathcal{L}_+^\Gamma \,|\, \det\left(G_1^*\Lambda G_1 + I\right) = 0, \quad \forall e^{j\vartheta} \in \mathbb{T}\right\}$ *and consider its complement set* $\mathcal{B}^c := \left\{\Lambda \in \partial\mathcal{L}_+^\Gamma \,|\, \Lambda \notin \mathcal{B}\right\}$. *Then, under feasibility assumption:*

1. $J_\Psi^\infty(\Lambda)$ *is bounded from below on* $\overline{\mathcal{L}_+^\Gamma}$;
2. $J_\Psi^\infty(\Lambda) = J_\Psi(\Lambda)$ *on* $\mathcal{L}_+^\Gamma$;
3. $J_\Psi^\infty(\Lambda)$ *is finite over* $\mathcal{B}^c$.

The proof can be found in the Appendix.

Proof.

1. As a consequence of the previous lemma,

$$\begin{aligned} J_\Psi^n(\Lambda) &= \int \mathrm{Tr}\left[\Lambda - \log(I + G_1^*\Lambda G_1 + \frac{1}{n}I)\right] \\ &\geqslant \int \mathrm{Tr}\left[\mu(I + G_1^*\Lambda G_1) - \log(I + G_1^*\Lambda G_1 + \frac{1}{n}I)\right] + \alpha. \end{aligned} \tag{3.86}$$

   Let $\{x_i\}$ be the eigenvalues of $(I + G_1^*\Lambda G_1)$. Then,

$$\begin{aligned} J_\Psi^n(\Lambda) &= \int \mathrm{Tr}\left[\mu(I + G_1^*\Lambda G_1) - \log(I + G_1^*\Lambda G_1 + \frac{1}{n}I)\right] + \alpha \\ &= \int \mu \sum_{i=1}^m x_i - \sum_{i=1}^m \log\left(x_i + \frac{1}{n}\right) + \alpha = \int \rho\left(x_1, \ldots, x_m\right) + \alpha, \end{aligned} \tag{3.87}$$

   where $\rho(x_1, \ldots, x_m) := \mu \sum_{i=1}^m x_i - \sum_{i=1}^m \log\left(x_i + \frac{1}{n}\right)$. Moreover,

$$\frac{\partial}{\partial x_i}\left[\rho(x_1, \ldots, x_m)\right] = \mu - \frac{1}{x_i + \frac{1}{n}} \quad \forall\, i. \tag{3.88}$$

The minimum of ρ is thus attained by choosing $x_i = \frac{1}{\mu} - \frac{1}{n}, \forall\, i$. Therefore,

$$\rho(x_1, \ldots, x_m) \geqslant m - \frac{\mu m}{n} + m \log \mu \tag{3.89}$$

The fact that $J^n_\Psi(\Lambda)$ is bounded from below over $\overline{\mathcal{L}^\Gamma_+}$ now follows:

$$J^n_\Psi(\Lambda) \geqslant \alpha + m + m \log \mu - \frac{\mu m}{n} \geqslant \alpha + m + m \log \mu. \tag{3.90}$$

2. Beppo Levi's Theorem allows to conclude that $J^\infty_\Psi(\Lambda) = J_\Psi(\Lambda)$ in $\mathcal{L}^\Gamma_+$:

$$J^\infty_\Psi(\Lambda) = \int \mathrm{Tr}[\Lambda] - \int \mathrm{Tr}\left[\lim_{n\to\infty} \log(I + G_1^* \Lambda G_1 + \frac{1}{n} I)\right] = J_\Psi(\Lambda). \tag{3.91}$$

3. Since, for $\Lambda \in \mathcal{B}^c$, the rational function $\det(I + G_1^* \Lambda G_1)$ is not identically zero, its logarithm is integrable over $\mathbb{T}$. Hence, $J^\infty_\Psi(\Lambda)$ is finite. $J^\infty_\Psi(\Lambda) = +\infty$ instead for $\Lambda \in \mathcal{B}$.

■

Lemma 3.7.4. *If the feasibility hypothesis holds, then, for $\Lambda \in \mathcal{L}^\Gamma_+$,*

$$\lim_{\|\Lambda\| \to +\infty} J_\Psi(\Lambda) = +\infty. \tag{3.92}$$

Proof. In view of lemma 3.7.2

$$\mathrm{Tr}\,[\Lambda] \geqslant \mu \mathrm{Tr}\left[\int (G_1^* \Lambda G_1 + I)\right] + \alpha > \alpha, \tag{3.93}$$

so that $\mathrm{Tr}\,[\Lambda]$ is bounded from below. Consider a sequence $\{\Lambda_k\}_{k\in\mathbb{N}} \in \mathcal{L}^\Gamma_+$, such that

$$\lim_{k\to\infty} \|\Lambda_k\| = +\infty. \tag{3.94}$$

Let $\Lambda^0_k := \frac{\Lambda_k}{\|\Lambda_k\|}$. Since $\mathcal{L}^\Gamma_+$ is convex and $\Lambda = 0$ belongs to $\mathcal{L}^\Gamma_+$, $\forall\, \xi \in [0,1]$, $\xi\Lambda \in \mathcal{L}^\Gamma_+$. Therefore $\Lambda^0_k \in \mathcal{L}^\Gamma_+$ for sufficiently large k. Let $\eta := \liminf \mathrm{Tr}\left[\Lambda^0_k\right]$ In view of (3.93)

$$\mathrm{Tr}\Lambda^0_k = \frac{1}{\|\Lambda_k\|}\mathrm{Tr}\Lambda_k > \frac{1}{\|\Lambda_k\|}\alpha \to 0, \tag{3.95}$$

for $\|\Lambda_k\| \to \infty$, so $\eta \geqslant 0$. Thus, the sequence $\left\{\Lambda^0_k\right\}$ has a subsequence such that the limit of its trace is η. Given that Λ^0_k belongs to the surface of the unit ball, which is compact, the subsequence contains a subsubsequence $\left\{\Lambda^0_{k_m}\right\}_{k_m\in\mathbb{N}}$ that is convergent. Define

$$\Lambda^0_\infty := \lim_{k_m\to\infty} \Lambda^0_{k_m}. \tag{3.96}$$

The next step is to prove that $\Lambda_\infty^0 \in \mathcal{L}_+^\Gamma$. To this aim, notice that Λ_∞^0 is the limit of a convergent sequence in the finite-dimensional linear space Range (Γ). Therefore it belongs to Range (Γ). Moreover, recall that the primary sequence $\{\Lambda_k\}_{k\in\mathbb{N}}$ has elements belonging to $\mathcal{L}_+^\Gamma$. It means that, for each Λ_k, $(I + G_1^* \Lambda_k G_1) > 0$. As a consequence, it holds that, for each m,

$$\left(\frac{1}{\|\Lambda_{k_m}\|} I + G_1^* \Lambda_{k_m}^0 G_1\right) > 0 \quad \text{on } \mathbb{T}\,. \tag{3.97}$$

Taking the pointwise limit for $m \to \infty$, it results that $G_1^* \Lambda_\infty^0 G_1$ is positive semidefinite on $\mathbb{T}$, and so $(I + G_1^* \Lambda_\infty^0 G_1)$ is strictly positive definite on $\mathbb{T}$. Therefore, $\Lambda_\infty^0 \in \mathcal{L}_+^\Gamma$.

The next step is to prove that $\mathrm{Tr}\Lambda_\infty^0 > 0$. If the feasibility condition (3.60) holds, there exists Φ_I such that $I = \int G\Phi_I G^*$. Therefore, it is possible to write:

$$\begin{aligned} \mathrm{Tr}\Lambda_\infty^0 &= \mathrm{Tr}\int G\Phi_I G^* \Lambda_\infty^0 = \int \mathrm{Tr}\left[W_\Psi^{-*} W_\Psi^* G^* \Lambda_\infty^0 G W_\Psi W_\Psi^{-1} \Phi_I\right] \\ &= \int \mathrm{Tr}\left[G_1^* \Lambda_\infty^0 G_1 \underbrace{W_\Psi^{-1} \Phi_I W_\Psi^{-*}}_{\Xi}\right] = \int \mathrm{Tr}\left[\Xi^{\frac{1}{2}} G_1^* \Lambda_\infty^0 G_1 \Xi^{\frac{1}{2}}\right], \end{aligned} \tag{3.98}$$

where the coercive spectral density Ξ is defined as in lemma 3.7.2. Since $G_1^* \Lambda_\infty^0 G_1 \geqslant 0$, in order to prove that $\mathrm{Tr}\left[\Lambda_\infty^0\right]$ is positive, in view of (3.98) it is sufficient to show that $G_1^* \Lambda_\infty^0 G_1$ is not identically zero. Assume by contradiction that $G_1^* \Lambda_\infty^0 G_1 \equiv 0$. As a consequence, $\forall e^{j\vartheta} \in \mathbb{T}$,

$$0 \equiv G_1^* \Lambda_\infty^0 G_1 = W_\Psi^* G^* \Lambda_\infty^0 G W_\Psi. \tag{3.99}$$

Therefore, $G^* \Lambda_\infty^0 G \equiv 0$. However, this means that $\Lambda_\infty^0 \in \mathrm{Range}(\Gamma)^\perp$. But it has already been proven that $\Lambda_\infty^0 \in \mathrm{Range}(\Gamma)$. Moreover, $\Lambda_\infty^0 \neq 0$, since it belongs to the surface of the unit ball. This is a contradiction. Thus, $G_1^* \Lambda_\infty^0 G_1$ is not identically zero, and from (3.98) it follows that $\eta = \mathrm{Tr}\Lambda_\infty^0 > 0$. It follows that there exists K such that $\mathrm{Tr}\Lambda_k^0 > \frac{\eta}{2}$ for all $k > K$. Notice that $G_1^* G_1$ is positive definite on $\mathbb{T}$ (and indeed coercive). Moreover, $G_1^* \Lambda_k^0 G_1 \leqslant G_1^* G_1$, since Λ_k^0 belongs to the unit ball. Therefore,

$$\begin{aligned} \liminf_{k\to\infty} J_\Psi(\Lambda_k) &= \liminf_{k\to\infty} \int \mathrm{Tr}\left[\Lambda_k - \log(I + G_1^* \Lambda_k G_1)\right] \\ &= \liminf_{k\to\infty} \mathrm{Tr}\left[\|\Lambda_k\| \Lambda_k^0\right] - \liminf_{k\to\infty} \int \mathrm{Tr}\left[\log\left[\|\Lambda_k\|\left(\frac{1}{\|\Lambda_k\|} I + G_1^* \Lambda_k^0 G_1\right)\right]\right] \\ &\geqslant \liminf_{k\to\infty} \|\Lambda_k\| \frac{\eta}{2} - \liminf_{k\to\infty} \int \log\|\Lambda_k\| - \liminf_{k\to\infty} \int \mathrm{Tr}\left[\log\left(\frac{1}{\|\Lambda_k\|} I + G_1^* G_1\right)\right] \\ &= \liminf_{k\to\infty} \frac{\eta}{2}\left(\|\Lambda_k\| - \frac{4\pi}{\eta}\log\|\Lambda_k\|\right) - \liminf_{k\to\infty} \int \mathrm{Tr}\left[\log\left(\frac{1}{\|\Lambda_k\|} I + G_1^* G_1\right)\right] \\ &= +\infty. \end{aligned}$$

■

Then, by Weierstrass' Theorem we can conclude that there exists a minimum point $\Lambda^\circ \in \overline{\mathcal{L}_+^\Gamma}$. More can be proven:

Theorem 3.7.2. *If the feasibility condition (3.60) holds, then the problem of minimizing* $J_\Psi(\Lambda)$ *over* $\mathcal{L}_+^\Gamma$ *admits a unique solution* $\Lambda^\circ \in \mathcal{L}_+^\Gamma$.

Proof. Since $J_\Psi(\Lambda)$ is inf-compact over $\overline{\mathcal{L}_+^\Gamma}$, it admits a minimum point Λ° there. Obviously, $\Lambda^\circ \notin \mathcal{B}$, since $J_\Psi(\Lambda) = +\infty$ on $\mathcal{B}$ (lemma 3.7.3). Suppose $\Lambda^\circ \in \mathcal{B}^c$. In this case, $\det\left(G_1^*\Lambda G_1 + I\right)$ is a non-zero rational function, whose inverse is then a well defined rational function having (a finite number of) poles on the unit circle $\mathbb{T}$. Hence, $\left(I + G_1^*\Lambda^\circ G_1\right)^{-1}$ is a well defined matrix-valued rational function having (a finite number of) poles on the unit circle $\mathbb{T}$ and taking positive definite values at all $z = e^{j\vartheta}$ that are not poles. Hence $\operatorname{Tr}\left[\left(I + G_1^*\Lambda^\circ G_1\right)^{-1}\right]$ is a positive rational function having (a finite number of) poles on $\mathbb{T}$. This gives $\int \operatorname{Tr}\left[\left(I + G_1^*\Lambda^\circ G_1\right)^{-1}\right] = \infty$. Moreover, by lemma 3.7.3 again, it follows that $J_\Psi(\Lambda^\circ)$ is finite. By convexity of $\overline{\mathcal{L}_+^\Gamma}$, $\forall\, \varepsilon \in [0,1]$, $\Lambda^\circ + \varepsilon(I - \Lambda^\circ) \in \overline{\mathcal{L}_+^\Gamma}$, since the feasibility condition (3.60) ensures that $I \in \mathcal{L}_+^\Gamma$. The one-sided directional derivative is

$$\begin{aligned}
\delta J_{\Psi_+}(\Lambda^\circ; I - \Lambda^\circ) &= \lim_{\varepsilon \searrow 0}\left[\frac{J_\Psi\left(\Lambda^\circ + \varepsilon\left(I - \Lambda^\circ\right)\right) - J_\Psi(\Lambda^\circ)}{\varepsilon}\right] \\
&= \operatorname{Tr}\left[I - \Lambda^\circ\right] - \int \operatorname{Tr}\left[\left(I + G_1^*\Lambda^\circ G_1\right)^{-1} G_1^*\left(I - \Lambda^\circ\right) G_1\right] \\
&= \operatorname{Tr}\left[I - \Lambda^\circ\right] - \int \operatorname{Tr}\left[\left(I + G_1^*\Lambda^\circ G_1\right)^{-1}\left(I + G_1^* G_1\right) - I\right] \\
&\leqslant \operatorname{Tr}\left[2I - \Lambda^\circ\right] - \int \operatorname{Tr}\left[\left(I + G_1^*\Lambda^\circ G_1\right)^{-1}\right] = -\infty.
\end{aligned} \tag{3.100}$$

As a consequence, the minimum point cannot belong to $\partial\mathcal{L}_+^\Gamma$. Thus, $\Lambda^\circ \in \mathcal{L}_+^\Gamma$. ■

Tu sum up, this section shows that

- A $\Lambda^\circ \in \mathcal{L}_+^\Gamma$ that minimizes $J_\Psi(\Lambda)$ in (5.44) *does* exist;
- Λ° is unique;
- Λ° annihilates the directional derivative (3.73) in any direction $\delta\Lambda \in \mathcal{Q}(n)$, namely

$$\left\langle I - \int G_1 (I + G_1^*\Lambda^\circ G_1)^{-1} G_1^*, \delta\Lambda \right\rangle = 0 \quad \forall \delta\Lambda \in \mathcal{Q}(n), \tag{3.101}$$

or, equivalently,

$$I = \int G_1 (I + G_1^* \Lambda^\circ G_1)^{-1} G_1^* = \int G \Phi^\circ(\Lambda^\circ) G^*. \tag{3.102}$$

This means that the corresponding spectral density

$$\Phi^\circ(\Lambda^\circ) = \left[\Psi^{-1} + G^* \Lambda^\circ G\right]^{-1} \tag{3.103}$$

satisfies constraint (3.14)

Therefore, the solution of the dual problem Λ° is in one-to-one correspondence with the solution of RER spectrum estimation problem 3.6.1.

3.8 EFFICIENT IMPLEMENTATION OF A MATRICIAL NEWTON-LIKE ALGORITHM

In the same vein of Ferrante, Pavon, and Zorzi [41], Ramponi, Ferrante, and Pavon [89], we propose a matricial Newton-like algorithm in order to compute the minimizer of the dual functional $J_\Psi(\Lambda)$. First we set the starting point for the minimizing sequence $\{\Lambda_i\}_{i\in\mathbb{N}}$ to $\Lambda_0 = 0$. Then, at each step of we perform the following tasks:

1. Compute the Newton *search direction* $\Delta\Lambda_i$
2. Once the search direction is found, compute the Newton *step length* t_i^k.

Remark 3.8.1. By setting $\Lambda_0 = 0$, the algorithm starts from the candidate solution $\Phi_0 = \Psi$. Thus, it keeps adjusting the *prior* spectral density at each step until it satisfies the integral constraints (3.14).

3.8.1 Search Direction

The matricial nature of the problem makes the the computation of the search direction rather delicate. Indeed, no matricial expression of the Hessian and the gradient, which would allow us to compute the search direction Δx as $\Delta x = -{H_x}^{-1} \nabla f_x$, is available. As a consequence, in order to compute $\Delta\Lambda_i$, given $\Lambda_i \in \mathcal{L}_+^\Gamma$, one has to solve the following equation for the unknown $\Delta\Lambda_i$:

$$H_{\Lambda_i}(\Delta\Lambda_i, \cdot) = -\nabla J_{\Psi,\Lambda_i}(\cdot). \tag{3.104}$$

This can be explicitly written as:

$$\int G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^*\Delta\Lambda_i G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^* = \int G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^* - I. \tag{3.105}$$

To this aim, compute a basis of Range(Γ). It can be readily obtained, by recalling that $\Sigma_k \in$ Range(Γ) if and only if $\exists H_k \in \mathbb{R}^{m\times n}$ s.t. $\Sigma_k - A\Sigma_k A^\top = BH_k + H_k{}^\top B^\top$ (see e.g. Georgiou [54]). Therefore, considering a basis $\{H_1,\dots,H_L\}$ for $\mathbb{R}^{m\times n}$, a set of generators $\{\Sigma_1',\dots,\Sigma_L'\}$ can be found by solving L Lyapunov equations. After that a basis $\{\Sigma_1',\dots,\Sigma_N'\}$ can be easily computed. Since $I \in$ Range Γ, we can add to each Σ_i the matrix $\alpha_i I$, and, for suitable (large) α_i, get a basis $\{\Sigma_1,\dots,\Sigma_N\}$ of Range(Γ) made of positive definite matrices. The search direction can now be computed by applying the following procedure:

1. Compute
$$Y = \int G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^* - I \tag{3.106}$$
2. For each generator Σ_k, compute
$$Y_k = \int G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^*\Sigma_k G_1(I+G_1^*\Lambda_i G_1)^{-1}G_1^* \tag{3.107}$$
3. Find $\{\alpha_k\}$ s.t. $Y = \sum_k \alpha_k Y_k$;
4. Set $\Delta\Lambda_i = \sum_k \alpha_k \Sigma_k$.

The most challenging step is to compute Y and Y_k. A sensible approach is to employ spectral factorization techniques in order to compute the integrals, along the same lines described in Ramponi, Ferrante, and Pavon [89, Section VI]. Indeed, the integrand that appears in equation (3.106) is a coercive spectral density and the same holds for the integrand in (3.107), since we have chosen the generators Σ_i to be positive definite. For the computation of Y, let us focus on $Q_{\Lambda_i}(z) = I + G_1^*(z)\Lambda_i G_1(z)$. Assume that a realization of the stable minimum phase spectral factor $W_\Psi(z)$ is given (or has been computed from Ψ). Then, we can easily obtain a state-space realization $G_1(z) = C_1(zI-A_1)^{-1}B_1$ of G_1. Since $\Lambda_i \in \mathcal{L}_+^\Gamma$, $Q_{\Lambda_i}(z)$ is positive definite on $\mathbb{T}$, so that the following ARE admits a positive definite stabilizing solution $P = P^\top > 0$ (see, e.g. Lemma 6.4 in Ramponi, Ferrante, and Pavon [89]):

$$P = A_1^\top P A_1 - A_1^\top P B_1 (B_1^\top P B_1 + I)^{-1} B_1^\top P A_1 + C_1^\top \Lambda_i C_1. \tag{3.108}$$

Moreover, $Q_{\Lambda_i}(z)$ can be factorized as $Q_{\Lambda_i}(z) = \Delta_{\Lambda_i}^*(z)\Delta_{\Lambda_i}(z)$, where $\Delta_{\Lambda_i}(z)$ can be explicitly written in term of the stabilizing solution P:

$$\Delta_{\Lambda_i}(z) = (B_1^\top P B_1 + I)^{-\frac{1}{2}} B_1^\top P A_1 (zI - A_1)^{-1} B_1 + (B_1^\top P B_1 + I)^{\frac{1}{2}}. \tag{3.109}$$

It is now easy to compute a state space realization of $\Delta_{\Lambda_i}^{-1}$ and then of the stable filter

$$W_Y := G_1 \Delta_{\Lambda_i}^{-1} = C_1 (zI - Z_1)^{-1} B_1 (B_1^\top P B_1 + I)^{-\frac{1}{2}}, \tag{3.110}$$

with

$$Z_1 := A_1 - B_1 (B_1^\top P B_1 + I)^{-1} B_1^\top P A_1 \tag{3.111}$$

being the closed-loop matrix. The computation of (3.106) is now immediate. In fact,

$$\begin{aligned} Y + I &= \int G_1 (I + G_1^* \Lambda_i G_1)^{-1} G_1^* \\ &= \int G_1 \Delta_{\Lambda_i}^{-1} \Delta_{\Lambda_i}^{-*} G_1^* = \int W_Y W_Y^*. \end{aligned} \tag{3.112}$$

The latter integral is thus the steady-state covariance of the output of the stable filter W_Y driven by normalized white noise. It can be obtained by computing the unique solution of the Lyapunov equation $R - Z_1 R Z_1^\top = B_1 (B_1^\top P B_1 + I)^{-1} B_1^\top$ and setting $Y + I = C_1 R C_1^\top$, so that

$$Y = C_1 R C_1^\top - I. \tag{3.113}$$

A similar procedure may be employed to compute also the matrices Y_k's.

3.8.2 Step length

The backtracking line search is implemented by halving the step t_i until both the following conditions are satisfied:

$$\Lambda_i + t_i^k \Delta\Lambda_i \in \mathcal{L}_+^\Gamma; \tag{3.114}$$

$$J_\Psi(\Lambda_i + t_i^k \Delta\Lambda_i) < J_\Psi(\Lambda_i) + \alpha t_i^k \nabla J_{\Psi,\Lambda_i} \Delta\Lambda_i, \tag{3.115}$$

where $0 < \alpha < 0.5$.

The first condition can be easily evaluated by testing whether $Q_{\Lambda_i + t_i^k \Delta\Lambda_i}$ admits a factorization of the kind introduced in the previous subsection or, equivalently, whether the corresponding ARE (3.108) admits a solution $P = P^\top > 0$.

The only difficulty in checking the second condition is in computing

$$\begin{aligned} J_\Psi(\Lambda) &= \mathrm{Tr} \int [\Lambda - \log(I + G_1^* \Lambda G_1)] \\ &= \mathrm{Tr}\Lambda - \int \log\det(I + G_1^* \Lambda G_1). \end{aligned} \tag{3.116}$$

The evaluation of the latter integral can be attained straightforwardly in the light of Szegö-Kolmogorov formula (3.3). In our case $Q(z) = Q_\Lambda(z)$ may be factorized as $Q_\Lambda =$

$\Delta^*\Delta$, where Δ is a stable and minimum phase filter for which a minimal realization can be computed as in 3.109 on page 37. Since $\log\det Q_\Lambda = \log\det[\Delta^*\Delta] = \log\det[\Delta\Delta^*]$, $\det R$ is given by $\det[\Delta(\infty)\Delta^*(\infty)]$ which may be explicitly written in terms the solution P of the corresponding ARE as $\det[B_1^\top P B_1 + I]$. Therefore,

$$\int \log\det(I + G_1^* \Lambda G_1) = \log\det\left(B_1^\top P B_1 + I\right). \tag{3.117}$$

3.8.3 Convergence of the proposed algorithm

A sufficient condition for global convergence of the algorithm is that the following requirements are satisfied Boyd and Vandenberghe [6, Chapter 9]:

1. $J_\Psi(\cdot)$ is twice continuously differentiable;
2. $\Lambda_0 \in \mathcal{L}_+^\Gamma$ and the sublevel set $S := \left\{\Lambda \in \mathcal{L}_+^\Gamma | J_\Psi(\Lambda) \leqslant J_\Psi(\Lambda_0)\right\}$ is closed;
3. $J_\Psi(\cdot)$ is *strongly* convex, i.e. $\exists\, m$ s.t. $H(J_\Psi)(\Lambda) > mI$, $\forall\, \Lambda \in S$.
4. The Hessian is Lipschitz continuous in S, i.e. $\exists L$ such that:

$$\left\|H_{\Lambda_1} - H_{\Lambda_2}\right\|_2 < L\left\|\Lambda_2 - \Lambda_1\right\|_2 \quad \forall \Lambda_1, \Lambda_2 \in S. \tag{3.118}$$

In this case, it is possible to prove not only that the algorithm converges, but also that, after a certain number of iterations, the backtracking line search always selects the full step (i.e. $t = 1$). During the last stage the rate of convergence is quadratic, since there exists a constant C such that $\|\Lambda_{i+1} - \Lambda^\circ\| \leqslant C\|\Lambda_i - \Lambda^\circ\|^2$.

Let us examine the requirements one by one.

- The continuous differentiability of the dual function has already been proven in Section 3.7.
- Theorem 3.7.2 states that the sublevel sets of the dual function J_Ψ are compact, and hence closed (recall that, in a finite dimensional vector space, a set is compact if and only if it is closed and bounded).
- It is possible to conclude straightforwardly on *strong* convexity and Lipschitz continuity of the Hessian. Indeed, let us consider the sublevel set

$$S = \left\{\Lambda \in \mathcal{L}_+^\Gamma \,|\, J_\Psi(\Lambda) \leqslant J_\Psi(\Lambda_0)\right\}. \tag{3.119}$$

  Notice that, assuming that Λ_0 is the starting point, the minimizing sequence computed by the Newton algorithm with backtracking line search is such that,

$\forall k \geqslant 0$, $\Lambda_k \in S$. The continuity of the Hessian over $\mathcal{L}_+^\Gamma$ has already been proven in Section 3.7. Moreover, since the map from a Hermitian matrix to its minimum eigenvalue is continuous (see lemma 5.1 in Ramponi, Ferrante, and Pavon [89]), the map from $\Lambda \in \mathcal{L}_+^\Gamma$ to the minimum eigenvalue of $H_\Lambda(\delta\Lambda, \delta\Lambda)$ is continuous, being a composition of continuous maps. Since S is compact, Weierstrass' Theorem holds. Therefore, there exists a minimum m in the set of eigenvalues of the Hessian $H_\Lambda(\delta\Lambda, \delta\Lambda)$, $\forall \Lambda \in S$. Recall that the hypothesis of *strict* convexity holds (as proven in Theorem 3.7.1). As a consequence, the Hessian H_Λ is a positive definite matrix $\forall \Lambda \in S$, therefore $m > 0$. In conclusion, there exists $m > 0$ such that $H_\Lambda > mI$, $\forall \Lambda \in S$, i.e. $J_\Psi(\Lambda)$ is *strongly* convex.

- Concerning the Lipschitz continuity of the Hessian of $J_\Psi(\Lambda)$, it is easy to see that H_Λ is $\mathcal{C}^1(\mathcal{L}_+^\Gamma)$. Indeed the third variation $\delta^3 J_\Psi(\Lambda; \delta\Lambda_1, \delta\Lambda_2, \delta\Lambda_3)$ can be explicitly computed and its continuity can be proven along the same line developed in the proof of Theorem 3.7.1 (the result can be extended, leading to the conclusion that $J_\Psi(\Lambda)$ is $\mathcal{C}^\infty(\mathcal{L}_+^\Gamma)$. Continuous differentiability implies Lipschitz continuity on a compact set. Therefore, the Hessian is Lipschitz continuous on S.

In conclusion, global convergence of the Newton algorithm is guaranteed, so that the proposed procedure is an effective computational tool to solve the spectral estimation problem 3.6.1.

3.9 SIMULATION RESULTS

Next we test RER spectral estimation procedure. In the user's perspective, it can be outlined as follows:

1. Consider a finite sequence $\{y_1, \dots, y_N\}$. It is assumed to be a sample realization of the zero-mean Gaussian process $\mathbf{y}(t) = \{\mathbf{y}(k);\, k \in \mathbb{Z}\}$ with values in $\mathbb{R}^m$, whose spectrum is $\Phi(e^{j\vartheta})$.

2. Design a filter $G(z)$ with the same structure as (3.12). Recall from Section 3.4 that the filter imposes interpolation constraints on the input spectral density. It also affects the complexity of Φ°, as shown by (3.69).

3. Feed the filter with the data sequence $\{y_1, \dots, y_N\}$, collect the output data x_i and compute a consistent estimate $\hat{\Sigma}$ of the output covariance matrix. As stated in Section 3.4, we will recur to the technique described in Ferrante, Pavon, and Zorzi [41].

4. Choose a prior spectral density Ψ, as suggested in Section 3.4, according to the available *a priori* information and keeping in mind that the order of Ψ is related to the complexity of Φ°, in light of (3.69).

5. Solve Problem 3.6.1 by running the algorithm described in Section 3.8 with $G(z)$,Ψ and $\tilde{\Sigma}$ as inputs.

3.9.1 Scalar Case

Next we test our estimation technique in a scalar case. Our aim is to highlight its high-resolution features. We will also compare it with original THREE method. One of the most interesting features of THREE-like estimation procedures is that an adequate choice of the filterbank poles can improve the estimate's resolution. Indeed, a higher resolution can be attained in a prescribed frequency range by selecting poles which are close to the unit circle and have arguments in the range of interest. This was already noticed in Byrnes, Georgiou, and Lindquist [12]. Recently, further results on poles placement policy were established in Karlsson and Georgiou [62]. Our aim is detecting spectral lines in colored noise. This is a classical problem in spectrum estimation. In particular, we are going to analyze the same setting as the one described in Byrnes, Georgiou, and Lindquist [12, Section IV.B].

We consider a process $\mathbf{y}(t)$ that obeys to the following difference equation:

$$\begin{aligned} y(t) &= 0.5\sin(\omega_1 t + \phi_1) + 0.5\sin(\omega_2 t + \phi_2) + z(t), \\ z(t) &= 0.8z(t-1) + 0.5\nu(t) + 0.25\nu(t-1), \end{aligned}$$

where the variables ϕ_1, ϕ_2 and $\nu(t)$ are Gaussian, independent, with zero-mean and unit variance. Matrix B is a column of ones. Matrix A was chosen as a block-diagonal matrix; its real eigenvalues are 0, 0.85 and -0.85 and there are also five pairs of complex eigenvalues, whose arguments are equispaced in a narrow range of frequency where the sinusoids lie. Firstly, the spectral lines were fixed in $\omega_1 = 0.42$ rad/s and $\omega_2 = 0.50$ rad/s, and so the complex poles of $G(z)$ were chosen as:

$$0.95e^{\pm j0.42},\ 0.95e^{\pm j0.44},\ 0.95e^{\pm j0.46},\ 0.95e^{\pm j0.48},\ 0.95e^{\pm j0.50}.$$

By considering the constant prior, equal to the sample covariance of the available data, the proposed method was able to approximately detect both lines, as shown in Figure 3.

Then we considered the more challenging task when $\omega_1 = 0.45$ rad/s and $\omega_2 = 0.47$ rad/s. This choice makes the value of the distance between the two lines lower than

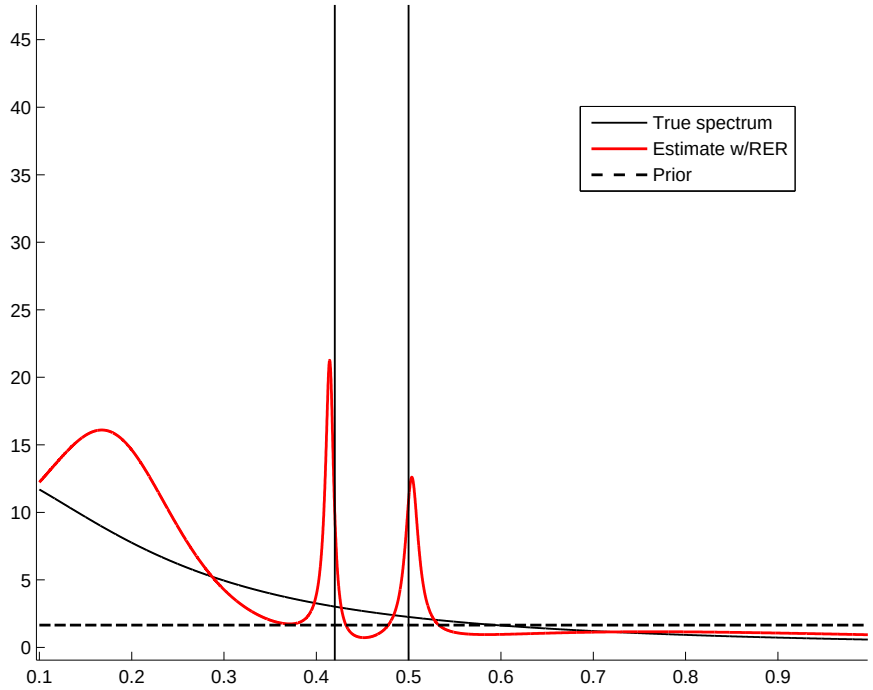


Figure 3: Estimation of spectral lines in colored noise ($\omega_1 = 0.42$ rad/s and $\omega_2 = 0.50$ rad/s). The chosen prior is the sample covariance of the data $\{y_k\}$. The radius of the complex poles is equal to 0.95.

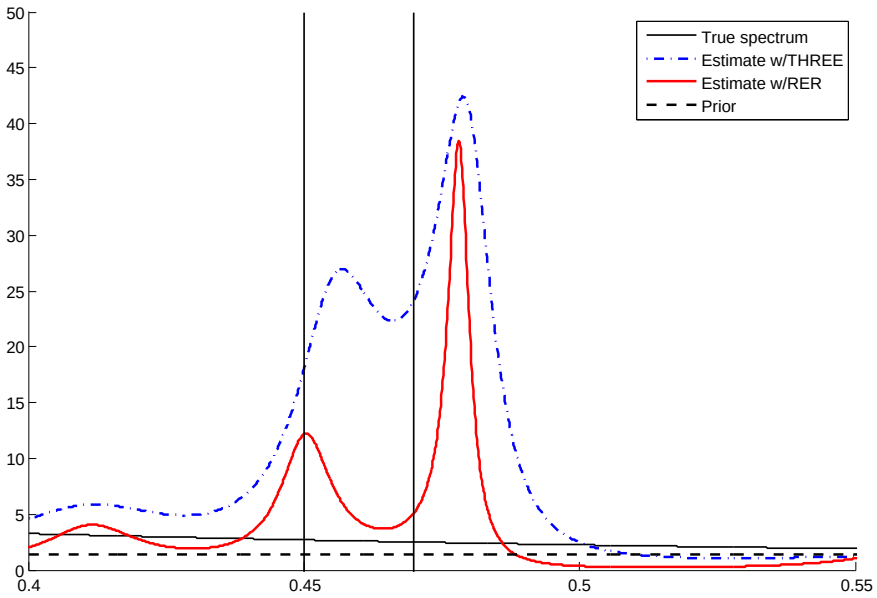


Figure 4: Estimation of close spectral lines in colored noise ($\omega_1 = 0.45$ rad/s and $\omega_2 = 0.47$ rad/s). The chosen prior is the sample covariance of the data $\{y_k\}$. The radius of the complex poles is equal to 0.95. Both RER and THREE indicate the presence of the two lines.

the resolution limit of the periodogram, which amounts to $\frac{2\pi}{N}$ (which in our case is $\frac{2\pi}{300} \simeq 0.021$ rad/s). Nevertheless, the RER estimator was still able to detect the presence of two lines. Figure 4 compares its performances with those achieved by the original THREE method for scalar spectral estimation. In simulations, RER turns out to perform at least as well as THREE in scalar spectral estimation. Another result is that, in general, poles which are closer to the unit circle imply both the resolution and the variance of the estimates get higher. The same trade-off was first described in Byrnes, Georgiou, and Lindquist [12] and seems to be typical of all THREE-like methods.

3.9.2 Multivariate Case

In order to test the performances of the proposed method in multivariate spectral estimation, we considered the same example described in Ramponi, Ferrante, and Pavon [89, Section VIII.C]. The process $\mathbf{y}(t)$ was obtained by filtering a bivariate Gaussian white noise process with zero mean and variance equal to the identity through a

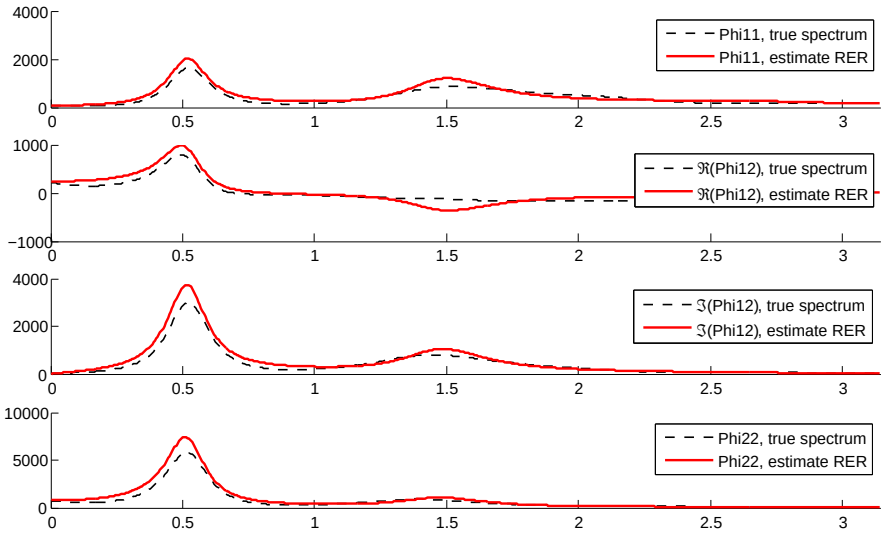


Figure 5: Multivariate spectrum estimation, $N = 300$, PEM(3) prior. Comparison between the approximant and the true spectrum.

square shaping filter of order 40. The filter coefficients were chosen at random, except for one fixed complex poles pair, $0.9e^{\pm j0.52}$ and the zeros pair $(1 - 10^{-5})e^{\pm j0.2}$.

The filter $G(z)$ was designed by choosing four complex poles pairs with radius 0.7 and arguments equispaced in the range $[0, \pi]$. We assumed $N = 300$ samples of the process $\mathbf{y}(t)$ to be available. As for the the prior, it was set equal to a simple PEM model of order 3, obtained by means of the standard function `pem` provided in MATLAB. Figure 5 shows the real spectrum and the estimate computed by the RER approach.

We then compared the performance of the proposed technique to those achieved by two other THREE-like approaches to multivariate spectral estimation:

- Maximum Entropy – ME – estimator, described in Georgiou [50];
- Hellinger distance-based estimator, see Ferrante, Pavon, and Ramponi [40].

In order to make the comparison as independent as possible of the specific data set, we performed 50 trials by feeding the shaping filter with independent realizations of the input noise process. Then we measured the average estimation error at each frequency, defined as

$$E_{\#}(\vartheta) := \frac{1}{50} \sum_{i=1}^{50} \|\hat{\Phi}_{\#}(e^{j\vartheta}) - \Phi(e^{j\vartheta})\|, \tag{3.120}$$

where the *spectral norm* (i.e. the largest singular value) is considered. Figure 6 shows that RER estimator performed better than ME estimator. It seems also to slightly

outperform the Hellinger-distance approach. It worthwhile that in Hellinger estimator computed a solution of order 19, while the solution provided by RER estimator had just McMillan degree 11. As for ME estimator, the corresponding solution had degree 8.

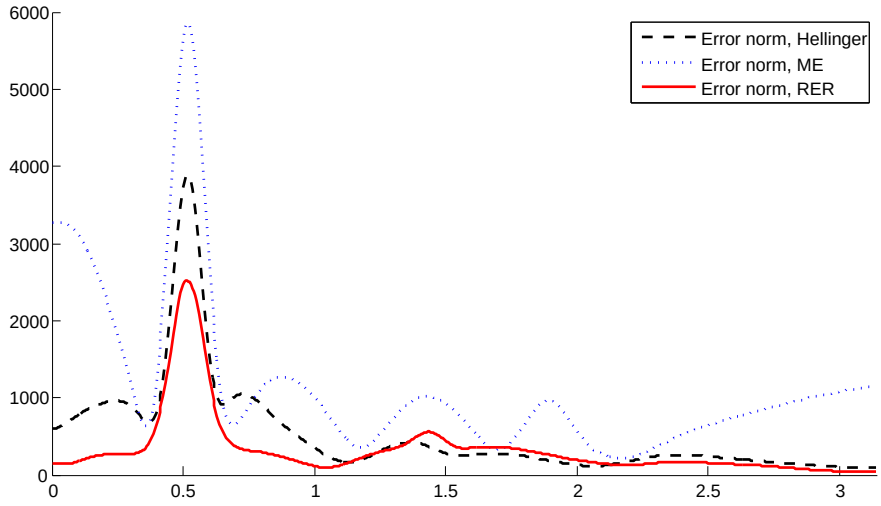


Figure 6: Comparison of THREE-like approaches in terms of average estimation error. $N = 300$ available data. Both RER and Hellinger estimator are provided with a PEM(3) prior. All the considered methods make use of the same filter $G(z)$

The case when only a few samples of the process $\mathbf{y}(t)$ are available, which is very relevant in practice, is a challenging scenario. Indeed, shortness of the available data record can heavily affect the estimates obtained by classical methods such as MATLAB's PEM and N4SID by introducing *artifacts*. On the contrary, RER method, as well as the other THREE-like approaches, seems to be quite robust. Figure 7 shows the results obtained in a case where only $N = 100$ samples are available. Both PEM and N4SID estimates are affected by artifacts. On the contrary, the proposed approach is not. This result seems to suggest that RER estimation is suitable to tackle spectral estimation issues where only short data records of the process of interest are available.

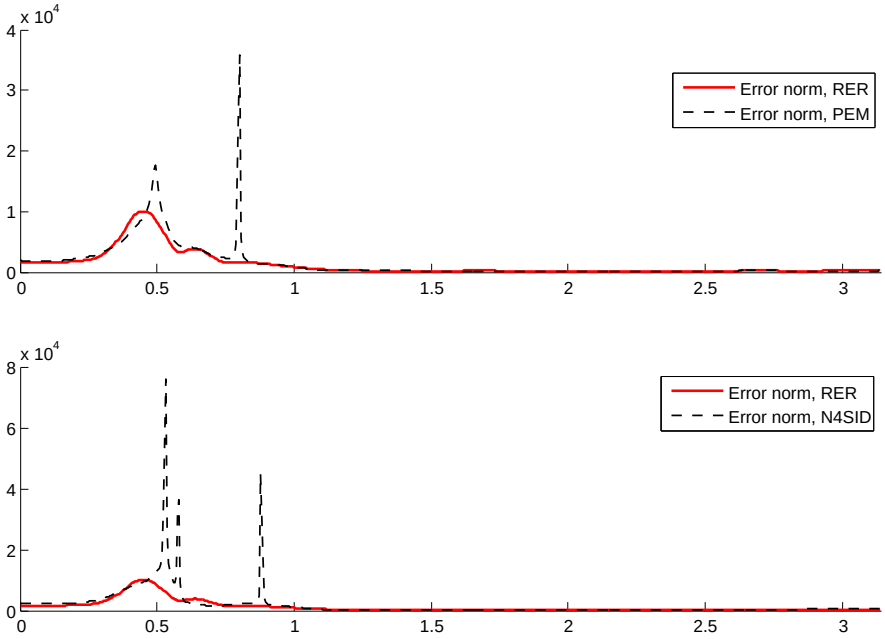


Figure 7: Comparison of MATLAB's PEM, MATLAB's N4SID and RER in terms of average estimation error. $N = 100$ available samples. The prior considered for RER is a PEM(2) model. The filter $G(z)$ has a pole in the origin and four complex conjugate poles pairs with radius 0.7. Notice that RER does not exhibit artifacts, whereas PEM and N4SID do.

3.10 CONCLUSIONS AND FUTURE WORK

Relative entropy rate (RER) estimator is a new approach to multivariate spectral estimation. It draws inspiration from a paradigm known as THREE (Tunable High Resolution Estimator, see Section 3.4), where *a priori* information on the process is taken into account and, by means of a bank of filters which can be designed arbitrarily, it is possible to impose interpolation constraints on the estimated spectral density. Therefore, spectrum estimation is recast as a generalized moment problem that can be solved efficiently by means of convex optimization techniques.

RER relies on a remarkable information-theoretic result that relates time and spectral domain relative entropy rates for stationary Gaussian processes. Thus, it provides a very natural extension of maximum entropy methods for multivariate spectral estimation when a prior estimate of the spectrum is available.

Moreover, it features an upper bound on the complexity of the estimate which is equal to the one provided by the original THREE method in the scalar context, and thus it improves sensibly on the best one so far available in the multichannel setting with prior estimate.

Finally, it inherits all the desirable properties of THREE-like methods. Indeed, it exhibits high resolution features, which can be tuned by placing the filterbank poles suitably. In addition, RER estimator is robust with regard to short observation records, outperforming PEM and N4SID (in their standard MATLABimplementation) in this framework, which is definitely significant in practice.

As for future research directions, one of them could be applying the theory underlying RER estimator to graphical models identification. Graphical models describe Gaussian multivariate stochastic processes with the property that some pairs of components are *conditionally* independent given the others. Such processes find interest in a widespread variety of fields, e.g. image processing, econometrics, bioinformatics, chemistry, medicine and so on (see e.g. Avventi, Lindquist, and Wahlberg [1], Brillinger [7], Dahlhaus [28], Materassi and Innocenti [75], Songsiri and Vandenberghe [95] and references therein). In this framework it is natural to represent the process by means of a graph, in which each component is a node and the lack of an edge between two nodes implies that they are conditionally independent, given the others. It is worthwhile to notice that conditionally independent joint Gaussian processes are also conditionally orthogonal. Consider the process $\mathbf{x}(t) = \{\mathbf{x}(k)\,;\, k \in \mathbb{Z}\}$ with values in $\mathbb{R}^m$. Denote by $\mathbf{x}_1, \ldots, \mathbf{x}_m$ its scalar components. Let $\mathcal{S}_+^m$ be the set of the spectral

densities which are positive definite and integrable on the interval $(-\pi, \pi]$ and assume the spectrum of $\mathbf{x}(t)$ is $\Phi(e^{j\vartheta}) \in \mathcal{S}_+^m$. It can be shown (see e.g. Dahlhaus [28]) that

$$\left[\Phi(e^{j\vartheta})^{-1}\right]_{kl} = 0, \quad \vartheta \in (-\pi, \pi] \tag{3.121}$$

for pairs (k, l) such that $\mathbf{x}_k$ and $\mathbf{x}_l$ are conditionally independent, given the rest of the components of $\mathbf{x}$.

Let us now introduce the *interaction graph* $G = (V, E)$, such that

$$V = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}; \quad E \subseteq V \times V \tag{3.122}$$

and

$$(k, l) \notin E \Leftrightarrow k \neq l, X_{\{k\}} \perp X_{\{l\}} \,|\, X_{V\setminus\{k,l\}} \tag{3.123}$$

where, given an arbitrary set $I \subset V$, $X_I := \operatorname{span}\{\mathbf{x}_j(t) : j \in I, t \in \mathbb{Z}\}$. An example is shown in Fig. 8.

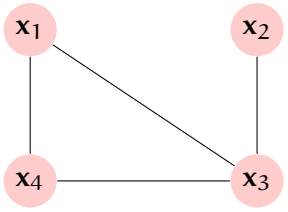


Figure 8: A graphical model for a Gaussian process of size 4. Note that x_1 is conditionally independent of $\mathbf{x}_2$, given $\mathbf{x}_3$ and $\mathbf{x}_4$

In order to identify a graphical model for a Gaussian process, we could draw inspiration from the approach described in Avventi, Lindquist, and Wahlberg [1], which computes a rational spectral density of the form

$$\Phi(e^{j\vartheta}) = \psi(e^{j\vartheta})Q(e^{j\vartheta})^{-1} \tag{3.124}$$

where $\psi(e^{j\vartheta})$ is a scalar pseudo-polynomial. Thus, the topology of the interaction graph is defined completely by the matricial pseudo-polynomial $Q(e^{j\vartheta})$. Indeed,

$$(k, l) \notin E \;\Leftrightarrow\; Q_{kl}(e^{j\vartheta}) \equiv 0 \text{ on } \mathbb{T}. \tag{3.125}$$

Then, three steps lead to identification of a graphical model for the Gaussian process of interest:

1. Compiling a list of candidate topological patterns;
2. Solving a spectral estimation issue for each of them, based on the available covariance lags sequence;

3. Selecting the solution that provides the best trade off between moments matching and model complexity.

Suppose that M covariance lags are available:

$$R(0), R(1), \dots, R(M-1).$$

Next, let us focus just on the second step. If, once the topology is fixed, we impose the interpolation conditions only to the entries of the spectral density corresponding to existing edges in the partial correlation graph, i.e.

$$\int \Phi_{(l,m)}(e^{j\vartheta}) e^{jk\vartheta} \frac{d\vartheta}{2\pi} = R(k)_{(l,m)} \quad \forall\, (l,m) \in E, \quad \forall\, k = 0, 1, \dots, M-1, \tag{3.126}$$

the solution turns out to automatically satisfy the conditional orthogonality structure imposed by the interaction graph (see Avventi, Lindquist, and Wahlberg [1, Section 4]). This fact can also be interpreted in light of the geometric results on constrained optimization of information theoretic indexes proved in Pavon and Ferrante [84]. We could exploit the same result in RER estimation framework. In principle, by using a RER-like approach, it should be possible to identify more general ARMA models, i.e. models where the moving average part can be an arbitrary matricial pseudo-polynomial instead of $\psi(e^{j\vartheta})I$. Suppose that M covariance lags are available:

$$R(0), R(1), \dots, R(M-1).$$

In this context, we aim at solving

Problem 3.10.1 (ARMA identification of graphical models based on RER estimation)**.** Let Ψ be an available prior spectral density. Minimize

$$d_{RER}(\Phi \| \Psi) := \frac{1}{2} \int \log\det \Phi^{-1}\Psi + \operatorname{Tr}\left[\Psi^{-1}(\Phi - \Psi)\right] \frac{d\vartheta}{2\pi}, \tag{3.127}$$

under the constraints (3.126).

By discarding the negligible terms, the corresponding Lagrangian is

$$\begin{aligned} \mathcal{L}_\Psi(\Phi, \{\lambda^k_{(l,m)}\}) = \int \left\{ -\log\det\Phi + \operatorname{Tr}\left[\Psi^{-1}\Phi\right] + \operatorname{Tr}\left[\Phi(e^{j\vartheta})\Lambda(e^{j\vartheta})\right] \right\} \frac{d\vartheta}{2\pi} \\ - \sum_{k=0}^{M-1} \operatorname{Tr}\left[\Lambda^*(k) R(k)\right]. \end{aligned} \tag{3.128}$$

Denote the vector of size m whose unique non-zero entry is the one in position k by e_k. Then,

$$\Lambda(k) := \sum_{(l,m)\in E} \lambda^k_{(l,m)} e_l e_m^\top \tag{3.129}$$

and

$$\Lambda(e^{j\vartheta}) := \frac{1}{2}\left[\sum_{k=0}^{M-1} \Lambda(k)e^{-jk\vartheta} + \sum_{k=0}^{M-1} \Lambda(k)^* e^{jk\vartheta}\right]. \tag{3.130}$$

By variational analysis, it turns out that the optimum solution has the form

$$\hat{\Phi} = [\Psi^{-1} + \Lambda]^{-1}. \tag{3.131}$$

Therefore, the inverse density is given by

$$\hat{\Phi}^{-1} = \Psi^{-1} + \Lambda. \tag{3.132}$$

By definition, all the entries of Λ whose position corresponds to a missing edge in the partial correlation graph are equal to zero. Thus, we can conclude that if the prior Ψ satisfies the topology constraints imposed by E, then $\hat{\Phi}$ exhibits the same topological pattern, too. First, existence and uniqueness of the solution of Problem 3.10.1 have to be investigated. When the solution exists and is unique, it can be computed by solving the corresponding dual problem, which is finite-dimensional. If that is the case, the next step is developing an efficient optimization procedure. Finally, choosing the prior Ψ so that it is consistent with the topology is another key issue to tackle.

4 MULTIVARIATE CIRCULANT RATIONAL COVARIANCE EXTENSION

4.1 INTRODUCTION TO MULTIVARIATE CIRCULANT RATIONAL COVARIANCE EXTENSION

This chapter focuses on rational covariance extension for multivariate periodic processes. Basically, rational covariance extension aims at estimating a spectral density in such a way that the estimate is rational and consistent with the available covariance lags. Given the rational spectral density of a stationary process, the latter can be modeled easily as the output of a finite memory linear filter fed by white noise (see Section A.10). Thus, rational covariance extension can be also interpreted as a technique that paves the way to filtering, estimation and prediction, for instance. Therefore, it plays a key role in systems and control, see e.g. Byrnes, Enqvist, and Linquist [9, 10], Byrnes, Gusev, and Lindquist [13, 14], Byrnes and Lindquist [20], Byrnes et al. [21], Enqvist [36], Georgiou [46, 47], Kalman [61], Pavon and Ferrante [84] and references therein.

Covariance extension was already mentioned in Section 3.4. In particular, it was proved that it could be recast in the framework of THREE-like spectral estimation. In the following we will follow a different path. Indeed, we will deal with covariance extension for periodic processes. First, this problem is interesting *per se*. Second, since it leads to partial stochastic realizations in the form of bilateral ARMA models, it also connects up to a rich realization theory for reciprocal processes, see e.g. Carli et al. [24], Carli et al. [25], Krener [66], Krener, Frezza, and Levy [67], Levy and Ferrante [70], Levy, Frezza, and Krener [71]. Finally, covariance extension for periodic processes also provides an efficient tool for approximating regular covariance extension, as it is based on fast Fourier transforms (FFT). This was first noticed in Lindquist and Picci [74]. In the following, we shall provide numerical evidence that this also holds in the multivariable case.

Rational covariance extension for periodic processes is strictly related to (block) circulant Toeplitz matrix completion problems. Thus, in the following it will be also referred to as *circulant* rational covariance extension. Recently it was discovered that this problem can be recast in the context of the optimization-based theory of moment problems with rational measures developed in Blomqvist, Lindquist, and Nagamune

[5], Byrnes, Enqvist, and Linquist [10], Byrnes, Georgiou, and Lindquist [11], Byrnes, Gusev, and Lindquist [13, 14], Byrnes and Lindquist [15, 16], Georgiou [49], Georgiou and Lindquist [56]. A complete theory for the scalar case was presented in Lindquist and Picci [74]. Here we provide a first step in generalizing this theory to the multivariable case.

The chapter is organized as follows: Section 4.2 recalls some preliminary results about regular multivariable rational covariance extension and harmonic analysis on the discrete unit circle. Then, in Section 4.3, we present our main result on the multivariable circulant rational covariance extension problem, that provides a complete parametrization the family of solutions. In Section 4.4 we show how logarithmic moments can be used to determine the best particular solution. In Section 4.6 we provide numerical examples which suggest that circulant covariance extension is a powerful tool for approximation. Finally, Section 4.7 concludes the chapter with some final remarks. It also proposes some future research directions.

4.2 PRELIMINARIES

4.2.1 Stationary periodic vector processes

We can consider a finite collection of $2N$ random vectors of dimension m as an m-dimensional stochastic process $\mathbf{y} = \{\mathbf{y}(k),\, k = -N+1, \dots, N\}$ defined on the finite interval $[-N+1, N] \subset \mathbb{Z}$. Let

$$\mathbf{y} := \begin{bmatrix} \mathbf{y}(-N+1) \\ \vdots \\ \mathbf{y}(0) \\ \vdots \\ \mathbf{y}(N) \end{bmatrix}. \tag{4.1}$$

Then the process $\mathbf{y}$ is said to be stationary if

$$\mathbb{E}\,[\mathbf{y}(j)\mathbf{y}(k)^*] = C_{j-k}, \quad \text{for all} \quad i, j = -N+1, \dots, N, \tag{4.2}$$

i.e. it only depends on the difference between the time indexes. In this case, the covariance matrix

$$\mathbf{\Sigma} := \mathbb{E}\,[\mathbf{y}\mathbf{y}^*] \tag{4.3}$$

is a block Toeplitz matrix.

Consider now a stationary multivariate process $\tilde{\mathbf{y}}$ defined on $\mathbb{Z}$, with period $2N$, i.e. such that

$$\tilde{\mathbf{y}}(k+2N) = \tilde{\mathbf{y}}(k), \quad \text{almost surely for all } k \in \mathbb{Z}. \tag{4.4}$$

Notice that a stationary process of period $2N$ can be considered as a process indexed on $\mathbb{Z}_{2N} := \{-N+1, \dots, 0, \dots, N\}$ with arithmetic mod $2N$, as shown in Fig. 9. The

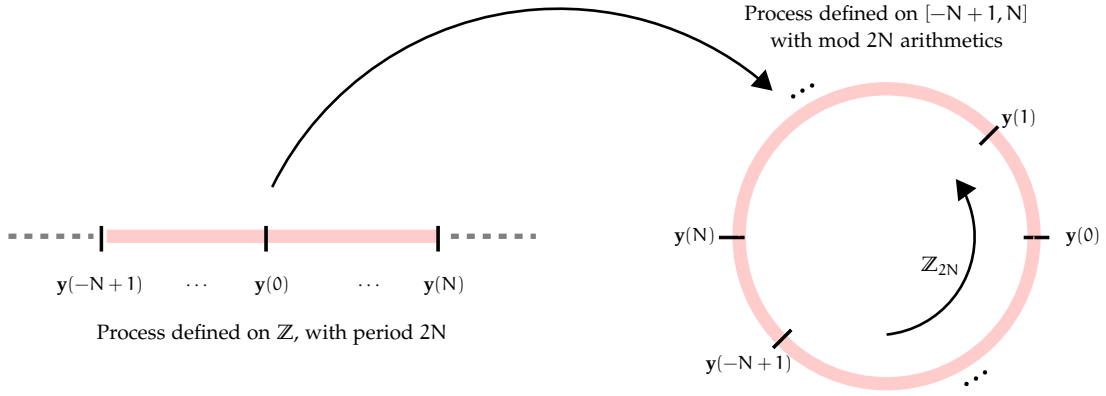


Figure 9: A periodic process of period $2N$ can be interpreted as a process indexed on $\mathbb{Z}_{2N}$

following important result holds (see Carli et al. [24]):

Proposition 4.2.1. *A stochastic process* $\mathbf{y}$ *defined on* $[-N+1, \dots, 0, \dots, N]$ *is the restriction to the interval of a wide-sense stationary periodic process of period* $2N$ *defined on* $\mathbb{Z}$*, if and only if its covariance matrix is Hermitian and block-circulant.*

Block-circulant matrices are block Toeplitz matrices whose block columns (or, equivalently, block rows) are shifted cyclically:

$$\mathrm{Circ}\{\Lambda_0, \Lambda_1, \dots, \Lambda_\nu\} := \begin{bmatrix} \Lambda_0 & \Lambda_\nu & \Lambda_{\nu-1} & \cdots & \Lambda_1 \\ \Lambda_1 & \Lambda_0 & \Lambda_\nu & \cdots & \Lambda_2 \\ \Lambda_2 & \Lambda_1 & \Lambda_0 & \cdots & \Lambda_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_\nu & \Lambda_{\nu-1} & \Lambda_{\nu-2} & \cdots & \Lambda_0 \end{bmatrix}. \tag{4.5}$$

Thus, we can write the covariance matrix $\mathbf{\Sigma}$ of a process which is the restriction on $[-N+1, \dots, 0, \dots, N]$ of a stationary process of period $2N$ defined on $\mathbb{Z}$ as

$$\mathbf{\Sigma} = \mathrm{Circ}\{C_0, C_1, C_2, \dots, C_N, C_{N-1}^*, \dots, C_2^*, C_1^*\}, \tag{4.6}$$

where

$$C_k := \mathbb{E}\left[\mathbf{y}(t+k)\mathbf{y}(t)^*\right]. \tag{4.7}$$

4.2.2 Harmonic analysis in $\mathbb{Z}_{2N}$

Harmonic analysis in $\mathbb{Z}_{2N}$ can be performed by means of the discrete Fourier transform (DFT). Let $\mathbf{g} = \{\mathbf{g}(k);\, k = -N+1, \dots, N\}$, with $\mathbf{g}(k) \in \mathbb{C}^m$ for $k = -N+1, \dots, N$. Let

$$D := \underbrace{\mathbb{C}^m \times \mathbb{C}^m \times \cdots \times \mathbb{C}^m}_{2N \text{ times}}. \tag{4.8}$$

Then, we can define DFT as a map

$$\begin{aligned} D &\to D \\ \mathbf{g} &\mapsto \mathcal{F}[\mathbf{g}] = \mathbf{G} \end{aligned}$$

where $\mathbf{G} = \{\mathbf{G}(\zeta_k);\, k = -N+1, \dots, N\}$ and

$$\mathbf{G}(\zeta_k) := \sum_{h=-N+1}^{N} \mathbf{g}(h)\zeta_k^{-h}, \qquad k = -N+1, -N+2, \dots, N. \tag{4.9}$$

and $\zeta_h := e^{jh\frac{\pi}{N}}$. Here we defined the discrete variable ζ running counterclockwise on the discrete unit circle $\mathbb{T}_{2N}$. In particular, we have $\zeta_h = (\zeta_1)^h$ and $\zeta_{-h} = \overline{\zeta_h}$. The inverse DFT $\mathcal{F}^{-1}$ is given by

$$\mathbf{g}(k) = \frac{1}{2N} \sum_{h=-N+1}^{N} \zeta_k^h \mathbf{G}(\zeta_h), \quad k = -N+1, -N+2, \dots, N, \tag{4.10}$$

which can also be written as a Stieltjes integral

$$\mathbf{g}(k) = \int_{-\pi}^{\pi} e^{ik\theta} \mathbf{G}(e^{i\theta}) d\nu(\theta), \quad k = -N+1, -N+2, \dots, N, \tag{4.11}$$

where ν is a step function with steps $\frac{1}{2N}$ at each ζ_h; i.e.,

$$d\nu(\theta) = \sum_{h=-N+1}^{N} \delta(e^{i\theta} - \zeta_h) \frac{d\theta}{2N}. \tag{4.12}$$

With $\mathbf{H}$ being the DFT of $\{\mathbf{h}\}$,

$$\begin{aligned} \sum_{j=-N+1}^{N} \mathbf{g}(j)\mathbf{h}(j)^* &= \frac{1}{2N} \sum_{k=-N+1}^{N} \mathbf{G}(\zeta_k)\mathbf{H}(\zeta_{-k})^* \\ &= \int_{-\pi}^{\pi} \mathbf{G}(e^{i\theta})\mathbf{H}(e^{i\theta})^* d\nu, \end{aligned} \tag{4.13}$$

which is *Plancherel's Theorem* for DFT. From this we see that

$$\langle G, H\rangle := \int_{-\pi}^{\pi} \operatorname{Tr}\left[\mathbf{G}(e^{i\theta})\mathbf{H}(e^{i\theta})^*\right] d\nu = \sum_{j=-N+1}^{N} \operatorname{Tr}\left[\mathbf{g}(j)\mathbf{h}(j)^*\right] \tag{4.14}$$

is computed exactly as in (4.46) despite the change of measure in the integral. Hence results such as (4.48) hold also with the Stieltjes measure $d\nu$.

We also provide a matricial expression for the discrete Fourier transform (4.9). With a slight abuse of notation we denote by $\mathbf{g}$ the vector

$$\left[\mathbf{g}(-N+1)^\top \mathbf{g}(-N+2)^\top \cdots \mathbf{g}(N)^\top\right]^\top \in \mathbb{C}^{2Nm} \tag{4.15}$$

and by $\mathbf{G}$ the vector

$$\left[\mathbf{G}(\zeta_{-N+1})^\top \mathbf{G}(\zeta_{-N+2})^\top \cdots \mathbf{G}(\zeta_N)^\top\right]^\top \in \mathbb{C}^{2Nm}. \tag{4.16}$$

Then we can write

$$\mathbf{G} = \mathbf{F}\mathbf{g}, \tag{4.17}$$

where $\mathbf{F}$ is the nonsingular $2mN \times 2mN$ block Vandermonde matrix

$$\mathbf{F} = \begin{bmatrix} \zeta_{-N+1}^{N-1} I_m & \zeta_{-N+1}^{N-2} I_m & \cdots & \zeta_{-N+1}^{-N} I_m \\ \vdots & \vdots & \cdots & \vdots \\ \zeta_0^{N-1} I_m & \zeta_0^{N-2} I_m & \cdots & \zeta_0^{-N} I_m \\ \vdots & \vdots & \cdots & \vdots \\ \zeta_N^{N-1} I_m & \zeta_N^{N-2} I_m & \cdots & \zeta_N^{-N} I_m \end{bmatrix}. \tag{4.18}$$

Likewise, it follows from (4.10) that

$$\mathbf{g} = \frac{1}{2N}\mathbf{F}^*\mathbf{G}, \tag{4.19}$$

i.e., $\mathcal{F}^{-1}$ corollary responds to $\frac{1}{2N}\mathbf{F}^*$. Consequently, $\mathbf{F}\mathbf{F}^* = 2N\,\mathbf{I}$, and hence $\mathbf{F}^{-1} = \frac{1}{2N}\mathbf{F}^*$ and $(\mathbf{F}^*)^{-1} = \frac{1}{2N}\mathbf{F}$.

Next consider a zero-mean stationary m-dimensional process $\mathbf{y}$ defined on $\mathbb{Z}_{2N}$; i.e., a stationary process defined on a finite interval $[-N+1, N]$ of the integer line $\mathbb{Z}$ and extended to all of $\mathbb{Z}$ as a periodic stationary process with period $2N$. Let $C_{-N+1}, C_{-N+2}, \ldots, C_N$ be the $m \times m$ covariance lags, defined as in (4.7), and define its discrete Fourier transformation

$$\Phi(\zeta_k) := \sum_{h=-N+1}^{N} C_h \zeta_k^{-h}, \qquad k = -N+1, \ldots, N, \tag{4.20}$$

which is a positive, Hermitian matrix-valued function of ζ. Then, as seen from (4.10) and (4.11),

$$\begin{aligned} C_k &= \frac{1}{2N} \sum_{h=-N+1}^{N} \zeta_k^h \Phi(\zeta_h) \\ &= \int_{-\pi}^{\pi} e^{ik\theta} \Phi(e^{i\theta}) d\nu, \quad k = -N+1, \dots, N. \end{aligned} \tag{4.21}$$

The $m \times m$ matrix function Φ is the *spectral density* of the vector process $\mathbf{y}$. In fact, let

$$\hat{\mathbf{y}}(\zeta_k) := \sum_{h=-N+1}^{N} \mathbf{y}(h) \zeta_k^{-h}, \quad k = -N+1, \dots, N, \tag{4.22}$$

be the discrete Fourier transformation of the process $\mathbf{y}$. Since

$$\frac{1}{2N} \sum_{h=-N+1}^{N} (\zeta_k \zeta_\ell^*)^h = \delta_{k\ell}, \tag{4.23}$$

the random variables (4.22) are uncorrelated, and

$$\frac{1}{2N} \mathbb{E}\{\hat{\mathbf{y}}(\zeta_k) \hat{\mathbf{y}}(\zeta_\ell)^*\} = \Phi(\zeta_k) \delta_{k\ell}. \tag{4.24}$$

This yields a spectral representation of $\mathbf{y}$ analogous to the usual one, namely

$$\mathbf{y}(k) = \frac{1}{2N} \sum_{h=-N+1}^{N} \zeta_k^h \, \hat{\mathbf{y}}(\zeta_k) = \int_{-\pi}^{\pi} e^{ik\theta} d\hat{\mathbf{y}}(\theta), \tag{4.25}$$

where $d\hat{\mathbf{y}} := \hat{\mathbf{y}}(e^{i\theta}) d\nu$.

4.2.3 Block-circulant matrices

In the multivariate circulant rational covariance extension problem we consider *Hermitian* circulant matrices of the same kind as (4.6). In general, they are defined as

$$\mathbf{M} := \mathrm{Circ}\{M_0, M_1, M_2, \dots, M_N, M_{N-1}^*, \dots, M_1^*\}, \tag{4.26}$$

and can be represented in the form

$$\mathbf{M} = \sum_{k=-N+1}^{N} S^{-k} \otimes M_k, \quad M_{-k} = M_k^* \tag{4.27}$$

where $\otimes$ is the Kronecker product and S is the nonsingular $2N \times 2N$ cyclic shift matrix

$$S := \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.28}$$

The $m \times m$ pseudo-polynomial

$$M(\zeta) = \sum_{k=-N+1}^{N} M_k \zeta^{-k}, \quad M_{-k} = M_k^* \tag{4.29}$$

is called the *symbol* of $\mathbf{M}$. Let $\mathbf{S}$ be the $2mN \times 2mN$ cyclic shift matrix

$$\mathbf{S} = S \otimes I_m = \begin{bmatrix} 0 & I_m & 0 & \dots & 0 \\ 0 & 0 & I_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & I_m \\ I_m & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.30}$$

Clearly $\mathbf{S}^{2N} = \mathbf{S}^0 = \mathbf{I} := I_{2mN}$, and

$$\mathbf{S}^{k+2N} = \mathbf{S}^k, \qquad \mathbf{S}^{2N-k} = \mathbf{S}^{-k} = (\mathbf{S}^k)^\top. \tag{4.31}$$

Moreover,

$$\mathbf{S}\mathbf{M}\mathbf{S}^* = \mathbf{M}, \tag{4.32}$$

is both necessary and sufficient for $\mathbf{M}$ to be circulant. With $\mathbf{g}$ defined as in (4.15), we have

$$[\mathbf{S}\mathbf{g}]_k = \mathbf{g}(k+1), \quad k \in \mathbb{Z}_{2N}. \tag{4.33}$$

Then, in view of (4.9), $\zeta \mathcal{F}(\mathbf{g})(\zeta) = \mathcal{F}(\mathbf{S}\mathbf{g})(\zeta)$, from which we have

$$\mathcal{F}(\mathbf{M}\mathbf{g})(\zeta) = M(\zeta)\mathcal{F}(\mathbf{g})(\zeta), \tag{4.34}$$

where the $m \times m$ matrix fuction $M(\zeta)$ is the symbol (4.29) of the circulant matrix $\mathbf{M}$. An important property of circulant block matrices is that they can be block-diagonalized by the discrete Fourier transform. More precisely, it follows from (4.34) that

$$\mathbf{M} = \frac{1}{2N}\mathbf{F}^* \operatorname{diag}\big(M(\zeta_{-N+1}), \dots, M(\zeta_N)\big)\mathbf{F}, \tag{4.35}$$

where "diag" denotes block diagonal. Hence the inverse is

$$\mathbf{M}^{-1} = \frac{1}{2N}\mathbf{F}^* \operatorname{diag}\big(M(\zeta_{-N+1})^{-1}, \dots, M(\zeta_N)^{-1}\big)\mathbf{F}, \tag{4.36}$$

and, since

$$\begin{aligned} \mathbf{S} &= \frac{1}{2N}\mathbf{F}^* \operatorname{diag}\big(\zeta_{-N+1}, \dots, \zeta_N\big)\mathbf{F} \\ \mathbf{S}^* &= \frac{1}{2N}\mathbf{F}^* \operatorname{diag}\big(\zeta_{-N+1}^{-1}, \dots, \zeta_N^{-1}\big)\mathbf{F}, \end{aligned} \tag{4.37}$$

we have

$$\mathbf{S}\mathbf{M}^{-1}\mathbf{S}^* = \mathbf{M}^{-1}. \tag{4.38}$$

Hence $\mathbf{M}^{-1}$ is also a circulant block matrix with symbol $M(\zeta)^{-1}$. In general, in view of the circulant property (4.27) and (4.31), quotients of symbols are themselves pseudo-polynomials of degree at most N and hence symbols. More generally, if $\mathbf{A}$ and $\mathbf{B}$ are circulant block matrices of the same dimension with symbols $A(\zeta)$ and $B(\zeta)$ respectively, then $\mathbf{AB}$ and $\mathbf{A}+\mathbf{B}$ are circulant matrices with symbols $A(\zeta)B(\zeta)$ and $A(\zeta)+B(\zeta)$, respectively. In fact, the circulant matrices of a fixed dimension form an algebra, and the DFT is an *algebra homomorphism* of the set of circulant matrices onto the pseudo-polynomials of degree at most N in the variable $\zeta \in \mathbb{T}_{2N}$.

4.2.4 The multivariable rational covariance extension problem

Before stating multivariate *circulant* rational covariance extension problem, we review some basic results on *regular* rational covariance extension problem for multivariate stochastic processes. More details can be found in Blomqvist, Lindquist, and Nagamune [5], Byrnes and Lindquist [16].

Problem 4.2.1 (Multivariate rational covariance extension)**.** Given a sequence

$$C_0, C_1, \dots, C_n, \quad C_k \in \mathbb{C}^{m\times m} \text{ for } k = 0, \dots, n,$$

with C_0 Hermitian symmetric, such that the block Toeplitz matrix

$$\mathbf{T}_n = \begin{bmatrix} C_0 & C_1^* & C_2^* & \cdots & C_n^* \\ C_1 & C_0 & C_1^* & \cdots & C_{n-1}^* \\ C_2 & C_1 & C_0 & \cdots & C_{n-2}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_n & C_{n-1} & C_{n-2} & \cdots & C_0 \end{bmatrix} \tag{4.39}$$

is positive definite, find an infinite extension $C_{n+1}, C_{n+3}, C_{n+3}, \ldots$ such that, the series expansion

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{\infty} C_k e^{-ik\theta}, \quad C_{-k} = C_k^*, \tag{4.40}$$

converges for all $\theta \in [-\pi, \pi]$ to a positive $m \times m$ spectral density that takes the rational form

$$\Phi(z) = P(z)Q(z)^{-1}. \tag{4.41}$$

Notice that (4.40) imposes the following interpolation conditions on the spectral density Φ:

$$\int_{-\pi}^{\pi} e^{ik\theta}\Phi(e^{i\theta})\frac{d\theta}{2\pi} = C_k, \quad k = 0, 1, \ldots, n. \tag{4.42}$$

Thus, a generalized moment problem arises.

Here we take a first step in establishing a complete theory for the multivariable case. Indeed, for technical reasons, we confine our ARMA models to those whose transfer function has a matrix representation with a scalar numerator polynomial. Thus, P is a symmetric trigonometric polynomial of the form

$$P(e^{i\theta}) = \sum_{k=-n}^{n} p_k e^{-ik\theta}, \quad p_{-k} = \bar{p}_k, \tag{4.43}$$

of degree at most n, whereas Q is a symmetric trigonometric $m \times m$ matrix polynomial

$$Q(e^{i\theta}) = \sum_{k=-n}^{n} Q_k e^{-ik\theta}, \quad Q_{-k} = Q_k^*. \tag{4.44}$$

Let $\mathfrak{P}_+^{(m,n)}$ be the set of matrix polynomials (4.44) which are positive definite for all $\theta \in [-\pi, \pi]$. This is a convex cone, the closure of which we shall denote $\overline{\mathfrak{P}_+^{(m,n)}}$. Now, defining the trigonometric matrix polynomial

$$C(e^{i\theta}) = \sum_{k=-n}^{n} C_k e^{-ik\theta}, \quad C_{-k} = C_k^*, \tag{4.45}$$

we have

$$\langle C, Q\rangle := \int_{-\pi}^{\pi} \operatorname{Tr}\left[C(e^{i\theta})Q(e^{i\theta})^*\right]\frac{d\theta}{2\pi} = \sum_{k=-n}^{n} \operatorname{Tr}\left[C_k Q_k^*\right]. \tag{4.46}$$

If $Q \in \mathfrak{P}_+^{(m,n)}$, then there is a stable spectral factor

$$A(z) = A_0 + A_1 z^{-1} + \cdots + A_n z^{-n} \tag{4.47}$$

such that $Q(z) = A(z)A(z)^*$, and consequently

$$\langle C, Q\rangle = \int_{-\pi}^{\pi} \operatorname{Tr}\left[A(e^{i\theta})C(e^{i\theta})A(e^{i\theta})^*\right]\frac{d\theta}{2\pi} = \operatorname{Tr}\left[\mathbf{A}\mathbf{T}_n\mathbf{A}^*\right], \tag{4.48}$$

where $\mathbf{A} := (A_0, A_1, \dots, A_n)$. Let $\mathfrak{C}_+^{(m,n)}$ be the *interior* of the dual cone of all (4.45) such that

$$\langle C, Q\rangle \geqslant 0 \quad \text{for all } Q \in \overline{\mathfrak{P}_+^{(m,n)}}. \tag{4.49}$$

This is an open convex cone. It follows from (4.48) that $C \in \mathfrak{C}_+^{(m,n)}$ if and only if $\mathbf{T}_n$ is positive definite.

Next, consider the optimization problem to maximize the generalized entropy

$$\mathbb{I}_P(\Phi) = \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det \Phi(e^{i\theta})\frac{d\theta}{2\pi} \tag{4.50}$$

over all Φ that are positive definite on the unit circle subject to the moment conditions (4.42).

A fundamental result which allows to compute a solution for Problem 4.2.1 is the following

Theorem 4.2.1 (Blomqvist, Lindquist, and Nagamune [5]). *For each* $(P, C) \in \mathfrak{P}_+^{(1,n)} \times \mathfrak{C}_+^{(m,n)}$, *the problem to maximize* (4.50) *subject to the moment conditions* (4.42) *has a unique solution* $\hat{\Phi}$, *and it has the form*

$$\hat{\Phi}(z) = P(z)\hat{Q}(z)^{-1}, \tag{4.51}$$

where $\hat{Q} \in \mathfrak{P}_+^{(m,n)}$ *is the unique solution to the dual problem to minimize*

$$\mathbb{J}_P(Q) = \langle C, Q\rangle - \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det Q(e^{i\theta})\frac{d\theta}{2\pi} \tag{4.52}$$

over all $Q \in \mathfrak{P}_+^{(m,n)}$.

Consequently, a large subclass of all multivariable rational covariance extensions, namely those for which Φ takes the form (4.41), are completely parameterized by the $P \in \mathfrak{P}_+^{(1,n)}$.

4.3 THE MULTIVARIATE CIRCULANT RATIONAL COVARIANCE EXTENSION PROBLEM

Our purpose it to solve the following

Problem 4.3.1 (Multivariate circulant rational covariance extension problem)**.** Given $C := (C_0, C_1, \ldots, C_n) \in \mathfrak{C}_+^{(m,n)}$ for some $n < N$, find an $m \times m$ spectral density Φ of the form

$$\Phi(\zeta) = P(\zeta)Q(\zeta)^{-1}$$

such that

$$\int_{-\pi}^{\pi} e^{ik\theta}\Phi(e^{i\theta})d\nu = C_k, \quad k = 0, 1, 2, \ldots, n. \tag{4.53}$$

It turns out that this yields an extension

$$C_k = \int_{-\pi}^{\pi} e^{ik\theta}\Phi(e^{i\theta})d\nu, \quad k = n+1, n+2, \ldots, N \tag{4.54}$$

such that the banded Hermitian block-circulant matrix

$$\mathbf{C} = \mathrm{Circ}\{C_0, C_1, \ldots, C_n, 0, \ldots, 0, C_n^*, \ldots, C_1^*\} \tag{4.55}$$

with symbol (4.45) is extended to a Hermitian block-circulant matrix

$$\mathbf{\Sigma} := \mathrm{Circ}\{C_0, C_1, C_2, \ldots, C_N, C_{N-1}^*, \ldots, C_2^*, C_1^*\} \tag{4.56}$$

that is positive definite with symbol Φ.

Next we solve Problem 4.3.1 in terms of symbols. As already mentioned in Section 4.1, circulant rational covariance extension for multivariate processes can be recast as a block-circulant Toeplitz matrix completion problem. This will be the topic of Subsection 4.3.2.

4.3.1 Circulant rational covariance extension in terms of symbols

The main result of this section is that Problem 4.3.1 is feasible and that it can be solved by means of convex-optimization techniques.

Define the cone $\mathfrak{P}_+^{(m,n)}(N) \supset \mathfrak{P}_+^{(m,n)}$ of $m \times m$ matrix-valued trigonometric polynomials (4.44) such that

$$Q(\zeta_k) > 0 \quad k = -N+1, -N+2, \ldots, N. \tag{4.57}$$

Then $\mathfrak{P}_+^{(m,n)}(N) \supset \mathfrak{P}_+^{(m,n)}(2N) \supset \mathfrak{P}_+^{(m,n)}(4N) \supset \cdots \supset \mathfrak{P}_+^{(m,n)}$, and the corresponding dual cones satisfy

$$\mathfrak{C}_+^{(m,n)}(N) \subset \mathfrak{C}_+^{(m,n)}(2N) \subset \mathfrak{C}_+^{(m,n)}(4N) \subset \cdots \subset \mathfrak{C}_+^{(m,n)}. \tag{4.58}$$

The next theorem states that Problem 4.3.1 is feasible.

Theorem 4.3.1. *Let* $C \in \mathfrak{C}_+^{(m,n)}(N)$. *Then, for each* $P \in \mathfrak{P}_+^{(1,n)}(N)$, *there is a unique* $Q \in \mathfrak{P}_+^{(m,n)}(N)$ *such that*

$$\Phi = PQ^{-1} \tag{4.59}$$

satisfies the moment conditions (4.53).

Proof of Theorem 4.3.1. We begin by proving that the moment map $F^P : \mathfrak{P}_+^{(m,n)}(N) \to \mathfrak{C}_+^{(m,n)}(N)$, defined by

$$F_k^P(Q) = \int_{-\pi}^{\pi} e^{ik\theta} P(e^{i\theta}) Q(e^{i\theta})^{-1} d\nu, \quad k = 0, 1, \dots, n, \tag{4.60}$$

is proper for each fixed $P \in \mathfrak{P}_+^{(1,n)}(N)$; i.e., the inverse image $(F^P)^{-1}(K)$ is compact for any compact $K \subset \mathfrak{C}_+^{(m,n)}(N)$. To this end, we first show that $(F^P)^{-1}(K)$ is bounded. Clearly,

$$\langle F_P(Q), Q \rangle = \sum_{k=-n}^{n} \operatorname{Tr}\left[\left(Q_k^* \int_{-\pi}^{\pi} e^{ik\theta} P Q^{-1} d\nu\right)\right] = m \int_{-\pi}^{\pi} P d\nu =: \kappa \tag{4.61}$$

However, in view of (4.48) modified as in (4.14), $\langle F_P(Q), Q \rangle = \langle C, Q \rangle = \operatorname{Tr}[\mathbf{A}\mathbf{T}_n\mathbf{A}^*]$. Moreover, since K is compact, the eigenvalues of $\mathbf{T}_n > 0$ are bounded away from zero, and hence there is an $\varepsilon > 0$ such that $\mathbf{T}_n \geqslant \varepsilon I$ for any $C \in K$. Hence $\langle F_P(Q), Q \rangle \geqslant \varepsilon \|\mathbf{A}\|^2$, and therefore $\|\mathbf{A}\|^2 \leqslant \kappa/\varepsilon$. Since $\mathbf{A}$ is bounded, then so is $Q \in (F^P)^{-1}(K)$. Consequently, for any convergent sequence $\left(C^{(k)}\right)$ in K converging to $\hat{C}$, there is a convergent subsequence $\left(Q^{(k)}\right)$ in $(F^P)^{-1}(K)$ (for convenience also indexed by k) converging to some limit $\hat{Q}$. To prove properness, we need to show that $\hat{Q} \in (F^P)^{-1}(K)$, which can fail only if $\hat{Q}$ ends up on the boundary of $\mathfrak{P}_+^{(m,n)}(N)$; i.e., only if $\hat{Q}(\zeta_j) \geqslant 0$ is singular for some $j = -N+1, \dots, N$. We need to rule this out. Indeed, taking the limit,

$$\lim_{k\to\infty} \langle P I_m, C^{(k)} \rangle = \langle P I_m, \hat{C} \rangle = \int_{-\pi}^{\pi} P^2 \hat{Q}^{-1} d\nu, \tag{4.62}$$

which is finite, since $\left(C^{(k)}\right)$ belongs to the compact set K. However,

$$\int_{-\pi}^{\pi} P^2 \hat{Q}^{-1} d\nu = \frac{1}{2N} \sum_{j=-N+1}^{N} P(\zeta_j)^2 \hat{Q}(\zeta_j)^{-1}, \tag{4.63}$$

where $P(\zeta_j) > 0$. Hence $\hat{Q}(\zeta_j)$ cannot be singular. This establishes that the moment map F^P is proper.

Next we show that F^P is injective. To this end, note that

$$\delta^2 \mathbb{J}_P(Q; \delta Q) = \sum_{k=-n}^{n} \sum_{\ell=-n}^{n} \operatorname{Tr}\left[\left(\delta Q_\ell \int_{-\pi}^{\pi} e^{i(\ell-k)\theta} P Q^{-2} d\nu\, \delta Q_k^*\right)\right] > 0 \tag{4.64}$$

That is, the Hessian is positive definite, and hence $\mathbb{J}_P$ is strictly convex. Therefore the moment map F^P is injective.

Since $F^P : \mathfrak{P}_+^{(m,n)}(N) \to \mathfrak{C}_+^{(m,n)}(N)$ is a continuous, injective, proper map between spaces of the same finite dimension, it is a homeomorphism Byrnes and Linquist [18, Theorem 2.6]. ∎

The following theorem provides an algorithm for computing the solution.

Theorem 4.3.2. *For each* $(P, C) \in \mathfrak{P}_+^{(1,n)}(N) \times \mathfrak{C}_+^{(m,n)}(N)$, *the problem to maximize the functional*

$$\mathbb{I}_P(\Phi) = \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det \Phi(e^{i\theta}) d\nu \tag{4.65}$$

subject to the moment conditions (4.53) *has a unique solution* $\hat{\Phi}$, *and it has the form*

$$\hat{\Phi}(z) = P(z)\hat{Q}(z)^{-1}, \tag{4.66}$$

where $\hat{Q} \in \mathfrak{P}_+^{(m,n)}(N)$ *is the unique solution to the dual problem to minimize*

$$\mathbb{J}_P(Q) = \langle C, Q\rangle - \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det Q(e^{i\theta}) d\nu \tag{4.67}$$

over all $Q \in \mathfrak{P}_+^{(m,n)}(N)$.

Proof of Theorem 4.3.2. consider the (primal) problem to maximize the generalized entropy gain

$$\mathbb{I}_P(\Phi) = \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det \Phi(e^{i\theta}) d\nu \tag{4.68}$$

subject to the moment conditions (4.53). The corresponding Lagrangian is then given by

$$\begin{aligned} L(\Phi, Q) &= \mathbb{I}_P(\Phi) + \operatorname{Tr}\left[\sum_{k=-n}^{n} Q_k^* \left(C_k - \int_{-\pi}^{\pi} e^{ik\theta}\Phi(e^{i\theta}) d\nu\right)\right] \\ &= \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det \Phi(e^{i\theta}) d\nu + \langle C, Q\rangle - \int_{-\pi}^{\pi} \operatorname{Tr}\left[Q(e^{i\theta})\Phi(e^{i\theta})\right] d\nu, \end{aligned} \tag{4.69}$$

where $Q_0, Q_1, \dots, Q_n$ are Lagrange multipliers, and where Q is defined as in (4.44) with $Q_{-k} = Q_k^*$. Since $\log\det\Phi = \operatorname{Tr}[\log\Phi]$, the Lagrangian may be written

$$L(\Phi, Q) = \operatorname{Tr}\left[\int_{-\pi}^{\pi} \left[P(e^{i\theta}) \log\Phi(e^{i\theta}) - Q(e^{i\theta})\Phi(e^{i\theta})\right] d\nu\right] + \langle C, Q\rangle \tag{4.70}$$

Since the dual functional $\sup_\Phi L(\Phi, Q)$ is finite only if $Q \in \overline{\mathfrak{P}_+(N)}$, we may restrict the Lagrange multipliers to that set. Therefore, for each $Q \in \overline{\mathfrak{P}_+(N)}$, consider the directional derivative

$$\delta L(\Phi, Q; \delta\Phi) = \operatorname{Tr}\left[\int_{-\pi}^{\pi} \left(P\Phi^{-1} - Q\right) \delta\Phi \, d\nu\right],$$

which equals zero for all variations $\delta\Phi$ if and only if (4.59), which inserted into (4.70) yields

$$\sup_\Phi L(\Phi, Q) = \mathbb{J}_P(Q) + \int_{-\pi}^{\pi} P(e^{i\theta}) \left[\log\det P(e^{i\theta}) - 1\right] d\nu, \tag{4.71}$$

where

$$\mathbb{J}_P(Q) = \langle C, Q\rangle - \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det Q(e^{i\theta}) d\nu \tag{4.72}$$

and the last term is constant. Hence we may take (4.72) as the dual functional.

Taking the directional derivative of (4.72) we have

$$\delta\mathbb{J}_P(Q; \delta Q) = \operatorname{Tr}\left[\sum_{k=-n}^{n} \left(C_k - \int_{-\pi}^{\pi} e^{ik\theta} P Q^{-1} d\nu\right) \delta Q_k^*\right], \tag{4.73}$$

which shows that provided there is a stationary point $\hat{Q}$ it will have the property that $\hat{\Phi} := P\hat{Q}^{-1}$ satisfies the moment conditions (4.53). However, by Theorem 4.3.1, there is a unique such solution. By standard duality theory we see that $\hat{\Phi}$ is the optimal solution of the primal problem. ■

4.3.2 Circulant rational covariance extension in terms of matrices

Next we reformulate the optimization problems in terms of circulant matrices. To this end, we define the circulant matrix

$$\mathbf{\Sigma} = \frac{1}{2N} \mathbf{F}^* \operatorname{diag}\big(\Phi(\zeta_{-N+1}), \dots, \Phi(\zeta_N)\big)\mathbf{F} \tag{4.74}$$

with symbol (4.59) and the banded numerator matrix

$$\mathbf{P} = \frac{1}{2N} \mathbf{F}^* \operatorname{diag}\big(I_m \otimes P(\zeta_{-N+1}), \dots, I_m \otimes P(\zeta_N)\big)\mathbf{F} \tag{4.75}$$

of degree at most n with symbol $P(\zeta)I_m$, where the scalar pseudo-polynomial P is given by (4.43). It can also be shown that

$$\log \mathbf{\Sigma} = \frac{1}{2N} \mathbf{F}^* \operatorname{diag}\big(\log \Phi(\zeta_{-N+1}), \dots, \log \Phi(\zeta_N)\big)\mathbf{F}. \tag{4.76}$$

Therefore, since $\log\det\Phi = \operatorname{Tr}[\log\Phi]$, the primal functional (4.68) may be written

$$\begin{aligned}\int_{-\pi}^{\pi} P(e^{i\theta})\log\det\Phi(e^{i\theta})\,d\nu &\\ = \frac{1}{2N}\sum_{j=-N+1}^{N}\operatorname{Tr}\left[P(\zeta_j)\log\Phi(\zeta_j)\right] &\\ = \frac{1}{2N}\operatorname{Tr}\left[\mathbf{P}\log\mathbf{\Sigma}\right] &\end{aligned} \tag{4.77}$$

and the moment conditions (4.53) as

$$\frac{1}{2N}\operatorname{Tr}\left[\mathbf{S}^k\mathbf{\Sigma}\right] = C_k, \quad k = 0, 1, \dots, n, \tag{4.78}$$

or, equivalently, as

$$\mathbf{E}_n^\top\mathbf{\Sigma}\mathbf{E}_n = \mathbf{T}_n, \quad \text{where } \mathbf{E}_n = \begin{bmatrix}\mathbf{I}_{mn}\\ \mathbf{0}\end{bmatrix}. \tag{4.79}$$

Consequently, the primal problem amounts to maximizing $\operatorname{Tr}[\mathbf{P}\log\mathbf{\Sigma}]$ over all Hermitian, positive definite $2mN\times 2mN$ block matrices subject to (4.78) or (4.79). This reduces to the primal problem presented in Carli et al. [24] in the special case $P\equiv 1$, except that in Carli et al. [24] there is an extra condition insuring that $\mathbf{\Sigma}$ is circulant. However, in Carli and Georgiou [23] it was shown that this condition is automatically satisfied and is therefore not needed.

Similarly the dual functional (4.72) can be written

$$\begin{aligned}\int_{-\pi}^{\pi} C(e^{i\theta})Q(e^{i\theta})^*\,d\nu - \int_{-\pi}^{\pi} P(e^{i\theta})\log\det Q(e^{i\theta})\,d\nu &\\ = \frac{1}{2N}\operatorname{Tr}[\mathbf{C}\mathbf{Q}] - \frac{1}{2N}\operatorname{Tr}[\mathbf{P}\log\mathbf{Q}], &\end{aligned} \tag{4.80}$$

where

$$\mathbf{Q} = \frac{1}{2N}\mathbf{F}^*\operatorname{diag}\big(Q(\zeta_{-N+1}),\dots,Q(\zeta_N)\big)\mathbf{F} \tag{4.81}$$

and $\mathbf{C}$ is the banded circulant block matrix (4.55) formed from $C_0, C_1, \dots, C_n$. Therefore, given $C\in\mathfrak{C}_+(N)$, it follows from Theorem 4.3.1 that, for each Hermitian, positive-definite circulant block matrix $\mathbf{P}$ with symbol of the form $P(\zeta)I_m$, where P is a pseudo-polynomial of degree at most n, there is a unique $\mathbf{\Sigma}$ given by

$$\mathbf{\Sigma} = \mathbf{Q}^{-1}\mathbf{P}, \tag{4.82}$$

where $\mathbf{Q}$ is the unique solution of the problem to minimize

$$\mathbb{J}_{\mathbf{P}}(\mathbf{Q}) = \frac{1}{2N}\operatorname{Tr}[\mathbf{C}\mathbf{Q}] - \frac{1}{2N}\operatorname{Tr}[\mathbf{P}\log\mathbf{Q}] \tag{4.83}$$

over all Hermitian, circulant block-banded matrices

$$\mathbf{Q} = \mathrm{Circ}\{Q_0, Q_1, \dots, Q_n, 0, \dots, 0, Q_n^*, Q_{n-1}^*, \dots, Q_1^*\}$$

that are positive definite. For the maximum-entropy solution corresponding to $\mathbf{P} = \mathbf{I}$ this reduces to an optimization problem that is different from the one presented in Carli et al. [24].

As observed in Carli et al. [24] the condition $\mathbf{T}_n > 0$ is necessary, but not a sufficient, for feasibility of the circulant block-banded covariance extension problem. In the present setting we see that the Toeplitz condition $\mathbf{T}_n > 0$ is equivalent to $C \in \mathfrak{C}_+^{(m,n)}$, whereas, by Theorem 4.3.1, $C \in \mathfrak{C}_+^{(m,n)}(N)$ is required for feasibility. Since $\mathfrak{C}_+^{(m,n)}(N) \subset \mathfrak{C}_+^{(m,n)}$, it follows that the Toeplitz condition cannot be sufficient in general. However, as proved in Carli et al. [24], feasibility is achieved for a sufficiently large N. This can also be seen by noting that the set $\{\zeta_j;\ j = -N+1, \dots, N\}$ becomes dense on the unit circle as $N \to \infty$, and therefore $\mathfrak{P}_+(N) \to \mathfrak{P}_+$. Consequently, $\mathfrak{C}_+(N) \to \mathfrak{C}_+$, and the convergence is monotone in the sense of (4.58). Therefore, since $\mathfrak{C}_+$ is an open set, there is an N_0 such that any $C \in \mathfrak{C}_+$ will sooner or later end up in $\mathfrak{C}_+(N)$ and remain there as $N \geqslant N_0$ increases.

4.4 DETERMINING $\mathbf{P}$ FROM LOGARITHMIC MOMENTS

We have parameterized a large class of solutions to the multivariable circulant rational covariance extension problem in a smooth manner by the numerator trigonometric polynomials $P \in \mathfrak{P}_+^{(1,n)}(N)$, or, equivalently, by their corresponding banded circulant matrices $\mathbf{P}$. Next, we show how P can be determined from the logarithmic moments

$$\gamma_k = \int_{-\pi}^{\pi} e^{ik\theta} \log\det \Phi(e^{i\theta}) d\nu, \quad k = 1, 2, \dots, n. \tag{4.84}$$

Such moments are known as *cepstral coefficients* in speech processing. Let $\Gamma(\zeta)$ be the pseudo-polynomial

$$\Gamma(\zeta) = \sum_{k=-n}^{n} \gamma_k \zeta^{-k}, \tag{4.85}$$

where $\gamma_{-k} = \bar{\gamma}_k$, $k = 1, 2, \dots, n$ and γ_0 is real.

Consider the problem of finding the spectral density Φ, or, equivalently, the circulant block matrix $\mathbf{\Sigma}$, that maximizes the entropy gain

$$\mathbb{I}(\Phi) = \int_{-\pi}^{\pi} \log\det \Phi(e^{i\theta}) d\nu = \frac{1}{2N} \mathrm{Tr}\,[\log \mathbf{\Sigma}] \tag{4.86}$$

subject to the two sets of moment conditions (4.53) and (4.84). Such a problem was apparently first considered in the usual trigonometric moment setting in an unpublished technical report Musicus and Kabel [79] and then, independently and in a more elaborate form, in Byrnes, Enqvist, and Linquist [9, 10], Enqvist [36].

Setting up the Lagrangian a straightforward calculation yields the dual problem to minimize

$$\begin{aligned}\mathbb{J}(P,Q) = \langle C, Q\rangle - \int_{-\pi}^{\pi} P(e^{i\theta}) \log\det Q(e^{i\theta}) d\nu \\ -\langle \Gamma, P\rangle + \int_{-\pi}^{\pi} P(e^{i\theta}) \log P(e^{i\theta}) d\nu,\end{aligned} \tag{4.87}$$

over $(P,Q) \in \hat{\mathfrak{P}}_+^{(1,n)}(N) \times \mathfrak{P}_+^{(m,n)}(N)$, where $\hat{\mathfrak{P}}_+^{(1,n)}(N)$ is the bounded subset

$$\hat{\mathfrak{P}}_+^{(1,n)}(N) := \{P \in \mathfrak{P}_+^{(1,n)}(N) \mid p_0 = 1\} \tag{4.88}$$

of the cone $\mathfrak{P}_+^{(1,n)}(N)$.

The following theorem is a multivariable version of Theorem 8 in Lindquist and Picci [74] and the proof is analogous.

Theorem 4.4.1. *Suppose that* $C \in \mathfrak{C}_+^{(m,n)}(N)$ *and that* $\gamma_1, \dots, \gamma_n$ *are complex numbers. Then there exists a solution* $(\hat{P}, \hat{Q})$ *that minimizes* $\mathbb{J}(P,Q)$ *over all* $(P,Q) \in \overline{\hat{\mathfrak{P}}_+^{(1,n)}(N)} \times \overline{\mathfrak{P}_+^{(m,n)}(N)}$, *and, for any such solution*

$$\hat{\Phi} = \hat{P}\hat{Q}^{-1} \tag{4.89}$$

satisfies the covariance moment conditions (4.53). *If, in addition,* $\hat{P} \in \mathfrak{P}_+^{(1,n)}(N)$, (4.89) *also satisfies the logarithmic moment conditions* (4.84) *and is an optimal solution of the primal problem to maximize the entropy gain* (4.86) *given* (4.53) *and* (4.84). *Then* $\hat{Q} \in \mathfrak{P}_+^{(m,n)}(N)$, *and the solution is unique. In fact,* $\mathbb{J}$ *is strictly convex on* $\hat{\mathfrak{P}}_+^{(1,n)}(N) \times \mathfrak{P}_+^{(m,n)}(N)$.

Provided $C \in \mathfrak{C}_+(N)$, minimizing $\mathbb{J}(P,Q)$ over all $(P,Q) \in \overline{\hat{\mathfrak{P}}_+^{(1,n)}(N)} \times \overline{\mathfrak{P}_+^{(m,n)}(N)}$ will always produce a spectral density with the prescribed covariance lags $C_0, C_1, \dots, C_n$. If the moments $C_0, C_1, \dots, C_n$ and $\gamma_1, \dots, \gamma_n$ come from the same theoretical spectral density, the optimal solution (4.89) will also match the cepstral coefficients. In practice, however, they will be estimated from different data sets, so there is no guarantee that $\hat{P}$ does not end up on the boundary of $\mathfrak{P}_+^{(1,n)}(N)$ without satisfying the logarithmic moment conditions. Then the problem needs to be regularized, leading to adjusted values of $\gamma_1, \dots, \gamma_n$ consistent with the covariances $C_0, C_1, \dots, C_n$.

Such a regularization was proposed in Enqvist [36] in the context of the usual rational covariance extension problem. The regularized dual problem to find a pair $(P, Q) \in \hat{\mathfrak{P}}_+^{(1,n)}(N) \times \mathfrak{P}_+^{(m,n)}(N)$ minimizing

$$\mathbf{J}_\lambda(P, Q) = \mathbb{J}(P, Q) - \lambda \int_{-\pi}^{\pi} \log P(e^{i\theta}) d\nu \tag{4.90}$$

for some $\lambda > 0$ will always lead to a solution where $P \in \mathfrak{P}_+^{(1,n)}(N)$. Indeed, (4.90) will take an infinite value for $P \in \partial\mathfrak{P}_+^{(1,n)}(N)$, since then $P(\zeta_k) = 0$ for some k, and hence the minimum will be in the interior. In circulant form (4.90) becomes

$$\begin{aligned} \mathbf{J}_\lambda(P, Q) = &\frac{1}{2N}\mathrm{Tr}\,[\mathbf{CQ}] - \frac{1}{2N}\mathrm{Tr}\,[\mathbf{\Gamma P}] \\ &+ \frac{1}{2N}\mathrm{Tr}\left[\mathbf{P}\log\mathbf{PQ}^{-1}\right] - \frac{\lambda}{2N}\mathrm{Tr}\,[\log\mathbf{P}], \end{aligned} \tag{4.91}$$

where

$$\mathbf{\Gamma} = \frac{1}{2N}\mathbf{F}^* \mathrm{diag}\big(I_m \otimes \Gamma(\zeta_{-N+1}), \dots, I_m \otimes \Gamma(\zeta_N)\big)\mathbf{F}. \tag{4.92}$$

Then both sets (4.53) and (4.84) of moments are matched provided one adjusts the logarithmic moments $\gamma_1, \gamma_2, \dots, \gamma_n$ to $\gamma_1 + \varepsilon_1, \gamma_2 + \varepsilon_2, \dots, \gamma_n + \varepsilon_n$, where

$$\begin{aligned} \varepsilon_k = \int_{-\pi}^{\pi} e^{ik\theta} \frac{\lambda}{P(e^{i\theta})} d\nu &= \frac{\lambda}{2N} \sum_{j=-N+1}^{N} \frac{\zeta_j^k}{P(\zeta_j)} \\ &= \frac{\lambda}{2N}\mathrm{Tr}\left[\mathbf{S}^k\mathbf{P}^{-1}\right]. \end{aligned} \tag{4.93}$$

4.5 IMPLEMENTATION DETAILS

Simulations suggest that a straightforward implementation of Newton-like algorithm with backtracking line search can be heavily affected by the high condition number of the Hessian. Unfortunately, ill-conditioning is often encountered when dealing with this kind of optimization problems, as pointed out in Enqvist [34]. In that paper, an effective homotopy continuation method was proposed for solving rational covariance extension with degree constraint. A generalization of this approach is not obvious in the case we are interested in, thus we propose to rephrase the problem so that it can be solved efficiently by means of robust off-the-shelf optimization methods as the ones featured by MATLAB.

For the sake of simplicity, next we deal with the minimization of (4.72) in case $\mathbf{y}$ is scalar and real-valued. The same procedure was extended to the general case with $\mathbf{y}$ taking values in $\mathbb{C}^m$ and to the minimization of (4.87). In order to be consistent

with the implementation of DFT featured by MATLAB's `fft` method, in this section we consider $T = 2N$ and $\zeta_k := e^{j\frac{2\pi}{T}k}$, with k on the interval $[0, T-1]$ instead of $[-N+1, N]$. Thus, the values $Q(\zeta_k)$ for $k = 0, \dots, T-1$, are given by the DFT of the T-dimensional sequence

$$\mathbf{q} := \begin{bmatrix} q_0 & q_1 & \dots & q_n & 0 & \dots & 0 & q_n & \dots & q_1 \end{bmatrix}. \tag{4.94}$$

Let us introduce the notations

$$\boldsymbol{\gamma} := \begin{bmatrix} C(\zeta_0) & C(\zeta_1) & \dots & C(\zeta_{T-1}) \end{bmatrix}^\top, \tag{4.95}$$

$$\boldsymbol{\rho} := \begin{bmatrix} P(\zeta_0) & P(\zeta_1) & \dots & P(\zeta_{T-1}) \end{bmatrix}^\top. \tag{4.96}$$

Now define the array $\mathbf{x}$ such that

$$x_k := Q(\zeta_k), \quad \text{for } k = 0, 1, \dots, T-1. \tag{4.97}$$

Then, minimizing (4.72) is equivalent to minimize

$$\tilde{J}_P(\mathbf{x}) := \boldsymbol{\gamma}^\top \mathbf{x} - \boldsymbol{\rho}^\top \log \mathbf{x}, \tag{4.98}$$

where

$$\log \mathbf{x} := \begin{bmatrix} \log x_0 & \log x_1 & \dots & \log x_{T-1} \end{bmatrix}$$

under the following constraints:

POSITIVITY : Since $Q(\zeta_k)$ has to be positive on the discretized unit circle, we require

$$x_k > 0 \quad \text{for k=0,1,...,T-1}$$

SYMMETRY : Symmetry of $Q(\zeta_k)$ requires that $x_k = x_{T-k}$ for $k = 1, \dots, \frac{T}{2} - 1$. Thus, we impose the constraint

$$A\mathbf{x} = 0 \tag{4.99}$$

where A is the $\frac{T}{2} - 1 \times T$ matrix defined by

$$A := \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & -1 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & -1 & 0 \\ \vdots & \vdots & & \ddots & & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 & -1 & \dots & 0 & 0 \end{bmatrix}. \tag{4.100}$$

ORDER OF $Q(\zeta)$: Recall that $Q(z) = \sum_{k=-n}^{n} q_k \zeta^{-k}$ and the sequence of q_k's is given by inverse DFT of $\mathbf{x}$. Thus, the constraint on the order of the pseudo-polynomial Q can be expressed in terms of the DFT matrix $\mathbf{F}$ as

$$[\mathbf{F}^*]_{n+1:N-n-1}\, \mathbf{x} = 0, \tag{4.101}$$

where $[\mathbf{F}^*]_{n+1:N-n-1}$ stands for matrix given by the rows of $\mathbf{F}^*$ of index $n+1, n+2, \dots, T-n-1$. However, standard optimization software requires the matrices appearing in the constraints to have real valued entries. Therefore, based on the structure of the inverse DFT matrix the last constraint can be written as

$$M\mathbf{x} = 0, \tag{4.102}$$

where

$$M := \begin{bmatrix} \Re\left\{[\mathbf{F}^*]_{n+1:N-n-1}\right\} \\ \Im\left\{[\mathbf{F}^*]_{n+1:N-n-1}\right\} \end{bmatrix}$$

In conclusion, we can rephrase the original problem of minimizing (4.72) over $\mathfrak{P}_+^{(1,n)}(N)$ so that it reads

Problem 4.5.1.

$$\mathbf{x}^\circ = \arg\min_{\mathbf{x}} \boldsymbol{\gamma}^\top \mathbf{x} - \boldsymbol{\rho}^\top \log \mathbf{x}$$
$$\text{s.t.} \quad \begin{cases} \begin{bmatrix} A \\ M \end{bmatrix} \mathbf{x} = 0 \\ \mathbf{x} > 0 \end{cases} \tag{4.103}$$

Thus standard optimization methods can be used for solving it efficiently. For example, we resorted to MATLAB's `fmincon`. Finally, the coefficients $q_0, q_1, \dots, q_n$ are obtained by inverse DFT of $\mathbf{x}^\circ$, according to the pattern shown in (4.94).

4.6 NUMERICAL EXAMPLES

Given a $P \in \mathfrak{P}_+^{(1,n)}$ and a sequence $C_0, C_1, \dots, C_n$ of $m \times m$ covariance lags with a positive definite block Toeplitz matrix (4.39), Theorem 4.2.1 states that there is a unique $Q \in \mathfrak{P}_+^{(m,n)}$ such that $\Phi := PQ^{-1}$ satisfies the moment conditions (4.42). As pointed out above, for a sufficiently large N the sequence C will also belong to the somewhat smaller cone $\mathfrak{C}_+^{(m,n)}(N)$, and then, by Theorem 4.3.1, there will be a unique $Q_N \in \mathfrak{P}_+^{(m,n)}(N)$ such that $\Phi_N := PQ_N^{-1}$ satisfies (4.53). Next we shall give some numerical results illustrating how Φ can be approximated by Φ_N for various values of N.

In our first example Φ is a 2×2 spectral density corresponding to an AR process of order $n = 8$ with poles as depicted in Fig. 10.

Given the theoretical covariance sequence $C_0, C_1, \dots, C_n$ from this Φ, we solve the corresponding circulant moment problem (4.53) for various values of N to obtain a

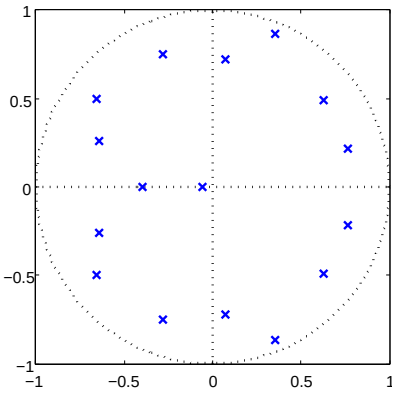


Figure 10: Autoregressive 2×2 model, with order $n = 8$.

bilateral AR representation of order $n = 8$ with spectral density Φ_N. Fig. 11 illustrates the approximation error $\|\Phi(e^{i\theta}) - \hat{\Phi}(e^{i\theta})\|_2$ for $N = 16, 32$ and 64. It turns out that there is no need to go for high values of N.

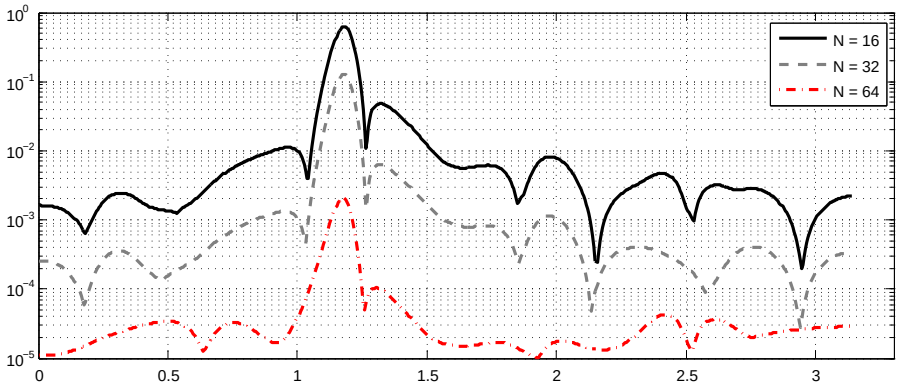


Figure 11: Norm of the spectral estimation error for bilateral AR models with $N = 16, 32, 64$.

In the second example we start from a two-dimensional ARMA process with a spectral density $\Phi := PQ^{-1}$, where P is a scalar pseudo-polynomial of degree three and Q is a 2×2 matrix-valued pseudo-polynomial of degree $n = 6$. Its zero poles map is illustrated in Fig. 12. Given its covariance sequence $C_0, C_1, \ldots, C_n$ and cepstral sequence $\gamma_1, \gamma_2, \ldots, \gamma_n$, we apply the combined covariance and cepstral procedure described in Section 4.4 to determine a pair (P_N, Q_N) for $n = 6$ and a corresponding bilateral ARMA model. For comparison we also compute an bilateral AR approximation with $n = 12$ fixing $P = 1$. As illustrated in Fig. 13, the bilateral ARMA model of order $n = 6$ computed for $N = 32$ compares favorably to the bilateral AR model with $n = 12$ which is obtained by fixing $N = 64$.

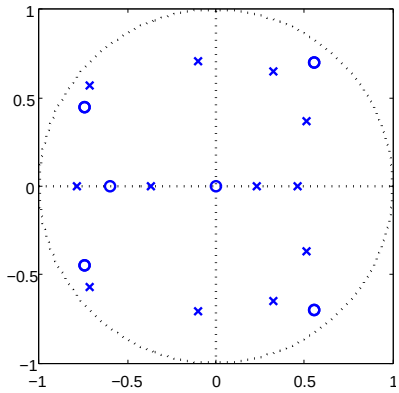


Figure 12: ARMA 2×2 model, with order $n = 6$.

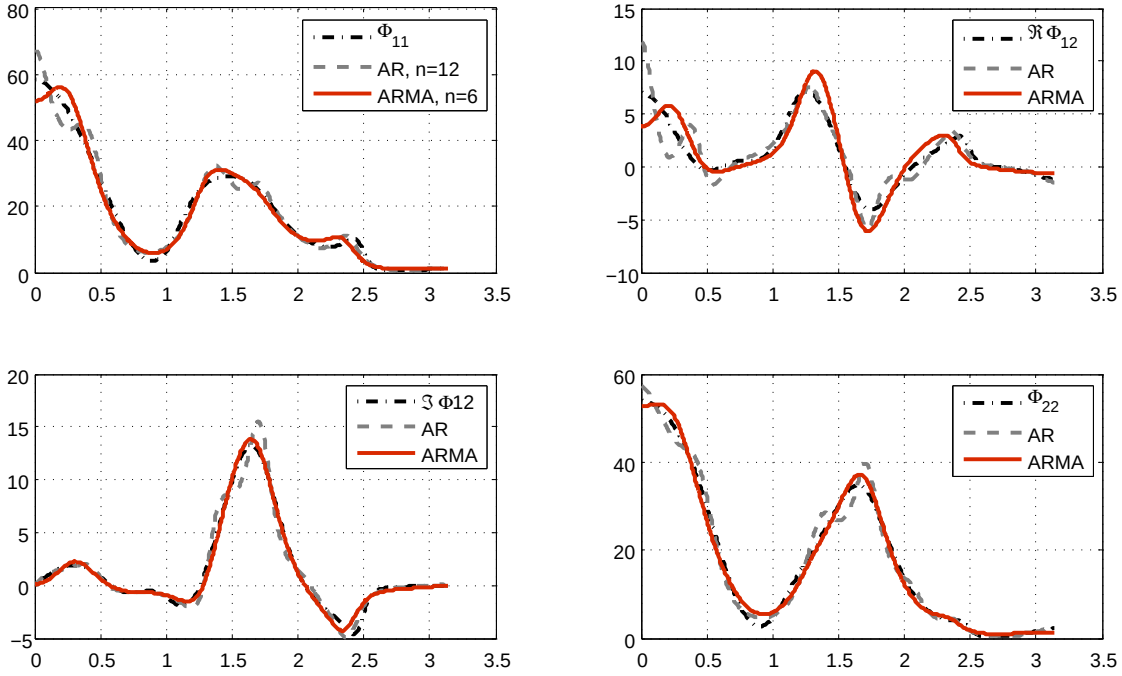


Figure 13: Comparison between a bilateral AR of order 12 for $N = 64$ and a bilateral ARMA of order 6 for $N = 32$: estimated spectral densities.

4.7 CONCLUSIONS AND FUTURE WORK

In this chapter we took a first step towards multivariate circulant rational covariance extension, which provides an effective method for computing an approximate solution of the regular covariance extension problem for multivariate, stationary processes. Indeed, it results in a convex optimization problem which can be solved efficiently by means of fast routines for DFT and off-the-shelf optimization algorithms.

The next step is to allow for arbitrary pseudo-polynomials $P(\zeta)$. In addition, as stated in Subsection 4.3.2, the set $\{\zeta_j;\ j = -N+1,\dots,N\}$ becomes dense on the unit circle as $N \to \infty$, and therefore $\mathfrak{P}_+(N) \to \mathfrak{P}_+$. Thus, there is an N_0 such that any $C \in \mathfrak{C}_+$ will end up in $\mathfrak{C}_+(N)$. Either determining such N_0 or being able of establishing whether, given N, a sequence $C \in \mathfrak{C}_+$ also belongs to $\mathfrak{C}_+(N)$ is essential in order to guarantee the feasibility of Problem 4.3.1. Moreover, it would be very significant to evaluate analytically the convergence of the approximate solution to the actual one (i.e. the one corresponding to the regular covariance extension problem).

The multivariate circulant rational covariance extension problem, however, is also interesting *per se*. In particular, one of the most promising research directions is establishing a connection between the proposed approach and the literature about reciprocal processes defined on the discretized unit circle (Carli et al. [24], Carli et al. [25], Chiuso, Ferrante, and Picci [26], Krener [66], Levy and Ferrante [70], Levy, Frezza, and Krener [71], Picci and Carli [86], Sand [93]). A regular reciprocal process $\mathbf{y}(t)$ can be considered as a one-dimensional Markov field. Here we are interested in stationary periodic reciprocal processes, which are defined on a finite interval $U := [-N+1,\dots,N]$ with arithmetic modulo $2N$. It is convenient to represent them as processes defined on the discretized circle by folding the interval U. Consider $t_1 < t_2$ such that $t_1, t_2 \in U$ and define

$$\begin{aligned} \mathbf{y}_{(t_1,t_2)} &: \text{values taken by } \mathbf{y}(t) \text{ on } (t_1,t_2); \\ \mathbf{y}_{(t_1,t_2)^c} &: \text{values taken by } \mathbf{y}(t) \text{ in the complementary set of } (t_1,t_2) \text{ in } U. \end{aligned}$$

Now let "$\perp$" denote independence. Then, a reciprocal process of order n defined on the discretized unit circle corresponding to U with arithmetic modulo $2N$ is characterized by the following property:

$$\mathbf{y}_{(t_1,t_2)} \perp \mathbf{y}_{(t_1,t_2)^c} \mid \{\mathbf{y}_{(t_1-n,t_1]} \vee \mathbf{y}_{[t_2,t_2+n)}\}.$$

For instance, if $n = 1$, we have that the values taken by the process inside the interval (t_1,t_2) are conditional independent of the values taken outside the interval, given $\mathbf{y}_{t_1}$

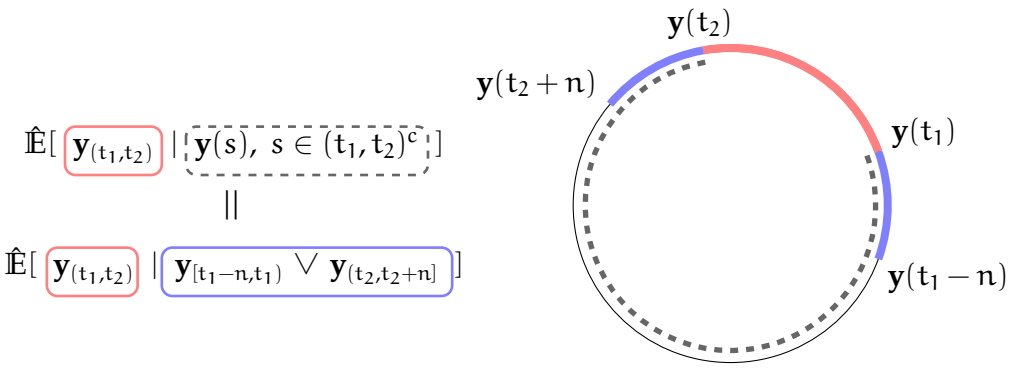


Figure 14: Gaussian reciprocal processes on the discretized unit circle

and $\mathbf{y}_{t_2}$. In case a reciprocal process of order n is Gaussian, conditional independence implies conditional uncorrelation, and so we have

$$\hat{\mathbb{E}}\left[\mathbf{y}_{(t_1,t_2)}\,|\,\mathbf{y}_{(t_1,t_2)^c}\right] = \hat{\mathbb{E}}\left[\mathbf{y}_{(t_1,t_2)}\,|\,\{\mathbf{y}_{(t_1-n,t_1]} \vee \mathbf{y}_{[t_2,t_2+n)}\}\right] \tag{4.104}$$

as show in Fig. 14.

A Gaussian reciprocal process $\mathbf{y}(t)$ of order n defined on $[-N+1, N]$ admits the representation (see e.g. Carli et al. [24])

$$\sum_{k=-n}^{n} Q_k \mathbf{y}(t-k) = \mathbf{e}(t), \quad t \in \mathbb{Z}_{2N}, \tag{4.105}$$

where $\mathbf{e}(t)$ is noise with correlation bandwidth equal to n. In terms of matrices, we have

$$Q\mathbf{y} = \mathbf{e}, \tag{4.106}$$

where

$$\mathbf{y} := \begin{bmatrix} \mathbf{y}(-N+1) \\ \vdots \\ \mathbf{y}(0) \\ \vdots \\ \mathbf{y}(N) \end{bmatrix}, \quad \mathbf{e} := \begin{bmatrix} \mathbf{e}(-N+1) \\ \vdots \\ \mathbf{e}(0) \\ \vdots \\ \mathbf{e}(N) \end{bmatrix}, \tag{4.107}$$

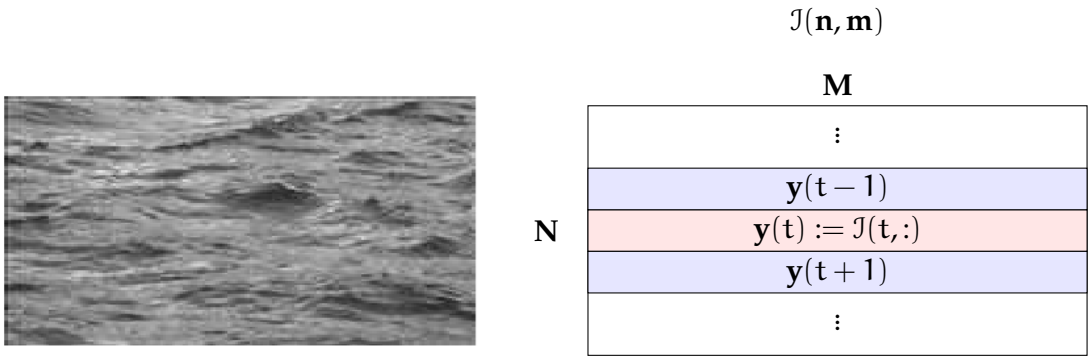
$$Q := \mathrm{Circ}\{Q_0, Q_1, \dots, Q_n, 0, \dots, 0, Q_n^*, \dots, Q_1^*\}. \tag{4.108}$$

Since $\mathbb{E}[\mathbf{y}\mathbf{e}] = I$, we have

$$\mathbb{E}[\mathbf{e}\mathbf{e}^*] = \mathbb{E}[Q\mathbf{y}\mathbf{e}^*] = Q \tag{4.109}$$

and

$$\Sigma := \mathbb{E}[\mathbf{y}\mathbf{y}^*] = Q^{-1}. \tag{4.110}$$



(a) Example of texture

(b) Texture modeled by means of a reciprocal process.

Figure 15: Application of reciprocal processes to image processing of textures

Recall that, after solving the circulant rational covariance extension problem, we end up with a *bilateral* ARMA model:

$$\sum_{k=-n}^{n} Q_k \mathbf{y}(t-k) = \sum_{k=-n}^{n} P_k \mathbf{e}(t-k), \quad t \in \mathbb{Z}_{2N}. \tag{4.111}$$

Thus, it seems that the approach we propose yields models that may somewhat generalize (4.105) and (4.106). This is a subject of current research.

If such connection can be established, we could take advantage of the proposed approach in dealing with the problems where reciprocal processes have been successfully applied, such as image processing of textures. Indeed, textures can be considered spatially stationary, periodic processes (see e.g. Chiuso, Ferrante, and Picci [26], Picci and Carli [86]). An example of texture is shown in Fig. 15a, taken from Picci and Carli [86]. In that paper, textures are modeled by introducing a Gaussian reciprocal process $\mathbf{y}(t) = \{\mathbf{y}(k); k \in [0,\dots,N]\}$ of order one, where $\mathbf{y}(k)$ is an M-dimensional random vector corresponding to the k-th row in the image, as shown in Fig. 15b. Then, a bilateral autoregressive model of the same kind as (4.105) is estimated by means of an *ad hoc* algorithm. This problem could be solved by means of the approach we propose. Moreover, it would be interesting to figure out whether the moving average part that appears in (4.111) can play a role in achieving better results.

Part II

Assessment of the performance of physical layer authentication over Rayleigh fading channels

5 ON THE ACHIEVABLE ERROR REGION OF PHYSICAL LAYER AUTHENTICATION

5.1 INTRODUCTION

Physical layer security refers to secure communication techniques which are implemented at the lowest layer of OSI (Open Systems Interconnection) reference model shown in Fig. 16 (ISO/IEC 7498-1, see e.g. Zimmermann [104]). This topic keeps arousing great interest in modern communications. Indeed, it provides an effective defense mechanism which is complementary to higher layer security techniques. On the one hand, it has the potential of resisting the attacks based on massive computational capabilities that may be feasible in the near future, e.g., by quantum computing. On the other, security implemented at the physical layer is usually based on information theoretic arguments, so it entails analytically predictable performance irrespective of the attacker capabilities.

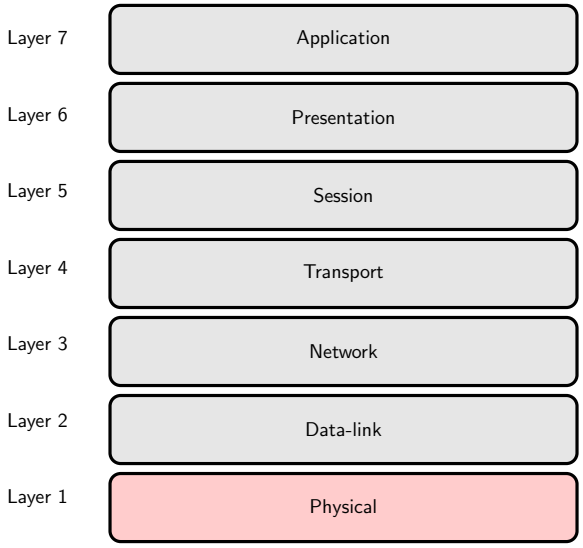


Figure 16: OSI reference model ISO/IEC 7498-1.

Next we focus on the authentication of the message source based on channel estimation, that hinges on using physical layer features as authentication keys. This is one

of the most desirable mechanisms of physical layer security. Indeed, its performance can be evaluated analytically by means of information theoretic results. Moreover, it significantly reduces the burden of authentication protocols, such as keyed hashed functions (see e.g Katz and Lindell [63], Menezes, Oorschot, and Vanstone [78]), at the network and higher layers.

In particular, we are interested in analyzing channel authentication performance in the framework of wireless communication systems. Under the assumption that there is correlation between the channels, physical layer authentication can be conveniently recast into a hypothesis testing problem (as in Lai, El Gamal, and Poor [69], Maurer [76]). Consider the scheme of Fig. 17: Whenever he receives a message, the receiver (Bob) has to decide between hypothesis $\mathcal{H}_0$ that the message was effectively sent by the legitimate source (Alice), and hypothesis $\mathcal{H}_1$ that it was forged by the attacker (Eve).

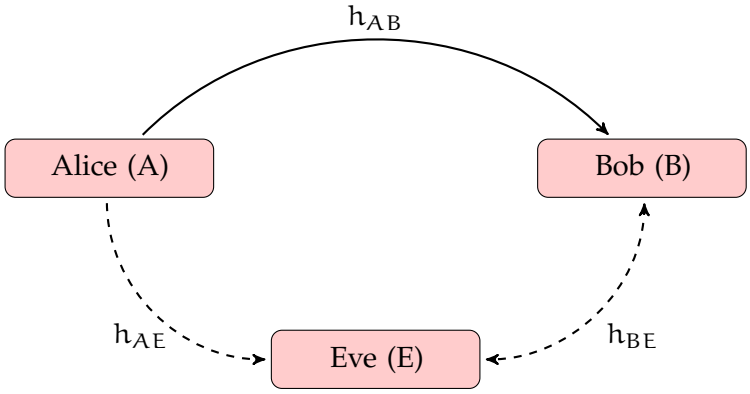


Figure 17: Sketch of the wireless channel authentication model

Physical layer authentication has been addressed by considering either device-specific non-ideal transmission parameters extracted from the received signal (Daniels, Mina, and Russell [30]), or channel characteristics in order to identify the link between a specific source and the receiver (Baracca, Laurenti, and Tomasin [2], Faria and Cheriton [38], Xiao et al. [103]). This chapter focuses on the latter case, which finds application in many wide-band wireless systems, where even small changes in the position of the transmitter have a significant impact on the channel, due to small scale fading effect (see Section 5.2). In particular, we consider the approach proposed in Baracca, Laurenti, and Tomasin [2], where the test is performed in two phases. In the first phase, the receiver gets an authenticated noisy estimate $\mathbf{x}$ of the channel with respect to the legitimate transmitter, h_{AB}. In the second phase, upon reception of a message, the receiver gets a new estimate $\mathbf{u}$ of the channel with respect to the source of the received

message and compares it with $\mathbf{x}$. Then, he must decide whether $\mathbf{u}$ is an estimate of the legitimate channel or the channel forged by an eavesdropper.

The performance of a binary hypothesis testing scheme is measured by the probability of type I (false alarm), and type II (missed detection) errors. Therefore, theoretical bounds on the achievable error probability region are of great importance to establish the effectiveness of practical schemes.

For instance, Maurer considered the traditional authentication scenario in which the legitimate parties can make use of a shared cryptographic key that is kept perfectly secret to the attackers. There, an outer bound on the achievable error region was derived, that holds irrespectively of the decision rule implemented by the receiver. Then, by fixing the false alarm probability, the outer bound is turned into a lower bound on the missed detection probability (see Maurer [76]). An analogous approach was used in Cachin [22] and Barni and Tondi [3] within the different contexts of steganography and fingerprinting, respectively. Similarly, in Lai, El Gamal, and Poor [69], such lower bound is paired with an asymptotic upper bound, and both are derived also in the case that the legitimate parties share correlated sequences, instead of an identical key.

In the above cases, since the attacker has no information on the shared sequences, the optimal attack strategy with respect to the outer bound is to present forged signals that, albeit independent of the legitimate shared key, are generated from the same marginal distribution as the legitimate signals. In our framework, on the contrary, the legitimate authentication signal is the actual realization of a fading wireless channel. In particular, the estimates $\mathbf{z}$ performed by the attacker provide Eve with some side information on the legitimate channel, because the channels h_{AE} and h_{BE} are in general correlated with h_{AB}. Therefore, our main contribution is threefold:

1. We derive an outer bound to the error probability region, in terms of the attacker strategy;

2. we prove the existence of a strategy $\mathbf{v}$, jointly Gaussian with $\mathbf{z}$, that yields the tightest bound, and characterize the joint covariance matrix through the solution of a system of two matrix equations;

3. we give an efficient technique for the numerical evaluation of the optimal attack strategy and the corresponding bound.

This chapter is outlined as follows: First, Section 5.2 provides some preliminaries. Then, Section 5.3 introduces the problem formally, so that the theoretical results can be derived in Section 5.4. Based on those results, in Section 5.5 we propose an efficient algorithm for the numerical evaluation of the optimal attack strategy. Next, in Section

5.6 we give examples of numerical results, and eventually we draw conclusions in Section 5.7.

5.2 PRELIMINARIES

Next we introduce the wireless channel model we are going to consider in the following sections. The wireless channel is affected by fading. While *large scale* fading represents the average signal power attenuation which occurs as the mobile moves over large areas, due to distance and shadowing, *small scale* fading is a consequence of multi-path propagation. Indeed, in wireless communication systems many objects in the propagation environment scatter the signal, so transmission takes place over multiple reflective paths. As a consequence interference occurs, and also small changes in spatial separation can give rise to dramatic oscillations in the received signal's amplitude and phase. Rayleigh fading provides a statistical model for these small scale effects. Starting from the continuous-time multi-path fading channel model, it is possible to derive a discrete-time baseband model which is described in terms of channel filter taps (see e.g. Tse and Viswanath [99]). Then, under the assumption that there is a large number of statistically independent reflected and scattered paths with random amplitudes in the delay window corresponding to a single tap, the Central Limit Theorem allows to conclude that the tap gains are circularly symmetric complex Gaussian distributed. As a consequence, the magnitude of the each tap can be modeled as a Rayleigh random variable. Rayleigh fading is the usual choice in modeling wireless transmissions, including those using orthogonal frequency division multiplexing (OFDM) or multi-antenna (multiple-input multiple-output, MIMO). In MIMO systems each agent may have an arbitrary number of antennas. Thus, the channel vectors are assumed to have different size, in general. All transmissions are corrupted by additive white Gaussian noise with zero mean. Therefore, we model channel estimates performed by each agent in Fig. 17 as zero-mean circular symmetric complex Gaussian vectors with correlated entries (see Section A.4).

In our notation, if $\mathbf{a} \in \mathcal{C}^n$ and $\mathbf{b} \in \mathcal{C}^m$ are (zero-mean) random vectors, K_{ab} denotes the $n \times m$ covariance matrix of vectors $\mathbf{a}$ and $\mathbf{b}$, i.e.

$$K_{ab} := \mathbb{E}\left[\mathbf{a}\mathbf{b}^*\right],$$

whereas $K_{\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]}$ stands for the $(n+m) \times (n+m)$ variance matrix

$$\mathbb{E}\left[\left[\begin{smallmatrix} \mathbf{a} \\ \mathbf{b} \end{smallmatrix}\right]\left[\begin{smallmatrix} \mathbf{a} \\ \mathbf{b} \end{smallmatrix}\right]^*\right].$$

The symbol A^* denotes the complex conjugate transpose of matrix A.

5.3 PROBLEM STATEMENT

5.3.1 Authentication Procedure

The authentication is performed via a two phase procedure, as detailed in Baracca, Laurenti, and Tomasin [2]:

FIRST PHASE In the first phase Alice transmits one or more messages, whose authenticity is guaranteed by higher layer techniques. Bob gets a noisy estimate $\mathbf{x}$ of the channel with respect to Alice (h_{AB}) and replies with acknowledge messages. Moreover, upon Alice and Bob transmissions, Eve gets (possibly noisy) estimates of her own channel with respect to both the other agents (h_{AE} and h_{BE}): the set of these two estimates is denoted with $\mathbf{z}$.

SECOND PHASE In the second phase, either Alice or Eve transmits messages. Bob authenticates the received messages by getting a new noisy channel estimate $\mathbf{u}$ and comparing it with his template $\mathbf{x}$. If this decision process $\mathfrak{D}$ deems the message as coming from Alice, the binary flag $\hat{b}$ is set to zero, otherwise it is set to one. In this phase, Alice performs transmissions in the same fashion of the first phase, while Eve performs a pre-processing of her own messages in order to induce an equivalent channel estimate by Bob that is as close as possible to $\mathbf{x}$.

This physical layer authentication scheme is shown in Fig 18.

In view of our assumptions, the channel estimates are complex, circular symmetric Gaussian random vectors. In particular, $\mathbf{x}$ and $\mathbf{y}$ are n-dimensional, while $\mathbf{z}$ is a m-dimensional. As for $\mathbf{v}$, i.e. the channel forged by the attacker, it is an n-dimensional, complex, random vector whose probability density is not specified as it will be chosen by the attacker in order to obtain better mimetic features.

An abstract representation of the authentication scenario in terms of an hypothesis testing problem is given in Fig. 19. We assume that Eve is able to forge any equivalent channel $\mathbf{v}$ to Bob, through a probabilistic strategy, based on her observations $\mathbf{z}$, which can be characterized by the conditional distribution $p_{v|z}$. Although constraints on power and channel characteristics may in practice prevent this, the assumption is a worst case scenario, which is of interest to derive performance bounds. On the other hand, estimates of Alice-Bob channel in both the first and the second phase are not identical, in general, due to independent noise that affects both estimates. Let $\mathbf{y}$ denote

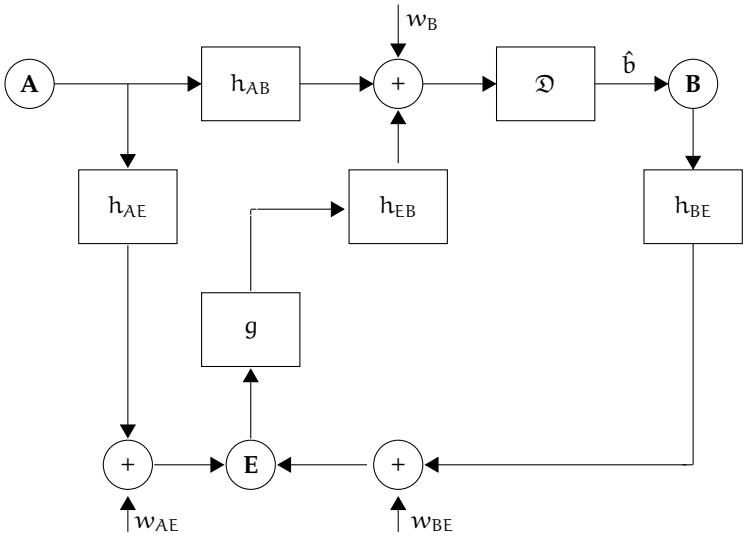


Figure 18: Physical layer authentication scheme. Agents collect noisy channels estimates. $\mathfrak{D}$ denotes the decision process, while g represents the preprocessing performed by the attacker in order to deceive the receiver.

a realization of the random channel estimate of the Alice-Bob channel. Given the channel estimate $\mathbf{u}$, Bob decides between the two hypotheses

$$\mathcal{H}_0 : \quad \mathbf{u} = \mathbf{y} \quad \text{message is from Alice,} \tag{5.1}$$

$$\mathcal{H}_1 : \quad \mathbf{u} = \mathbf{v} \quad \text{message was forged.} \tag{5.2}$$

In Fig. 19, being in hypothesis $\mathcal{H}_0$ or $\mathcal{H}_1$ is obtained by setting $b = 0$ or 1, respectively. Correct authentication is achieved when $\hat{b} = b$.

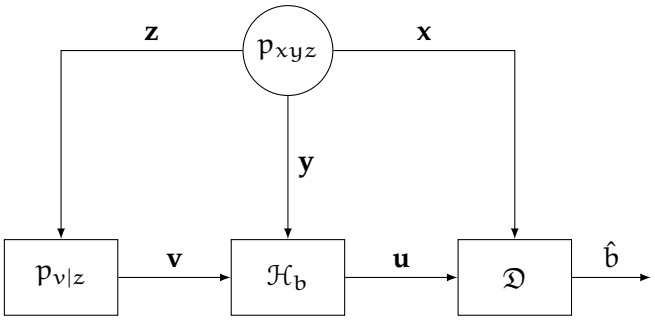


Figure 19: Abstract model for physical layer authentication recast as an hypothesis testing problem.

We denote the set of all possible conditional distributions (forging strategies) $p_{v|z}(\cdot|\cdot)$ as

$$\mathcal{Q} = \left\{ q(\cdot|\cdot) \,:\, \mathcal{C}^n \times \mathcal{C}^m \to \mathcal{R}, q(b|c) \geqslant 0, \int q(b|c)\, db = 1 \right\} . \tag{5.3}$$

Performance of the authentication system are assessed by type I error probability α, i.e., the probability that Bob discards a message as forged by Eve while it is coming from Alice

$$\alpha = P[\hat{b} = 1|\mathcal{H}_0] , \tag{5.4}$$

and the type II error probability β, i.e., the probability that Bob accepts a message coming from Eve as legitimate

$$\beta = P[\hat{b} = 0|\mathcal{H}_1] . \tag{5.5}$$

The aim of a clever design for the authentication scheme is to make both error probabilities α and β as small as possible. Since it is trivial to achieve $\alpha + \beta = 1$ with a random decision strategy that outputs $\hat{b} = 1$ with probability α, independently of the observation $\mathbf{u}$, we are only interested in values of α, β in the region

$$\mathcal{R}^0 = \{(\alpha, \beta) \,:\, \alpha \geqslant 0, \beta \geqslant 0, \alpha + \beta \leqslant 1\} . \tag{5.6}$$

5.3.2 Error Region Bounds for a Given Attacking Strategy

A first bound on the error region for a given attacking strategy can be obtained by applying the fundamental data processing inequality for the Kullback-Leibler (KL) divergence Kullback [68] to our binary hypothesis decision scheme $\mathfrak{D}$. In fact, from Cachin [22], Maurer [76] we have[1]

$$\mathcal{D}\left(p_{\hat{b}|\mathcal{H}_1} \,\|\, p_{\hat{b}|\mathcal{H}_0}\right) \leqslant \mathcal{D}\left(p_{xu|\mathcal{H}_1} \,\|\, p_{xu|\mathcal{H}_0}\right) . \tag{5.7}$$

First we observe that $p_{\hat{b}|\mathcal{H}_0}(1) = \alpha$, $p_{\hat{b}|\mathcal{H}_0}(0) = 1 - \alpha$, and similarly $p_{\hat{b}|\mathcal{H}_1}(0) = \beta$, $p_{\hat{b}|\mathcal{H}_1}(1) = 1 - \beta$. Therefore, introducing the function[2]

$$f(\varphi, \psi) = \varphi \log \frac{\varphi}{1-\psi} + (1-\varphi) \log \frac{1-\varphi}{\psi} , \quad \varphi, \psi \in [0, 1] \tag{5.8}$$

we can rewrite (5.7) as

$$f(\beta, \alpha) \leqslant \mathcal{D}\left(p_{xu|\mathcal{H}_1} \,\|\, p_{xu|\mathcal{H}_0}\right) . \tag{5.9}$$

1 Note that the symmetric bound $\mathcal{D}\left(p_{\hat{b}|\mathcal{H}_0} \,\|\, p_{\hat{b}|\mathcal{H}_1}\right) \leqslant \mathcal{D}\left(p_{xu|\mathcal{H}_0} \,\|\, p_{xu|\mathcal{H}_1}\right)$ also holds true (see also Baracca, Laurenti, and Tomasin [2]).

2 Notice that $f(\varphi, \psi)$ is the KL divergence between two Bernoulli probability distributions of parameters φ and $1 - \psi$, respectively.

Since the observation $\mathbf{z}$ encloses all the information the attacker can exploit in order to deceive the receiver, we can assume that the forging strategy $\mathbf{v}$ is *conditional independent* of the secure template $\mathbf{x}$, given $\mathbf{z}$. Then the divergence on the right side of (5.9) can be written explicitly for a given attacking strategy $q(\cdot|\cdot) \in \mathcal{Q}$ as

$$\begin{aligned} D(q) &= \mathcal{D}\left(p_{xu|\mathcal{H}_1} \| p_{xu|\mathcal{H}_0}\right) \\ &= \mathcal{D}\left(p_{xv} \| p_{xy}\right) \\ &= \iint \left\{ \left[\int p_{xz}(a,c) q(b|c)\, dc \right] \right. \\ &\quad \left. \times \left[\log\left(\int p_{xz}(a,c) q(b|c)\, dc \right) - \log p_{xy}(a,b) \right] \right\} da\, db\,. \end{aligned} \tag{5.10}$$

Let $f_0 \geqslant 0$ be given and set

$$\mathcal{R}(f_0) := \left\{ (\alpha,\beta) \in \mathcal{R}^0 \,:\, f(\beta,\alpha) \leqslant f_0 \right\}. \tag{5.11}$$

Then (5.9) can be rewritten as

$$(\alpha,\beta) \in \mathcal{R}(D(q))\ . \tag{5.12}$$

5.3.3 Error Region Bounds for Any Attacking Strategy

Each outer bound in (5.12) is clearly looser than

$$\mathcal{R}_\cap \quad = \quad \bigcap_{q \in \mathcal{Q}} \mathcal{R}(D(q)) = \mathcal{R}(D^\star) \tag{5.13}$$

where

$$D^\star = \inf_{q \in Q} D(q)\,. \tag{5.14}$$

Note that the region in (5.13) is not strictly speaking an outer bound of the type (5.12), since the infimum (5.14) may, in general, not be achievable. In that case, (5.13) represents a worst case performance for the authentication system, over all possible attacking strategies. On the other hand, for the attacker, approaching (5.14) represents the possibility to effectively carry out an impersonation attack.

Our main goal is to evaluate the tightest bound (5.13). Indeed, we provide an attacking strategy achieving (5.14), under the assumption that the observation $\mathbf{z}$ encodes all the information about the secure template $\mathbf{x}$ the attacker can rely on in order to deceive the receiver. We have just shown that this is equivalent to the following constrained optimization problem:

Problem 5.3.1. Given the zero-mean, circular symmetric, jointly Gaussian random vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ with joint covariance matrix

$$K_{\left[\begin{smallmatrix}x\\y\\z\end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xy} & K_{xz} \\ K_{yx} & K_{yy} & K_{yz} \\ K_{zx} & K_{zy} & K_{zz} \end{bmatrix}, \tag{5.15}$$

find a joint probability distribution $p_{xvz} \in L^1(\mathbb{C}^{2n+m})$ such that its marginal p_{xv} minimizes $\mathcal{D}(p_{xv} \| p_{xy})$ under the constraints:
1. The marginal distribution of $\mathbf{x}, \mathbf{z}$ (corresponding to p_{xvz}) is equal to the given distribution p_{xz}.
2. The random vectors $\mathbf{v}$ and $\mathbf{x}$ are conditionally independent given $\mathbf{z}$.

5.4 MAIN RESULTS

In this section, we address Problem 5.3.1. In particular, we show that the problem is feasible, that it admits an optimal solution and that this solution is Gaussian. Finally, we show how to reformulate this problem in terms of the solution of two coupled matrix equations. The first issue to be considered is the *feasibility* of Problem 5.3.1, namely the existence of a distribution p_{xvz} satisfying the constraints and such that $\mathcal{D}(p_{xv} \| p_{xy})$ is finite.

Lemma 5.4.1. *Problem 5.3.1 is feasible.*

Proof. Let $\mathbf{v}$ be an n-dimensional, complex, zero-mean, circular symmetric Gaussian random vector (with arbitrary covariance) independent of $\mathbf{x}$ and of $\mathbf{z}$. It is immediate to check that the corresponding p_{xvz} satisfies the constraints and is such that $\mathcal{D}(p_{xv} \| p_{xy})$ is finite. ∎

Lemma 5.4.2. *Let $\mathbf{x}$ and $\mathbf{z}$ be jointly Gaussian. For any attacking strategy p_{xv} having finite second moment and in which $\mathbf{v}$ and $\mathbf{x}$ are conditionally independent given $\mathbf{z}$, they are also conditionally orthogonal given $\mathbf{z}$, that is*

$$\mathbb{E}\left[(\mathbf{x} - \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}])(\mathbf{v} - \hat{\mathbb{E}}[\mathbf{v}|\mathbf{z}])\right] = 0, \tag{5.16}$$

where $\hat{\mathbb{E}}[\cdot|\mathbf{z}]$ stands for the best linear estimator of $\cdot$ given $\mathbf{z}$

Proof. We have

$$\begin{aligned} \mathbb{E}\left[(\mathbf{x} - \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}])(\mathbf{v} - \hat{\mathbb{E}}[\mathbf{v}|\mathbf{z}])\right] &= \mathbb{E}\left[\mathbb{E}\left[(\mathbf{x} - \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}])(\mathbf{v} - \hat{\mathbb{E}}[\mathbf{v}|\mathbf{z}])|\mathbf{z}\right]\right] && (5.17) \\ &= \mathbb{E}\left[\mathbb{E}\left[(\mathbf{x} - \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}])|\mathbf{z}\right]\mathbb{E}\left[(\mathbf{v} - \hat{\mathbb{E}}[\mathbf{v}|\mathbf{z}])|\mathbf{z}\right]\right] && (5.18) \\ &= \mathbb{E}\left[\left(\mathbb{E}[\mathbf{x}|\mathbf{z}] - \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}]\right)\left(\mathbb{E}[\mathbf{v}|\mathbf{z}] - \hat{\mathbb{E}}[\mathbf{v}|\mathbf{z}]\right)\right], && (5.19) \end{aligned}$$

where (5.17) and (5.18) follow from the Total Expectation Theorem and the definition of conditional independence, respectively. Since $\mathbf{x}$ and $\mathbf{z}$ are jointly Gaussian, we have that $\mathbb{E}[\mathbf{x}|\mathbf{z}] = \hat{\mathbb{E}}[\mathbf{x}|\mathbf{z}]$. Thus, we can conclude that the right-hand side of (5.19) is equal to zero. ■

In general, conditional independence does not imply conditional orthogonality, although for jointly Gaussian variables they are equivalent. However, we have proved that conditional independence of $\mathbf{x}$ and $\mathbf{v}$ given $\mathbf{z}$ implies that $\mathbf{x}$ and $\mathbf{v}$ are conditionally orthogonal given $\mathbf{z}$, thanks to $\mathbf{x}$ and $\mathbf{z}$ being jointly Gaussian.

Consider now the joint covariance matrix

$$K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}} := \begin{bmatrix} K_{xx} & K_{xv} & K_{xz} \\ K_{vx} & K_{vv} & K_{vz} \\ K_{zx} & K_{zv} & K_{zz} \end{bmatrix}. \tag{5.20}$$

Notice that, since the attacker knows the joint probability density p_{xyz}, the corner elements of (5.20) are known. For the sake of simplicity, we introduce the following symbols for the unknown blocks of (5.20):

$$X := K_{vv}, \quad Y := K_{xv}, \quad Z := K_{vz}. \tag{5.21}$$

Hence, we can write

$$K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}} = \begin{bmatrix} K_{xx} & Y & K_{xz} \\ Y^* & X & Z \\ K_{xz}^* & Z^* & K_{zz} \end{bmatrix}. \tag{5.22}$$

Recall that the conditional orthogonality of $\mathbf{x}$ and $\mathbf{v}$ given $\mathbf{z}$ is equivalent to the following zero-block pattern in its inverse[3]

$$K^{-1}_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}} = \begin{bmatrix} * & 0 & * \\ 0 & * & * \\ * & * & * \end{bmatrix}. \tag{5.23}$$

In this way we have expressed the second constraint of Problem 5.3.1 in terms of the structure of the inverse of the covariance matrix. Thus, a generalized moment problem arises: Notice that it is possible to enforce the zero pattern in the inverse by resorting to a "maximum entropy" completion as described e.g. in Dempster [32]. See also Ferrante and Pavon [39], Pavon and Ferrante [84] for more general results.

Lemma 5.4.3. *If q_G is a circular symmetric Gaussian distribution, then, among all distributions p that share the same mean vector μ and covariance matrix K, the one that minimizes $\mathcal{D}(p \,\|\, q_G)$ is circular symmetric and Gaussian.*

3 A proof can be worked out in the same vein of Speed and Kiiveri [96, Section 2].

Proof. Let p_G be a *circular symmetric Gaussian* probability density on $\mathbb{C}^n$ and let $p \neq p_G$ be any density having the same first and second moment as p_G. We denote by $h(p)$ the differential entropy of the density p, i.e. $h(p) := -\int p(a) \log p(a)\, da$. Then (see Neeser and Massey [81, Theorem 2]), we have the inequality

$$h(p) < h(p_G). \tag{5.24}$$

Now let q_G be any proper Gaussian density on $\mathbb{C}^n$. Under the same hypotheses, we have

$$\int \log q_G(x) p(x) dx = \int \log q_G(x) p_G(x) dx, \tag{5.25}$$

because $\log q_G(x)$ is a quadratic function of $\mathbf{x}$. In view of (5.24) and (5.25), we now have

$$\begin{aligned}
\mathcal{D}(p \,\|\, q_G) &= \int \log \frac{p(x)}{q_G(x)} p(x) dx \\
&= -h(p) - \int \log q_G(x) p(x) dx \\
&= -h(p) - \int \log q_G(x) p_G(x) dx \\
&\geqslant -h(p_G) - \int \log q_G(x) p_G(x) = \mathcal{D}(p_G \,\|\, q_G),
\end{aligned}$$

with equality iff p_G is circular symmetric and Gaussian. Thus, if p is the solution of any minimum entropy problem with circular symmetric Gaussian prior, p has to be circular symmetric and Gaussian. ∎

Lemma 5.4.4. *If the second moment of p_{XV} is not finite then $\mathcal{D}(p_{XV} \,\|\, p_{XY}) = \infty$.*

Proof. We assume that $\mathcal{D}(p_{XV} \,\|\, p_{XY})$ is finite and show that the second moment of p_{XV} is finite. Let us first recall the variational formula for the relative entropy Deuschel and Stroock [33, page 68]:

$$\mathcal{D}(p_{XV} \,\|\, p_{XY}) = \sup_{\varphi \in \Phi} \left\{ \int_{\mathbb{C}^{2n}} \varphi(a) p_{XV}(a) da - \log \left[\int_{\mathbb{C}^{2n}} \exp[\varphi(a)] p_{XY}(a) da \right] \right\} \tag{5.26}$$

where Φ is the set of bounded functions. Observe now that, since p_{XY} is a Gaussian probability density, there exists $\varepsilon > 0$ such that

$$L := \mathbb{E}_{p_{XY}}[\exp[\varepsilon \|a\|^2]] = \int_{\mathbb{C}^{2n}} \exp[\varepsilon \|a\|^2] p_{XY}(a) da$$

is finite. Let us now consider the following sequence of bounded functions:

$$\varphi_l(a) := \begin{cases} \varepsilon \|a\|^2, & \text{if } \|a\|^2 \leqslant l, \\ 0, & \text{if } \|a\|^2 > l. \end{cases}$$

From (5.26) we get that for all $l = 1, 2, \ldots,$

$$\mathcal{D}\left(p_{xv} \| p_{xy}\right) + \log\left[\int_{\mathbb{C}^{2n}} \exp[\varphi_l(a)] p_{xy}(a)\, da\right] \geqslant \int_{\mathbb{C}^{2n}} \varphi_l(a) p_{xv}(a)\, da, \tag{5.27}$$

or, equivalently,

$$\frac{1}{\varepsilon}\left\{\mathcal{D}\left(p_{xv} \| p_{xy}\right) + \log\left[\int_{\mathbb{C}^{2n}} \exp[\varphi_l(a)] p_{xy}(a)\, da\right]\right\} \geqslant \int_{\Omega_l} \|a\|^2 p_{xv}(a)\, da, \tag{5.28}$$

where $\Omega_l := \{a \in \mathbb{C}^{2n} : \ \|a\|^2 \leqslant l\}$. As $l \to \infty$, the left-hand side of (5.28) converges to $\frac{1}{\varepsilon}[\mathcal{D}\left(p_{xv} \| p_{xy}\right) + L]$ while the right hand side converges to the trace of the second moment of p_{xv}. Such a trace is therefore finite and thus also the second moment of p_{xv} is finite. ■

We are now ready to consider the *existence* problem. As in many optimization problems this is one of the most delicate issue.

Theorem 5.4.1. *There exists an optimal solution $p^\star_{xv}$ of Problem 5.3.1.*

Proof. Let $d^\star$ be the infimum of $\mathcal{D}\left(p_{xv} \| p_{xy}\right)$ over p_{xv}, satisfying the constraints of Problem 5.3.1. Let p^j_{xvz}, $j = 1, 2, \ldots,$ be a sequence of probability densities satisfying the constraints of Problem 5.3.1 and such that the corresponding marginals p^j_{xv} satisfy

$$\lim_{j\to\infty} \mathcal{D}\left(p^j_{xv} \| p_{xy}\right) = d^\star.$$

In view of Lemma 5.4.4, we can assume that all p^j_{xvz} have finite mean vector μ_j and covariance matrix $\bar{K}_j$. Let m_j and K_j be the mean and covariance of $\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]$, i.e. m_j are the first $2n$ components of μ_j and K_j is the $2n \times 2n$ upper-left block of $\bar{K}_j$. Now notice that, as $j \to \infty$, $\|K_j\|$ and $\|m_j\|$ remain bounded. In fact, in view of Lemma 5.4.3,

$$\mathcal{D}\left(p^j_{xv} \| p_{xy}\right) \geqslant \mathcal{D}\left(p^{Gj}_{xv} \| p_{xy}\right) = \operatorname{Tr}\left[K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} K_j\right] + m_j^* K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} m_j - \ln\left[\frac{\det[K_j]}{\det[K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}]}\right] - 2n, \tag{5.29}$$

where p^{Gj}_{xv} is the Gaussian distribution having mean vector m_j and covariance matrix K_j. It is easy to check that the right-hand side of (5.29) diverges if at least one of $\|K_j\|$ and $\|m_j\|$ does. Hence, both $\|K_j\|$ and $\|m_j\|$ remain bounded. Thus, also μ_j and $\bar{K}_j$ remain bounded. Therefore, there exists a subsequence $p^{j_i}_{xvz}$ such that $\bar{K}_{j_i}$ and μ_{j_i} converge. Let $\bar{K}^\star$ and $\mu^\star$ be their limits and let $K^\star$ and $m^\star$ be the limits of K_{j_i} and m_{j_i}. Notice now that each density of the corresponding sequence $p^{Gj_i}_{xvz}$ satisfies the constraints of Problem 5.3.1. In fact, the marginal p_{xz} does not change and, in view

of (5.23), the second constraint only depends on the variance matrix. Therefore, also the Gaussian distribution $p^{G\star}_{xvz}$, whose mean and variance are $\bar{K}^\star$ and $\mu^\star$, satisfies the constraints of Problem 5.3.1. Let $p^{G\star}_{xv}$ be the corresponding marginal. We have

$$\begin{aligned} d^\star &= \lim_{i\to\infty} \mathcal{D}\left(p^{j_i}_{xv} \| p_{xy}\right) \geqslant \lim_{i\to\infty} \mathcal{D}\left(p^{Gj_i}_{xv} \| p_{xy}\right) \\ &= \lim_{i\to\infty} \operatorname{Tr}[K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} K_{j_i}] + m^*_{j_i} K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} m_{j_i} - \ln \left[\frac{\det[K_{j_i}]}{\det[K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}]}\right] - 2n \\ &= \operatorname{Tr}[K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} K^\star] + (m^\star)^* K^{-1}_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} m^\star - \ln \left[\frac{\det[K^\star]}{\det[K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}]}\right] - 2n \\ &= \mathcal{D}\left(p^{G\star}_{xv} \| p_{xy}\right). \end{aligned} \tag{5.30}$$

Thus $p^{G\star}_{xvz}$ solves Problem 5.3.1.

■

Notice that from (5.30) it is immediate to see that the optimal solution not only exists but is Gaussian distributed with zero mean.

Corollary 5.4.1. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be jointly Gaussian. Then the solution of Problem 5.3.1 is zero mean and Gaussian.*

We are now ready to find the solution of our problem.

Theorem 5.4.2. *The solution of Problem 5.3.1 is the zero mean circular symmetric Gaussian density* $p^\star_{xvz}$ *whose covariance matrix is*

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}(Z, C) = \begin{bmatrix} K_{xx} & K_{xz}K^{-1}_{zz}Z^* & K_{xz} \\ ZK^{-1}_{zz}K^*_{xz} & ZK^{-1}_{zz}Z^* + CC^* & Z \\ K^*_{xz} & Z^* & K_{zz} \end{bmatrix}, \tag{5.31}$$

where Z *and* C *solve*

$$\begin{cases} C^* = C^*(ZK^{-1}_{zz}BK^{-1}_{zz}Z^* + CC^*)^{-1}A \\ Z^* = K_{zx}K^{-1}_{xx}K_{xy} + BK^{-1}_{zz}Z^*(ZK^{-1}_{zz}BK^{-1}_{zz}Z^* + CC^*)^{-1}A \end{cases} \tag{5.32}$$

with

$$A := K_{yy} - K^*_{xy}K^{-1}_{xx}K_{xy}, \tag{5.33}$$

$$B := K_{zz} - K^*_{xz}K^{-1}_{xx}K_{xz}. \tag{5.34}$$

Proof. We have already shown that the optimal solution is a zero-mean Gaussian distribution having covariance matrix of the form

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]} = \begin{bmatrix} K_{xx} & Y & K_{xz} \\ Y^* & X & Z \\ K_{xz}^* & Z^* & K_{zz} \end{bmatrix}, \tag{5.35}$$

where

$$K_{\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xz} \\ K_{xz}^* & K_{zz} \end{bmatrix} > 0$$

is given. Clearly in this way the first constraint of Problem 5.3.1 is automatically satisfied for any X, Y, Z. We now show that the second constraint is equivalent to impose

$$Y = K_{xz} K_{zz}^{-1} Z^*.$$

Indeed, in view of Lemma 5.4.2, $\mathbf{x}$ and $\mathbf{v}$ are conditional orthogonal given $\mathbf{z}$, so that the inverse of K_{xvz} must exhibit the zero-block pattern (5.23). Based on this information, we can compute Y as a function of Z and X by employing the block-matrix inversion formula:

$$M_1 = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \Rightarrow M_1^{-1} = \begin{bmatrix} (A_1 - B_1 D_1^{-1} C_1)^{-1} & -A_1^{-1} B_1 (D_1 - C_1 A_1^{-1} B_1)^{-1} \\ -D_1^{-1} C_1 (A_1 - B_1 D_1^{-1} C_1)^{-1} & (D_1 - C_1 A_1^{-1} B_1)^{-1} \end{bmatrix}. \tag{5.36}$$

We partition $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ as

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix}, \tag{5.37}$$

where

$$A_1 := K_{xx}, \quad B_1 := \begin{bmatrix} Y & K_{xz} \end{bmatrix}, \quad C_1 := \begin{bmatrix} Y^* \\ K_{xz}^* \end{bmatrix}, \quad D_1 := \begin{bmatrix} X & Z \\ Z^* & K_{zz} \end{bmatrix}.$$

Therefore, the block in position $(1,2)$ of K_{xvz}^{-1} (with respect to the partition (5.37)) is given by

$$\begin{aligned} -A_1^{-1} B_1 (D_1 - C_1 A_1^{-1} B_1)^{-1} &= -{K_{xx}}^{-1} \begin{bmatrix} Y & K_{xz} \end{bmatrix} \left(\begin{bmatrix} X & Z \\ Z^* & K_{zz} \end{bmatrix} - \begin{bmatrix} Y^* \\ K_{zx} \end{bmatrix} {K_{xx}}^{-1} \begin{bmatrix} Y & K_{xz} \end{bmatrix} \right)^{-1} \\ &= -{K_{xx}}^{-1} \begin{bmatrix} Y & K_{xz} \end{bmatrix} \left(\underbrace{\begin{bmatrix} X - Y^* {K_{xx}}^{-1} Y & Z - Y^* {K_{xx}}^{-1} K_{xz} \\ Z^* - K_{zx} K_{xx}^{-1} Y & K_{zz} - K_{zx} K_{xx}^{-1} K_{xz} \end{bmatrix}}_{:= M_2} \right)^{-1}. \end{aligned}$$

In order to impose the zero-block pattern (5.23) to the inverse, we make the block in position $(1,1)$ in $-A_1^{-1}B_1(D_1 - C_1A_1^{-1}B_1)^{-1}$ vanish. Note that we need to compute explicitely only the elements in the first column block of ${M_2}^{-1}$. Let

$$\begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} := \begin{bmatrix} X - Y^* {K_{xx}}^{-1} Y & Z - Y^* {K_{xx}}^{-1} K_{xz} \\ Z^* - K_{zx} K_{xx}^{-1} Y & K_{zz} - K_{zx} K_{xx}^{-1} K_{xz} \end{bmatrix} = M_2$$

Thus, in view of the matrix inversion lemma, the first column block in M_2^{-1} is given by

$$\begin{bmatrix} (A_2 - B_2 {D_2}^{-1} C_2)^{-1} \\ -{D_2}^{-1} C_2 (A_2 - B_2 D_2^{-1} C_2)^{-1} \end{bmatrix}.$$

Therefore, orthogonality of $\mathbf{x}$ and $\mathbf{v}$ given $\mathbf{z}$ implies

$$\begin{aligned} 0 &= -{K_{xx}}^{-1} \begin{bmatrix} Y & K_{xz} \end{bmatrix} \begin{bmatrix} (A_2 - B_2 {D_2}^{-1} C_2)^{-1} \\ -{D_2}^{-1} C_2 (A_2 - B_2 D_2^{-1} C_2)^{-1} \end{bmatrix} \\ &= -{K_{xx}}^{-1} Y (A_2 - B_2 {D_2}^{-1} C_2)^{-1} + {K_{xx}}^{-1} K_{xz} {D_2}^{-1} C_2 (A_2 - B_2 D_2^{-1} C_2)^{-1} \\ &= Y - K_{xz} {D_2}^{-1} C_2, \end{aligned}$$

so that

$$\begin{aligned} Y &= K_{xz} \left(K_{zz} - K_{zx} K_{xx}^{-1} K_{xz}\right)^{-1} \left(Z^* - K_{zx} K_{xx}^{-1} Y\right) \\ &= \left[\left(I + K_{xz} \left(K_{zz} - K_{zx} K_{xx}^{-1} K_{xz}\right)^{-1} K_{zx} {K_{xx}}^{-1}\right)\right]^{-1} K_{xz} \left(K_{zz} - K_{zx} K_{xx}^{-1} K_{xz}\right)^{-1} Z^* \\ &= K_{xz} K_{zz}^{-1} Z^*. \end{aligned}$$

In this way, we have parametrized all the matrices $K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]}$ whose inverse has the specified structure. At this point, we could minimize the divergence $\mathcal{D}(p_{xv} \| p_{xy})$ over Z and X. This turns out to be an easy problem that can be solved in closed form. This solution, however, is not the solution[4] of our original problem since there is yet another (hidden) constraint that we need to impose. Namely we have to impose that the matrix

$$K_{\left[\begin{smallmatrix} x \\ v \\ z \end{smallmatrix}\right]} = \begin{bmatrix} K_{xx} & K_{xz} K_{zz}^{-1} Z^* & K_{xz} \\ (K_{xz} K_{zz}^{-1} Z^*)^* & X & Z \\ K_{xz}^* & Z^* & K_{zz} \end{bmatrix} \tag{5.38}$$

4 Here we mention this simplified optimization problem because, as discussed later, it turns out to be very useful as the first step of an efficient numerical procedure that computes the solution of our original problem.

is a *bona fide* covariance matrix, i.e. it is positive semidefinite. Since $K_{\left[\begin{smallmatrix} x \\ z \end{smallmatrix}\right]}$ is positive definite, this constraint is equivalent to

$$X - \begin{bmatrix} (K_{xz}K_{zz}^{-1}Z^*)^* & Z \end{bmatrix} \begin{bmatrix} K_{xx} & K_{xz} \\ K_{xz}^* & K_{zz} \end{bmatrix}^{-1} \begin{bmatrix} K_{xz}K_{zz}^{-1}Z^* \\ Z^* \end{bmatrix} \geqslant 0$$

which, with simple algebraic manipulations, is seen to be equivalent to

$$X - ZK_{zz}^{-1}Z^* \geqslant 0. \tag{5.39}$$

The positivity constraint is then automatically satisfied if we re-parametrize the unknown matrix X in term of a new matrix C in the form

$$X = ZK_{zz}^{-1}Z^* + CC^*. \tag{5.40}$$

The optimal solution can be now easily obtained by solving the following *unconstrained* optimization problem

$$\arg\min_{C,Z} \mathcal{D}\left(p_{xv} \,\|\, p_{xy}\right). \tag{5.41}$$

Since

$$K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C) := \begin{bmatrix} K_{xx} & K_{xz}K_{zz}^{-1}Z^* \\ Z(K_{xz}K_{zz}^{-1})^* & ZK_{zz}^{-1}Z^* + CC^* \end{bmatrix}, \quad K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy}^* & K_{yy} \end{bmatrix}, \tag{5.42}$$

solving (5.41) is equivalent to compute

$$\arg\min_{Z,C} \left\{ -\log\det\left(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C) K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}^{-1} \right) + \operatorname{Tr}\left[K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}^{-1} K_{xv} K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C) \right] \right\}. \tag{5.43}$$

We are then led to the formulation of Problem 5.3.1. Let

$$J(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C)) := -\log\det\left(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C) K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}^{-1} \right) + \operatorname{Tr}\left[K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}^{-1} K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z,C) \right]. \tag{5.44}$$

Its first variation is provided by

$$\begin{aligned}
&D[J(K_{xv}(Z,C)); \delta K_{xv}(Z,C)] \\
&= \operatorname{Tr}\left[\left[(-K_{xv}^{-1} + K_{xy}^{-1}) \delta K_{xv} \right] \right] \\
&= \operatorname{Tr}\left[\left[\underbrace{(-K_{xv}^{-1} + K_{xy}^{-1})}_{=:\Delta} \begin{bmatrix} 0 & K_{xz}K_{zz}^{-1}\delta Z^* \\ \delta Z(K_{xz}K_{zz}^{-1})^* & \delta Z K_{zz}^{-1}Z^* + ZK_{zz}^{-1}\delta Z^* + \delta C C^* + C\delta C^* \end{bmatrix} \right] \right] \\
&= \operatorname{Tr}\left[\left[\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \begin{bmatrix} 0 & K_{xz}K_{zz}^{-1}\delta Z^* \\ \delta Z(K_{xz}K_{zz}^{-1})^* & \delta Z K_{zz}^{-1}Z^* + ZK_{zz}^{-1}\delta Z^* + \delta C C^* + C\delta C^* \end{bmatrix} \right] \right] \\
&= \operatorname{Tr}\left[\begin{bmatrix} \Delta_{12}\delta Z(K_{xz}K_{zz}^{-1})^* & * \\ * & \Delta_{21}K_{xz}K_{zz}^{-1}\delta Z^* + \Delta_{22}\left[\delta Z K_{zz}^{-1}Z^* + ZK_{zz}^{-1}\delta Z^* + \delta C C^* + C\delta C^* \right] \end{bmatrix} \right].
\end{aligned}$$

By the properties of the trace and the Hermitian symmetry, we get that the first variation vanishes if and only if

$$\operatorname{Tr}\left[\left[\left((K_{xz}K_{zz}^{-1})^{*}\Delta_{12}+Z^{*}K_{zz}^{-1}\Delta_{22}\right)\delta Z+C^{*}\Delta_{22}\delta C\right]\right]=0. \tag{5.45}$$

This holds for all δZ, δC if and only if

$$\begin{cases}(K_{xz}K_{zz}^{-1})^{*}\Delta_{12}+K_{zz}^{-1}Z^{*}\Delta_{22}=0\\ C^{*}\Delta_{22}=0\end{cases} \tag{5.46}$$

The first equation in (5.46) can be simplified so that it reads

$$K_{xz}\Delta_{12}+Z^{*}\Delta_{22}=0. \tag{5.47}$$

The matrix inversion lemma allows to compute an explicit expression for matrix Δ

$$\begin{aligned}\Delta_{12}&=-K_{xx}^{-1}K_{xy}(K_{yy}-K_{yx}K_{xx}^{-1}K_{xy})^{-1}+\\&\quad K_{xx}^{-1}K_{xz}K_{zz}^{-1}Z^{*}\left[ZK_{zz}^{-1}(K_{zz}-K_{zx}K_{xx}^{-1}K_{xz})K_{zz}^{-1}Z^{*}+CC^{*}\right]^{-1},\\ \Delta_{22}&=\left(K_{yy}-K_{yz}K_{xx}^{-1}K_{xy}\right)^{-1}-\left[ZK_{zz}^{-1}(K_{zz}-K_{zx}K_{xx}^{-1}K_{xz})K_{zz}^{-1}Z^{*}+CC^{*}\right]^{-1}.\end{aligned}$$

Now, let $A:=K_{yy}-K_{yx}K_{xx}^{-1}K_{xy}$, and $B:=K_{zz}-K_{zx}K_{xx}^{-1}K_{xz}$. Then we can write

$$\begin{aligned}\Delta_{12}&=-K_{xx}^{-1}K_{xy}A^{-1}+K_{xx}^{-1}K_{xz}K_{zz}^{-1}Z^{*}\left[ZK_{zz}^{-1}BK_{zz}^{-1}Z^{*}+CC^{*}\right]^{-1},\\ \Delta_{22}&=A^{-1}-\left(ZK_{zz}^{-1}BK_{zz}^{-1}Z^{*}+CC^{*}\right)^{-1}.\end{aligned}$$

Therefore, after some manipulation, we conclude that the optimum solution is provided by C,Z such that

$$\begin{cases}C^{*}=C^{*}(ZK_{zz}^{-1}BK_{zz}^{-1}Z^{*}+CC^{*})^{-1}A\\ Z^{*}=K_{zx}K_{xx}^{-1}K_{xy}+BK_{zz}^{-1}Z^{*}(ZK_{zz}^{-1}BK_{zz}^{-1}Z^{*}+CC^{*})^{-1}A\end{cases}. \tag{5.48}$$

■

In view of (5.10) and (5.13), Theorem 5.4.2 provides the tightest bound to the error region (5.13). Indeed, let $K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}$ be a shorthand notation for the $2n\times 2n$ upper-left corner of (5.31). Then, $D^{\star}$ is given by

$$D^{\star}=\mathcal{D}\left(p_{xv}^{\star}\,\|\,p_{xy}\right)=-\log\det\left(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1}\right)+\operatorname{Tr}\left[K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}^{-1}\left(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}-K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]}\right)\right]. \tag{5.49}$$

Consider the circular symmetric Gaussian density $p^{\star}_{xvz}$, with zero mean and variance

$$K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}} = \begin{bmatrix} K_{xx} & K_{xz}K_{zz}^{-1}Z^{*} & K_{xz} \\ ZK_{zz}^{-1}K_{xz}^{*} & ZK_{zz}^{-1}Z^{*} + CC^{*} & Z \\ K_{xz}^{*} & Z^{*} & K_{zz} \end{bmatrix}. \tag{5.50}$$

Note that it is such that $\mathbf{x}$ and $\mathbf{v}$ are conditionally independent given $\mathbf{z}$. Then, by marginalizing and conditioning, we can obtain an optimum attacking strategy $p^{\star}_{v|z}(\cdot|a)$ which achieves (5.13). It is given by the proper Gaussian density whose mean and variance are defined by

$$\mu_{v|z} := ZK_{zz}^{-1}a \tag{5.51}$$

$$K_{v|z} := K_{vv} - K_{vz}K_{zz}^{-1}K_{vz}^{*} = CC^{*} \tag{5.52}$$

5.5 EFFICIENT COMPUTATION OF THE TIGHTEST BOUND

In view of Theorem 5.4.2, in order to provide the expression of the optimal solution $p^{\star}_{xvz}$, we have to compute matrices C, Z which solve the system of nonlinear matrix equations (5.32). This appears however to be a highly non trivial task. Thus, we propose a two stage algorithm:

1. **Feasible (projected) Solution.** To begin with, we deal with an optimization problem which can be considered a relaxed version of Problem 5.3.1, since no positivity constraints on the matrix $K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}}$ are imposed. This task turns out to be much simpler to achieve. Indeed, the solution can be computed in closed form. Then, we project the solution to the relaxed problem onto the feasible set, i.e. the set of pairs (X, Z) which make $K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}}$ positive definite.

2. **Iterative Algorithm.** We use the projection as a starting point for an iterative update procedure whose fixed point satisfies (5.32).

Next we provide some details for each phase.

Feasible Solution. Minimizing (5.10) with no constraints on the positivity of $K_{\begin{bmatrix} x \\ v \\ z \end{bmatrix}}$ is equivalent to solve

Problem 5.5.1.

$$\arg\min_{X,Z} J(K_{\begin{bmatrix} x \\ v \end{bmatrix}}(Z, X)) := \left\{ -\log\det\left(K_{\begin{bmatrix} x \\ v \end{bmatrix}}(Z, X) K^{-1}_{\begin{bmatrix} x \\ y \end{bmatrix}} \right) + \mathrm{Tr}\left[K^{-1}_{\begin{bmatrix} x \\ y \end{bmatrix}} K_{\begin{bmatrix} x \\ v \end{bmatrix}} \right] \right\} \tag{5.53}$$

where

$$K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}(Z,X) := \begin{bmatrix} K_{xx} & K_{xz}K_{zz}^{-1}Z^* \\ Z(K_{xz}K_{zz}^{-1})^* & X \end{bmatrix}, \quad K_{\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]} := \begin{bmatrix} K_{xx} & K_{xy} \\ K_{xy}^* & K_{yy} \end{bmatrix}. \tag{5.54}$$

In the same vein of the proof of Theorem 5.4.2, based on the analysis of the first variation $D[J(K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}(Z,X);\delta K_{\left[\begin{smallmatrix}x\\v\end{smallmatrix}\right]}]$, we work out the optimality conditions that have to be satisfied by X and Z. Some easy algebraic calculations lead us to the closed form of an optimal solution (Z,X):

$$\begin{cases} Z = K_{xy}^* {K_{xx}}^{-1} K_{xz} (K_{xz}^* K_{xx}^{-1} K_{xz})^\dagger K_{zz}, \\ X = K_{yy} - K_{xy}^* {K_{xx}}^{-\frac{1}{2}} \left[I - K_{xx}^{-\frac{1}{2}} K_{xz} (K_{xz}^* {K_{xx}}^{-1} K_{xz})^\dagger K_{xz}^* K_{xx}^{-\frac{1}{2}} \right] {K_{xx}}^{-\frac{1}{2}} K_{xy} \end{cases} \tag{5.55}$$

where " † " denotes Moore-Penrose pseudo inverse.

If the obtained $\mathbf{x}$ and $\mathbf{z}$ are such that $X - ZK_{zz}^{-1}K^* \geqslant 0$, the algorithm terminates. Otherwise, a pair (C,Z) is obtained as follows. Let T be a unitary matrix such that $\Sigma_T := T^*(X - ZK_{zz}^{-1}K^*)T = \operatorname{diag}(d_1, d_2, \ldots, d_k, \delta_1, \delta_2, \ldots, \delta_h)$, where d_i are positive and in decreasing order, and δ_i are negative or zero. Let $\Sigma'_T := \operatorname{diag}(d_1, d_2, \ldots, d_k, \varepsilon, \varepsilon, \ldots, \varepsilon)$, where $\varepsilon := (d_k/100) > 0$ is a "small" parameter. Let $\Sigma' := T\Sigma'_T T^* > 0$ and C be such that $CC^* = \Sigma'$.

Iterative Algorithm. We use the pair (C,Z) as a starting point for the iterations

$$\begin{cases} C^*(k+1) = C^*(k)(Z(k)K_{zz}^{-1}BK_{zz}^{-1}Z^*(k) + C(k)C^*(k))^{-1}A, \\ Z^*(k+1) = K_{zx}K_{xx}^{-1}K_{xy} + BK_{zz}^{-1}Z^*(k)(Z(k)K_{zz}^{-1}BK_{zz}^{-1}Z^*(k) + C(k)C^*(z))^{-1}A \end{cases} \tag{5.56}$$

where

$$A := K_{yy} - K_{xy}^* K_{xx}^{-1} K_{xy} \tag{5.57}$$

$$B := K_{zz} - K_{xz}^* K_{xx}^{-1} K_{xz}. \tag{5.58}$$

By the the iterative process we aim at finding a fixed point for (5.56), which provides the solution of Problem 5.3.1. The iterative process can be stopped either after a fixed number of iterations, or when the variation of D^* over one iteration is smaller than a given percentage.

5.6 NUMERICAL RESULTS

5.6.1 Uncorrelated Channels

In order to assess the performance of the proposed algorithm for the computation of the tightest bound, we first consider the case where $m = n$ and the covariance matrices are identities, i.e.

$$K_{\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]} = \begin{bmatrix} I_n & \sigma I_n & \rho I_n \\ \sigma^* I_n & I_n & \tau I_n \\ \rho^* I_n & \tau^* I_n & I_n \end{bmatrix}$$

This scenario corresponds for example to an OFDM transmission with uncorrelated channel frequency response. Beyond being an asymptotic case widely considered in the literature, this is also a practical scenario, when a subset of subcarriers with cardinality smaller than the number of channel taps is considered, and the channel taps are independent Gaussian variables. The parameter ρ dictates the correlation between channel estimates performed by Eve and the legitimate channel.

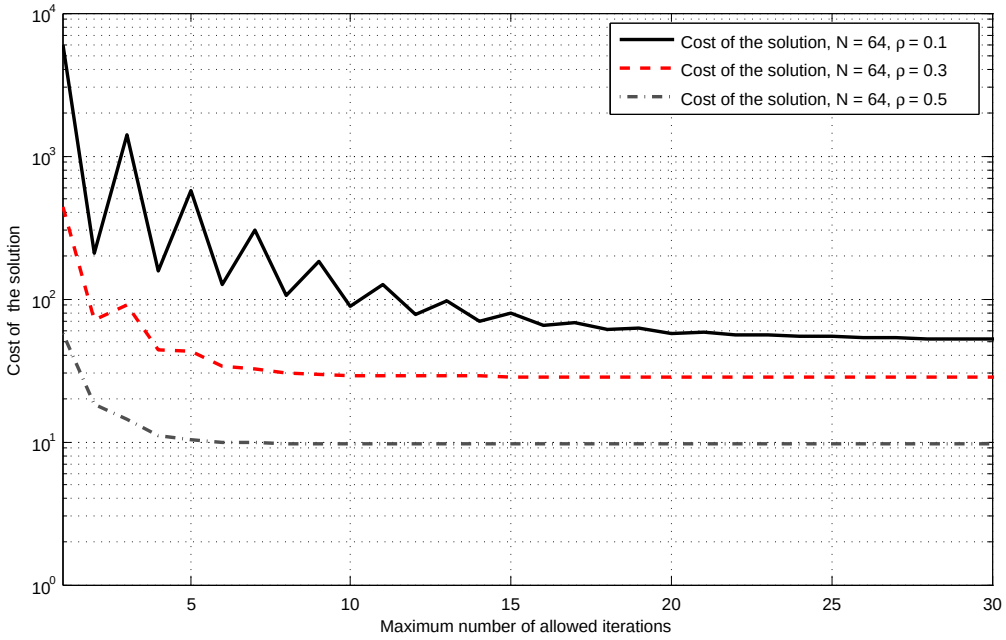


Figure 20: Cost of the solution computed by the iterative algorithm as a function of the maximum number of iterations, with $n = m = 64$, for $\rho = 0.1, 0.3, 0.5$

First we assess the performance of the iterative algorithm. Fig 20 shows the values of the cost of the optimum solution D^* as a function of the number of iterations for the iterative algorithm, with $n = m = 64$, and various values of ρ. We observe

that the iterative algorithm always converges to a fixed point for (5.56) and that the convergence to a solution with good accuracy takes less than 100 iterations. Thus, in the following we consider this value for the maximum number of iterations.

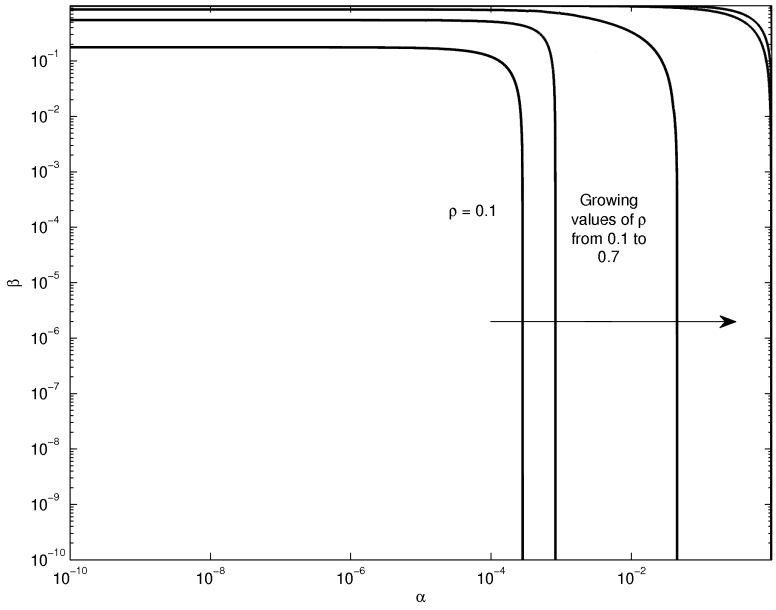


Figure 21: Bound of the region type II (β) vs type I (α) error probability for various values of the correlation parameter ρ, with $K_{xx} = I_{n\times n}$, $K_{zz} = I_{m\times m}$, and $K_{xz} = \rho I_{n\times m}$.

Fig. 21 shows the bound of the type II (β) – type I (α) error probability region for various values of the correlation parameter ρ, and for $n = m = 64$, as obtained from the proposed iterative approach. As expected, we observe that for increasing values of ρ, the region of achievable values of α and β gets wider. In particular, for the considered scenario, the type II error probability is larger than 10^{-1} already for $\rho = 0.4$.

In Fig. 22 we report the results obtained for both the initial feasible solution (projection of the solution of (5.55)) and final solution of the iterative algorithm, as a function of n, for $\rho = 0.1,0.5,0.7$. For the sake of clarity, we also show the cost of the solutions provided by the iterative algorithm in Tab. 1.

Table 1: Cost of the solution provided by the iterative solution.

D^*	$n = 2$	$n = 4$	$n = 8$	$n = 16$	$n = 32$	$n = 64$
$\rho = 0.1$	1.6099	3.2199	6.4397	12.8795	25.7589	51.5179
$\rho = 0.5$	0.3047	0.6094	1.2189	2.4378	4.8756	9.7511
$\rho = 0.7$	0.0005	0.0011	0.0021	0.0042	0.0085	0.0169

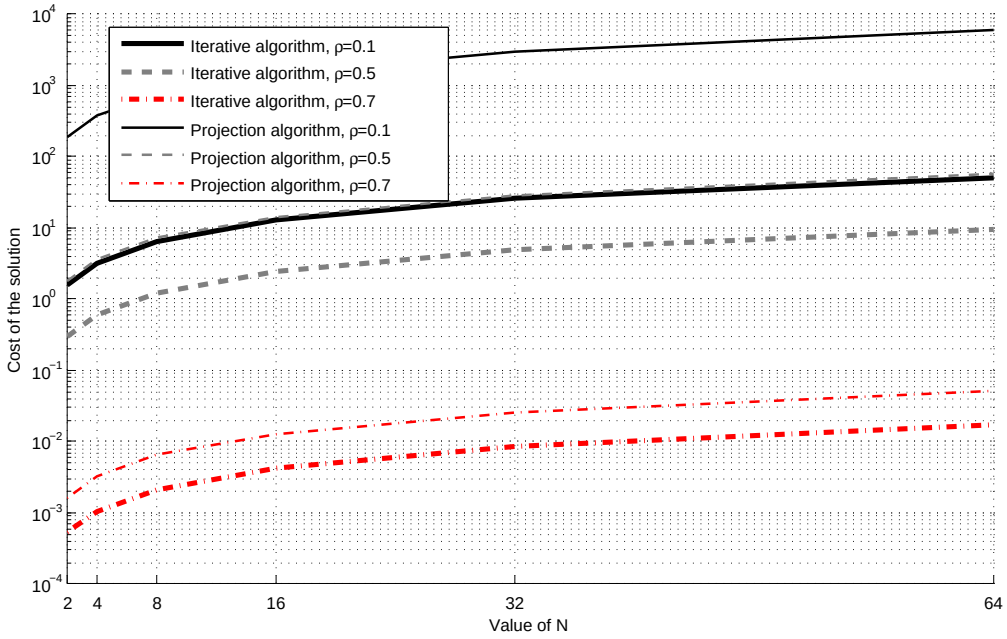


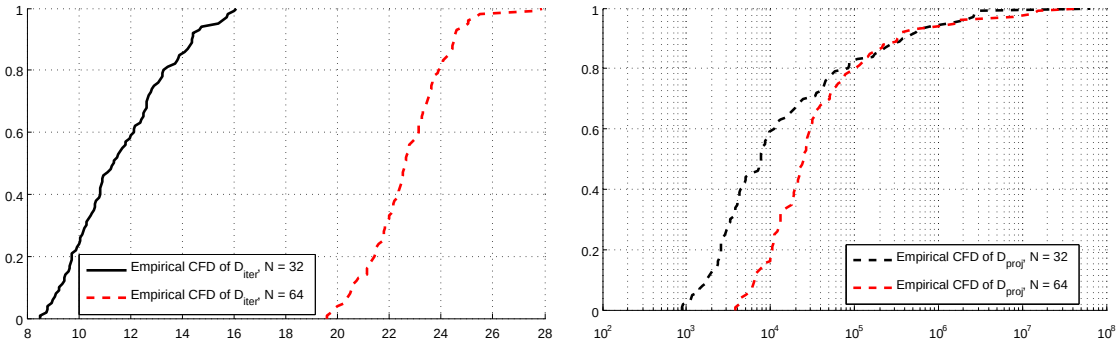
Figure 22: Cost function D^* as a function of n for various values of the correlation parameter ρ, with $K_{xx} = I_{n\times n}$, $K_{zz} = I_{m\times m}$, and $K_{xz} = \rho I_{n\times m}$. Both projection and iterative algorithms are considered.

We note that the iterative algorithm remarkably lowers the value of the cost function from the initial feasible solution, thus motivating its use, although it comes at a cost of more computations. Also, as expected, the cost function increases with n. For the considered case of OFDM transmission, this means that more dispersive channels having independent taps provide potentially a better authentication system. This phenomenon has been already seen in Baracca, Laurenti, and Tomasin [2].

5.6.2 Correlated Channels

We now consider channels with random correlation. We let $m = n$ and generate $K_{\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]}$ as a realization of a $3n \times 3n$ real Wishart matrix[5]. Even in this case we verified that setting the maximum number of iteration to 100 is enough for the convergence of the iterative algorithm. Fig. 23.a shows the cumulative distribution function (CDF) of D^* for two values of $n = m$, at the convergence of the iterative algorithm. Also in this

5 A $n \times n$ real (resp., complex) *Wishart matrix* is a random matrix W that can be written as $W = AA^*$, where A is a $n \times n$ random matrix with independent identically distributed (iid) real (resp., circularly symmetric complex) Gaussian entries. In our case, the entries of A have zero mean and unit variance.



(a) *Iterative algorithm.* **(b)** *Initial feasible solution.*

Figure 23: CDF of the cost function for two values of n.

case we observe that a larger n provides a larger value of D^*. We also report in Fig. 23.b the CDF for the initial feasible solution obtained by projection.

For the random correlation case, Tab. 2 shows the probability that the closed form solution of the relaxed problem (5.55) satisfies the positivity constraint, as a function of n.

Table 2: Probability that (5.55) is feasible, as a function of n.

n	2	4	8	16	32	64
p [%]	43	10	0	0	0	0

Note that as n increases this probability goes fast to zero, thus making the projection step necessary to obtain an initial feasible solution for the iterative algorithm.

In order to compare the iterative solution to the one provided by (5.55), which may not fulfill the positivity constraints on the joint covariance matrix, Fig. 24 shows the percentage increase of the cost (5.44) defined as

$$\eta := 100 \times \left[\frac{J^*_{iter}}{J^*_{cf}} - 1 \right], \tag{5.59}$$

where J^*_{iter} is the cost of the solution provided by the iterative algorithm, whereas J^*_{cf} is the cost of the one computed in closed form through (5.55). The analyzing of the increment with regard to J^* is convenient because D^*_{cf} can vanish. Indeed, recall that, if $K_{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]}$ is a $n \times n$ matrix, it holds that $D^* = J^* - 2n$. We note that the increase is in the range of 20% to 30% for the considered scenario. Moreover, it is diminishing as n increases. This seems to suggest that, for growing values of n, the solution computed by means of (5.55) corresponds to a matrix of the form (5.38) which gets closer to the cone of positive definite matrices of size $(2n + m)$.

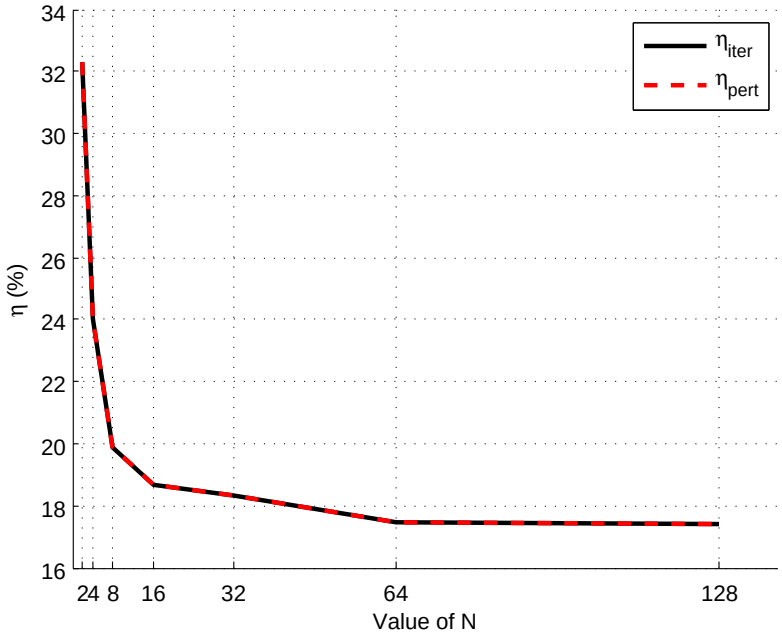


Figure 24: Percentage improvement η as a function of n. Random correlation matrices and $n = m$. Perturbation analysis results are included.

We also provide results for the perturbation analysis. In particular, we evaluate the effects of small perturbations of Z and C generated as Gaussian random variables with norm $0.01\|Z\|$ and $0.01\|C\|$, respectively. Fig. 24 reports the maximum cost function achieved for all perturbed values, showing that it provides negligible improvement with respect to the solution of the iterative approach. This supports the conclusion that the iterative approach reaches a minimum point for $\mathbb{J}(K_{\left[\begin{smallmatrix} x \\ v \end{smallmatrix}\right]}(Z, C))$. We also applied the iterative algorithm starting from the perturbed solutions which led to cost improvements. Results, not reported here, show that this procedure achieves very small improvements with an increase of the cost function of 0.01% .

5.7 CONCLUSIONS

We have considered the problem of deriving a universal performance bound for a message source authentication scheme based on channel estimates in a wireless fading scenario, where an attacker may have correlated observations available and possibly unbounded computational power. We have formulated an outer bound to the region of achievable false alarm and missed detection probabilities, which is universal across all possible decision rules by the receiver.

Under the assumption that the channels are represented by multivariate complex Gaussian variables, we have proved that the tightest bound corresponds to a forging strategy that produces a zero mean signal that is jointly Gaussian with the attacker observations. Furthermore, we have derived a characterization of their joint covariance matrix through the solution of a system of two nonlinear matrix equations. Based upon this characterization, we have also devised an efficient iterative algorithm for its computation: The solution to the matricial system appears as fixed point of the iteration.

From numerical results, we conclude that the proposed iterative approach for the best attacking strategy always converges. Moreover, from the perturbation analysis, we deduce that the limit point is a local minimum. We have therefore provided an effective method for the attacking strategy that yields the tightest bound on the error region of the message authentication procedure.

A APPENDIX

For the sake of completeness, here we collect some preliminary notions in Probability, Information Theory, and Identification which will be useful throughout this dissertation. For more details, the reader is referred to Cover and Thomas [27], Lindquist and Picci [73], Papoulis and Pillai [83], Rozanov [92], Stoica and Moses [97]

A.1 RANDOM VARIABLES AND VECTORS

We first introduce real-valued random variables. Consider the probability space $(\Omega, \mathcal{F}, P)$, where Ω is a set (also known as *sample space*), $\mathcal{F}$ is a σ-algebra defined on Ω and P is a probability measure on $\mathcal{F}$[1]. Given the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra defined on $\mathbb{R}$, the function $\mathbf{y} : \Omega \to \mathbb{R}$ is a (real-valued) random variable if and only if, for each B in $\mathcal{B}(\mathbb{R})$

$$\{\omega : \mathbf{y}(\omega) \in B\} \in \mathcal{F}, \tag{A.1}$$

or, equivalently, if and only if

$$\{\omega : \mathbf{y}(\omega) \leqslant b\} \in \mathcal{F}, \quad \forall\, b \in \mathbb{R}. \tag{A.2}$$

In the following, we will often write $\mathbf{y}$ instead of $\mathbf{y}(\omega)$.

The probability *distribution* of $\mathbf{y}$ is given by the function

$$F_{\mathbf{y}}(b) := P\left[\{\omega : \mathbf{y}(\omega) \leqslant b\}\right], \quad \forall\, b \in \mathbb{R}. \tag{A.3}$$

If it is absolutely continuous, we can introduce the probability *density* function $f_{\mathbf{y}}$ such that

$$F_{\mathbf{y}}(b) = \int_{-\infty}^{b} f_{\mathbf{y}}(x)dx, \quad \forall\, b \in \mathbb{R}. \tag{A.4}$$

1 Recall that $P : \mathcal{F} \to [0, 1]$ is a probability measure if and only if

- $P(\Omega) = 1$;
- $P(\cup_{n=1}^{\infty} F_n) = \sum_{n=1}^{\infty} P(F_n)$ for every sequence $\{F_n\}_{n=1}^{\infty}$ of pairwise disjoint elements.

Then the mean and variance of $\mathbf{y}$ are given by

$$\mathbb{E}[\mathbf{y}] := \int_{\mathbb{R}} x f_{\mathbf{y}}(x) dx, \tag{A.5}$$

and

$$\text{Var}[\mathbf{y}] := \int_{\mathbb{R}} x^2 f_{\mathbf{y}}(x) dx, \tag{A.6}$$

respectively.

Analogously, we can define an $m-$dimensional random vector

$$\mathbf{y} : \Omega \to \mathbb{R}^m \tag{A.7}$$

by considering the same probability space $(\Omega, \mathcal{F}, P)$ and the measurable set $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Along the same line, it is possible to introduce complex-valued random variables and vectors.

Example A.1.1 (Multivariate normal distribution)**.** Let $\mathbf{y}$ be an $m-$dimensional, real-valued Gaussian vector, with mean $\mu = \mathbb{E}[\mathbf{y}]$, and variance $\Sigma := \mathbb{E}\left[\mathbf{y}\mathbf{y}^\top\right]$. Then, its density function is given by

$$f_{\mathbf{y}}(x_1, \ldots, x_m) := \frac{1}{\sqrt{(2\pi)^m \det \Sigma}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right], \tag{A.8}$$

where $x := [x_1 \ \cdots \ x_n]^\top$. Thus, Gaussian random vectors are completely described by their second order properties.

A.2 NOTIONS IN INFORMATION THEORY

Next a few fundamental notions in Information Theory are recalled.

Definition A.2.1 (Entropy rate)**.** Let $\mathbf{y}$ be a real-valued random variable, described by the probability density f. Then, the *entropy rate* of $\mathbf{y}$ is given by

$$h(\mathbf{y}) := -\int_{\mathbb{R}} f(x) \log x dx. \tag{A.9}$$

Example A.2.1 (Gaussian vector)**.** Consider an n-dimensional Gaussian random vector $\mathbf{x}$ of variance Σ. Its entropy rate is

$$h(\mathbf{x}) = \frac{n}{2}(1 + \log 2\pi) + \frac{1}{2}\log\det\Sigma. \tag{A.10}$$

Definition A.2.2 (Relative entropy)**.** Let f, g be two probability density functions defined on the same set S. Then, the *relative entropy* $\mathcal{D}(f \,\|\, g)$ is defined as

$$\mathcal{D}(f \,\|\, g) = \int_{\mathbb{R}} f(x) \log \frac{f(x)}{g(x)} dx. \tag{A.11}$$

Relative entropy is also known as *Kullback-Leibler divergence*. It exhibits the following properties:

- $\mathcal{D}(f \,\|\, g) = 0$ if and only if $f = g$ (a.e.);
- $\mathcal{D}(f \,\|\, g) \geqslant 0$ for all probability density functions f and g.

Note that it is not a proper distance, because it is not symmetric neither it obeys to the triangular inequality.

Example A.2.2 (Relative entropy between Gaussian distributions)**.** Consider the random the variables $\mathbf{x} \sim \mathcal{N}(\mu_x, \Sigma_x)$, $\mathbf{y} \sim \mathcal{N}(\mu_y, \Sigma_y)$. Then, we have

$$\begin{aligned} \mathcal{D}(\mathcal{N}(\mu_x, \Sigma_x) \,\|\, \mathcal{N}(\mu_y, \Sigma_y)) = -\frac{1}{2} \log \frac{\det \Sigma_x}{\det \Sigma_y} + \frac{1}{2} \mathrm{Tr}\left[{\Sigma_y}^{-1} \Sigma_x\right] - \frac{n}{2} + \\ \frac{1}{2} (\mu_x - \mu_y)^\top {\Sigma_y}^{-1} (\mu_x - \mu_y). \end{aligned} \tag{A.12}$$

A.3 COMPLEX GAUSSIAN RANDOM VECTORS

In the following, *complex-Gaussian* random vectors will play an important role. Here we recall some basic notions. Let $\mathbf{x}$ and $\mathbf{y}$ with values in $\mathbb{R}^k$ be jointly Gaussian random vectors, i.e. such that $\forall i, j \in \{1, \ldots, k\}$ the vector $\left[\mathbf{x}_1 \cdots \mathbf{x}_i \, \mathbf{y}_1 \cdots \mathbf{y}_j\right]^\top$ is Gaussian distributed. Then, we say that $\mathbf{z} := \mathbf{x} + j\mathbf{y}$ has the *complex normal distribution*. Its statistical description is given by the following three parameters:

- Mean $\mu := \mathbb{E}[\mathbf{z}]$;
- Covariance matrix $\Gamma := \mathbb{E}\left[(\mathbf{z} - \mu)(\bar{\mathbf{z}} - \bar{\mu})^\top\right]$;
- Relation matrix $C := \mathbb{E}\left[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^\top\right]$.

Here $\bar{\mathbf{z}}$ and denotes the complex conjugate of $\mathbf{z}$. Notice that the covariance matrix is Hermitian and positive semi-definite, while the relation matrix is symmetric. It is useful to define the *extended* covariance matrix

$$R := \begin{bmatrix} \Gamma & C \\ \bar{C}^\top & \bar{\Gamma} \end{bmatrix}.$$

The density function of the $\mathbf{z}$ is the joint probability density of the $2n$-dimensional compound vector $\mathbf{v} = [\mathbf{x}^\top \mathbf{y}^\top]^\top$. The differential entropy of the n-dimensional, zero mean, complex Gaussian vector $\mathbf{z}$, whose probability density is p, is given by

$$\begin{aligned} H(p) := H(\mathbf{z}) &= -\int_{\mathbb{R}^{2n}} \log(p(x))p(x)dx \\ &= \frac{1}{2}\log(\det R) + \frac{1}{2}(2n)\left(1+\log(2\pi)\right), \end{aligned} \tag{A.13}$$

where R is the covariance matrix of the $2n$-dimensional vector γ. Similarly, the relative entropy between two zero-mean n-dimensional complex Gaussian densities p and q is given by

$$\mathcal{D}(p \,\|\, q) := \frac{1}{2}\left[\log\det(R_p^{-1}R_q) + \mathrm{Tr}(R_q^{-1}R_p) - 2n\right], \tag{A.14}$$

where R_p and R_q are the covariance matrices of the $2n$-dimensional vectors $\mathbf{z}_p$ and $\mathbf{z}_q$ corresponding to the densities p and q, respectively.

A.4 CIRCULAR SYMMETRIC COMPLEX GAUSSIAN RANDOM VECTORS

If the zero-mean $\mathbb{C}^k$-valued Gaussian random vector $\mathbf{z}$ has the property that the relation matrix is zero, i.e. $\mathbb{E}\left[zz^\top\right] = 0$, we say that $\mathbf{z}$ is a *circular symmetric* normally distributed random vector. Then, we have

$$p(\mathbf{z}) = \frac{1}{\pi^k \det\Gamma}\exp\{-\bar{\mathbf{z}}^\top\Gamma^{-1}\mathbf{z}\}. \tag{A.15}$$

Remark A.4.1. Given the $\mathbb{C}^k-$valued vector $\mathbf{z} = \mathbf{x} + j\mathbf{y}$, the fact that it is circular symmetric implies that, $\forall\, n,\, m \in [1,\ldots,k]$, the following conditions hold:

$$\begin{aligned} \mathbb{E}\left[\mathbf{x}_n\mathbf{x}_m\right] &= \mathbb{E}\left[\mathbf{y}_n\mathbf{y}_m\right] \\ \mathbb{E}\left[\mathbf{x}_n\mathbf{y}_m\right] &= -\mathbb{E}\left[\mathbf{x}_m\mathbf{y}_n\right] \end{aligned}$$

If p and q are two n-dimensional complex Gaussian distribution with circular symmetry, the expression of the relative entropy simplifies to the formula

$$\mathcal{D}(p \,\|\, q) = \log\det(P^{-1}Q) + \mathrm{Tr}(Q^{-1}P) - n, \tag{A.16}$$

where P and Q are the covariance matrices of $\mathbf{z}_p := \mathbf{x}_p + j\mathbf{y}_p$ and $\mathbf{z}_q := \mathbf{x}_q + j\mathbf{y}_q$, respectively.

A.5 STOCHASTIC PROCESSES

Let $T \subseteq \mathbb{R}$. Then $\mathbf{y}(t,\omega) = \{\mathbf{y}(k,\omega),\ k \in T\}$, where $\mathbf{y}(k,\omega)$, for each k, is a random variable defined with regard to $(\Omega, \mathcal{F}, P)$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, is a (real-valued) stochastic process. A stochastic process can be seen as a family of deterministic signals parametrized by ω. Indeed, by fixing the value of ω, we obtain a particular trajectory (or realization) of the process. Next, we will be mainly interested in discrete-time processes, that occur whenever $T \subseteq \mathbb{Z}$.

Consider a discrete time random process $\mathbf{y}(t,\omega)$. Its complete statistical description is given by the family $\{F_n,\ n \in \mathbb{N}\}$ of the probability distributions of the random variables $\mathbf{y}(t_1,\omega),\dots,\mathbf{y}(t_n,\omega)$, for each possible choice of the sequence $\{t_k\}_{k=1}^n$, such that $t_i \neq t_j$ for $i \neq j$[2]. Given n, we have

$$F_n : \mathbb{R}^n \times \mathbb{Z}^n \to [0,1]$$
$$F_n(x_1,\dots,x_n;t_1,\dots,t_n) = P[\mathbf{y}(t_1,\omega) \leqslant x_1,\dots,\mathbf{y}(t_n,\omega) \leqslant x_n], \quad \forall\, x_i \in \mathbb{R}, \forall\, t_i \in \mathbb{Z}.$$

If F_n is absolutely continuous, an equivalent description of the process is given by the probability density of order n

$$f_n(x_1,\dots,x_n;t_1,\dots,t_n) := \frac{\partial^n F_n(x_1,\dots,x_n;t_1,\dots,t_n)}{\partial x_1 \cdots \partial x_n}, \quad \forall\, x_i \in \mathbb{R}, \forall\, t_i \in \mathbb{Z}. \tag{A.17}$$

In the following, we will drop the dependency on ω and denote a stochastic process by the short hand notation $\mathbf{y}(t)$. Then, we can compute the mean as

$$m(t) := \mathbb{E}[\mathbf{y}(t)] := \int_{\mathbb{R}} x f_1(x;t)\,dx, \quad \forall\, t \in \mathbb{Z} \tag{A.18}$$

and the correlation as

$$r(t,s) := \mathbb{E}[\mathbf{y}(t)\mathbf{y}(s)] := \int_{\mathbb{R}^2} x_1 x_2 f_2(x_1,x_2;t,s)\,dx_1\,dx_2, \quad \forall\, t,s \in \mathbb{Z}. \tag{A.19}$$

Notice that correlation is a symmetric and positive semi-definite function.

A key property for stochastic processes is stationarity. A process is *strict-sense* stationary if, for all $n \in \mathbb{N}$, for all $\tau \in \mathbb{Z}$,

$$F_n(x_1,\dots,x_n;t_1,\dots,t_n) = F_n(x_1,\dots,x_n;t_1+\tau,\dots,t_n+\tau), \quad \forall\, x_i \in \mathbb{R}, \forall\, t_i \in \mathbb{Z}. \tag{A.20}$$

In the following, we assume that a weaker condition holds. Indeed, we require the process to be *wide-sense* stationary, i.e. such that

2 Notice that the family $\{F_n,\ n \in \mathbb{N},\dots,n\}$ must satisfy some compatibility conditions.

1. $m(t) = m, \forall\, t \in \mathbb{Z}$
2. $r(t, s) = r(t + \tau, s + \tau), \forall\, t, s, \tau \in Z$

From now on, the mean is assumed to be zero. Since for wide-sense stationary processes the correlation only depends on the difference between time indexes, with a slight abuse of notation we will drop the dependence on the second index and write simply $r(t - s)$ instead of $r(t, s)$. A *second order process* is the equivalence class of all the processes with the same mean and correlation.

The stochastic properties of a wide-sense stationary random process can also be studied in the frequency domain. Indeed, under the assumption that the Fourier series converges we can define the *spectral density* of the process

$$\Phi(e^{j\vartheta}) := \sum_{k=-\infty}^{\infty} r(k)e^{-jk\vartheta}, \quad \vartheta \in \mathbb{T}, \tag{A.21}$$

where $\mathbb{T}$ denotes the interval $(-\pi, \pi]$. Then, it holds that

$$r(\tau) := \int_{\mathbb{T}} \Phi(e^{j\vartheta})e^{jk\vartheta}\frac{d\vartheta}{2\pi}, \quad \vartheta \in \mathbb{T}, \tag{A.22}$$

By Bochner's theorem, since the correlation function is positive semi-definite it turns out that $\Phi(e^{j\vartheta}) \geqslant 0$ on $\mathbb{T}$. Since the $r(\tau)$ is even and real, we also have that $\Phi(e^{j\vartheta}) = \overline{\Phi(e^{j\vartheta})} = \Phi(e^{-j\vartheta})$, for all $\vartheta \in \mathbb{T}$.

We briefly introduce multivariate processes. Given the probability space $(\Omega, \mathcal{F}, P)$, a discrete-time m-dimensional stochastic process is defined as a function

$$\mathbf{y}(t, \omega) : \mathbb{Z} \times \Omega \to \mathbb{R}^m \tag{A.23}$$

such that its scalar components $\mathbf{y}_1(t, \omega), \ldots, \mathbf{y}_m(t, \omega)$ are random variables defined on the probability space, for all $t \in \mathbb{Z}$. In the following we will just write $\mathbf{y}(t)$ instead of $\mathbf{y}(t, \omega)$. For wide-sense stationary, multivariate stochastic processes we have

$$R(\tau) := \mathbb{E}\left[\mathbf{y}(t + \tau)\mathbf{y}(t)^\top\right],$$
$$R(\tau) = \overline{R(-\tau)}^\top.$$

As for the spectral density, it satisfies

$$\Phi(e^{j\vartheta}) = \Phi(e^{j\vartheta})^*. \tag{A.24}$$

where the notation A^* denotes the complex conjugate and transpose of A.

Consider now the *positive real part* Φ_+ of the spectral density Φ:

$$\Phi_+(z) = \frac{1}{2}\int_{\mathbb{T}} \Phi(e^{j\vartheta})\frac{z+e^{j\vartheta}}{z-e^{j\vartheta}}\frac{d\vartheta}{2\pi} \tag{A.25}$$

This is a positive real function, i.e. it is analytic with positive real part in $|z| > 1$. Thus, the power spectral density can be expressed as

$$\Phi(e^{j\vartheta}) = \Phi_+(e^{j\vartheta}) + \Phi_+(e^{j\vartheta})^*, \tag{A.26}$$

and Φ_+ admits the following representation for $|z| > 1$:

$$\Phi_+ = \frac{1}{2}R(0) + \sum_{k=1}^{\infty} R(k)z^{-k}. \tag{A.27}$$

As we already mentioned, each set of second order properties defines a class of equivalence rather than a specific process. Among the elements of such class, it is often convenient to consider the Gaussian representative, because it is completely specified by the second order description. Recall that a process $\mathbf{y}(t)$ is Gaussian if and only if, for every finite set of indexes $t_1, \dots, t_k$ in the index set $\mathbb{Z}$, we have that $\left[\mathbf{y}(t_1)^\top \cdots \mathbf{y}(t_k)^\top\right]^\top$ is a multivariate Gaussian vector.

A.6 CIRCULAR SYMMETRIC GAUSSIAN PROCESSES

Consider a discrete time multivariate stochastic process $\mathbf{z}(t)$ with values in $\mathbb{C}^k$. Let $\mathbf{x}(t) := \Re\mathbf{z}(t)$ and $\mathbf{y}(t) := \Im\mathbf{z}(t)$, so that

$$\mathbf{z}(t) = \mathbf{x}(t) + j\mathbf{y}(t).$$

Assume the processes is zero-mean. Then, its second order description is given by

$$\begin{aligned} R_{\mathbf{xx}}(t_1, t_2) &:= \mathbb{E}\left[\mathbf{x}(t_1)\mathbf{x}(t_2)^\top\right], \\ R_{\mathbf{yy}}(t_1, t_2) &:= \mathbb{E}\left[\mathbf{y}(t_1)\mathbf{y}(t_2)^\top\right], \\ R_{\mathbf{xy}}(t_1, t_2) &:= \mathbb{E}\left[\mathbf{x}(t_1)\mathbf{y}(t_2)^\top\right]. \end{aligned}$$

Thus, the vector covariance $\Gamma_{\mathbf{z}}(t_1, t_2) := \mathbb{E}\left[\mathbf{z}(t_1)\mathbf{z}(t_2)^*\right]$ is not a sufficient statistics: Indeed, in order to convey the same information given by the previously introduced functions, it is necessary to consider also the *relation function* (also known as *complementary covariance*)

$$C_{\mathbf{z}}(t_1, t_2) := \mathbb{E}\left[\mathbf{z}(t_1)\mathbf{z}(t_2)^\top\right]. \tag{A.28}$$

A special case is given by *circular symmetric* (or *proper*) processes, see Fuhrmann [43], Picinbono and Bondon [87], Wahlberg and Schreier [102]: A multivariate complex-valued process $\mathbf{z}(t)$ is circular symmetric if

$$\mathbb{E}\left[\mathbf{z}(t_1)\mathbf{z}(t_2)\right] = 0 \quad \forall\, t_1, t_2 \in \mathbb{Z} \tag{A.29}$$

This means that each scalar process $\mathbf{z}_i(t)$, $i \in \{1, \dots, k\}$ is circular symmetric. Moreover, $\mathbf{z}_i(t)$, $\mathbf{z}_j(t)$ are jointly circular symmetric $\forall\, i, j \in \{1, \dots, k\}$. Denote by $\mathbf{x}(t)$ and $\mathbf{y}(t)$ the real and imaginary part of $\mathbf{z}(t)$, respectively. Then a circular symmetric process is such that:

1. $\forall\, t_1, t_2, \in \mathbb{Z}$, $R_{\mathbf{xx}}(t_1, t_2) = R_{\mathbf{yy}}(t_1, t_2)$;
2. $\forall\, t_1, t_2, \in \mathbb{Z}$, $R_{\mathbf{xy}}(t_1, t_2) = -R_{\mathbf{yx}}(t_1, t_2)$;

Let us focus on second order stationary complex-valued processes, i.e. such that $\Gamma(t_1, t_2) = \Gamma(t_1 - t_2)$ and $C(t_1, t_2) = C(t_1 - t_2)$. Circular symmetry implies that the cross-covariance function is such that $R_{\mathbf{xy}}(0) = 0$, i.e. $\forall\, t \in \mathbb{Z}$, $\mathbf{x}(t)$ and $\mathbf{y}(t)$ are uncorrelated.

A.7 A GEOMETRIC PERSPECTIVE

Consider a family of real-valued random variables defined on the probability space $(\Omega, \mathcal{F}, P)$, such that their mean is zero and they have finite variance. They generate a linear vector space $\mathbf{H}$ and we can define the inner product $\langle \cdot, \cdot \rangle_{\mathbf{H}}$ by

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{H}} = \mathbb{E}\left[\mathbf{x}\mathbf{y}\right]. \tag{A.30}$$

It induces the norm

$$\|\mathbf{y}\|_{\mathbf{H}} := \mathbb{E}\left[\mathbf{y}^2\right] = \text{var}\left[\mathbf{y}\right]. \tag{A.31}$$

It can be proved that $\mathbf{H}$, equipped with $\|\cdot\|_{\mathbf{H}}$, is a complete space, so it is Hilbert. Convergence with regard to the induced norm is called *quadratic mean* convergence. This notion can be extended to random vectors. Analogously, we can deal with the $m-$dimensional stochastic process $\mathbf{y}(t)$, by introducing the the Hilbert space $\mathbf{H}(\mathbf{y})$ as the closure in $L^2(\Omega, P)$ of all the finite linear combinations of the random variables $\mathbf{y}_k(t)$, for $k = 1, \dots, m$, $t \in \mathbb{Z}$, i.e.

$$\mathbf{H}(\mathbf{y}) := \overline{\text{span}\{y_1(t), \dots, y_m(t); t \in \mathbb{Z}\}}. \tag{A.32}$$

Then, linear estimation has a very natural interpretation in terms of projection. Of course, we can consider also subspaces of $\mathbf{H}(\mathbf{y})$. For instance, in facing prediction

problems, it is useful to introduce the vector space generated by past values of the process $\mathbf{x}(t)$:

$$\mathbf{H}_t(\mathbf{y}) := \overline{\text{span}\{y_1(s), \dots, y_m(s); s < t\}}. \tag{A.33}$$

A.8 LINEAR SYSTEMS

For the sake of clarity, next we recall some notions about discrete-time LTI systems. A linear time invariant map from u to y is represented by the convolution

$$y(k) = \sum_{l \in \mathbb{Z}} h(l)u(k-l), \tag{A.34}$$

where h is called *impulse response* of the system, by considering the Kronecker delta as input signal for the system, we get

$$y(k) = \sum_{l=-\infty}^{\infty} h(l)\delta(k-l) = h(k). \tag{A.35}$$

If we introduce the formal series

$$\begin{aligned} U(z) &= \sum_{k=\mathbb{Z}} u(k)z^{-k}, \\ Y(z) &= \sum_{k=\mathbb{Z}} y(k)z^{-k}, \\ H(z) &= \sum_{k=\mathbb{Z}} h(k)z^{-k}, \end{aligned}$$

where z^{-1} denotes the unitary delay operator, we can write

$$Y(z) = H(z)U(z). \tag{A.36}$$

With a slight abuse of notation, we will often write $y(t) = H(z)u(t)$ instead of (A.36). In the following, we will consider multivariate signals and rational transfer function, so we can write

$$H(z) = B(zI - A)^{-1}C + D, \tag{A.37}$$

and thus obtain a state-space model for the system:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k) \end{cases}.$$

A.9 PURELY NON DETERMINISTIC PROCESSES

Consider the *remote past* of a second order process $\mathbf{y}(t)$, i.e.

$$\mathbf{H}_{-\infty}(\mathbf{y}) := \cap_{t \leqslant k} \mathbf{H}_t(\mathbf{y}). \tag{A.38}$$

In light of Wold's theorem, the processes whose remote past is trivial, i.e. such that $\mathbf{H}_{-\infty}(\mathbf{y}) = 0$, are *purely non deterministic*. By Szëgo-Kolmogorov theorem, full rank processes whose spectral power distribution is absolutely continuous are purely non deterministic if and only if

$$\int_{\mathbb{T}} \log \det \Phi(e^{j\vartheta}) \frac{d\vartheta}{2\pi} > -\infty. \tag{A.39}$$

Another fundamental result in Wold's theory is that wide-sense stationary, purely non deterministic processes can be represented as the output of a causal ℓ^2-stable time invariant linear filter driven by white noise, i.e.

$$\mathbf{y}(t) = \sum_{k=0}^{\infty} w(k)\mathbf{e}(t-k), \tag{A.40}$$

with $\mathbb{E}\left[\mathbf{e}(k)\mathbf{e}(s)^*\right] = I\delta_{sk}$.

A.10 LINEAR FILTERING OF STOCHASTIC PROCESSES

The main result on linear filtering of stochastic processes is that the output signal of a BIBO linear time-invariant filter with impulse response H fed by a wide-sense stationary process $\mathbf{x}(t)$ is a stochastic process defined in mean square sense by

$$\mathbf{y}(t) = \sum_{k \in \mathbb{Z}} h(k)\mathbf{x}(t-k) \tag{A.41}$$

This process is jointly stationary with regard to the input process $\mathbf{x}(t)$. Let $H(e^{j\vartheta})$ be the frequency response of the filter. Then, the spectral density of $\mathbf{y}(t)$ is given by the Wiener-Kintchine formula:

$$\Phi_y(e^{j\vartheta}) = H(e^{j\vartheta})\Phi_x(e^{j\vartheta})H(e^{j\vartheta})^*. \tag{A.42}$$

As a consequence of Wold's theorem, if

$$\mathbf{y}(t) = \sum_{k=0}^{\infty} w(k)\mathbf{e}(t-k), \tag{A.43}$$

with Var $[\mathbf{e}(t)] = I$, we obtain the spectral factorization

$$\Phi_y(e^{j\vartheta}) = W(e^{j\vartheta})W(e^{j\vartheta})^*. \tag{A.44}$$

If W is rational we can write it as

$$W(z) = Q(z)^{-1}P(z), \tag{A.45}$$

where $Q(z)$ and $P(z)$ are multivariate matrix-valued polynomials. Thus, we can introduce the auto regressive moving average (ARMA) representation

$$\sum_{k=0}^{n} Q_k \mathbf{y}(t-k) = \sum_{k=0}^{n} P_k \mathbf{e}(t-k) \tag{A.46}$$

where e is m-dimensional white noise. If $Q(z) = I$ we get a moving average (MA) representation

$$\mathbf{y}(t) = P(z)\mathbf{e}(t), \tag{A.47}$$

whereas if $P(z) = I$ we obtain an auto regressive (AR) representation

$$Q(z)\mathbf{y}(t) = \mathbf{e}(t). \tag{A.48}$$

A.11 MOMENT PROBLEMS

Suppose we want to compute a measure μ such that it is consistent with some given moments

$$m_k := \int x^k d\mu(x), \quad k \in \mathcal{K}. \tag{A.49}$$

Then, a moment problem arises. Generalized moment problems occur when the task is inverting the map from the measure μ to a sequence of generalized moments

$$g_k := \int G_k(x) d\mu(x), \tag{A.50}$$

where $\{G_k\}_{k\in\mathcal{K}} \in$ is a family of arbitrary functions.

To our purposes, generalized moment problems play a key role. In particular, the measures we are going to deal with will be either spectral densities or probability density functions.

A.12 HYPOTHESIS TESTING

Hypothesis testing is a well-established procedure in decision theory. Here we consider the simplest case. Assume we collect n observations $x_1, \ldots, x_n$ of i.i.d. random variables distributed according to the density function $Q(x)$. We want to decide which of the following hypotheses best fits the observations:

- $Q = H_0$, where H_0 is the *null hypothesis*,
- $Q = H_1$, where H_1 is the *alternative hypothesis*.

The decision can be modeled by means of a function

$$g : \mathbb{R}^n \to \{0, 1\}$$
$$g(x_1, \ldots, x_n) \mapsto \begin{cases} 0 & \text{if } H_0 \text{ is accepted} \\ 1 & \text{if } H_1 \text{ is accepted} \end{cases}$$

Since g can take only two values, it can be equivalently modeled by considering the region $A \subset \mathbb{R}^n$ such that $g(x_1, \ldots, x_n) = 0, \forall\, [x_1, \cdots, x_n]^\top \in A$. Let A^c be the complementary set of A in $\mathbb{R}^n$. Then, we can introduce two error probabilities:

TYPE I ERROR PROBABILITY $\alpha := P\left[g(x_1, \ldots, x_n) = 1 \,|\, H_0 \text{ is true}\right] = P_{H_0}(A^c)$

TYPE II ERROR PROBABILITY $\beta := P\left[g(x_1, \ldots, x_n) = 0 \,|\, H_1 \text{ is true}\right] = P_{H_1}(A)$

The capability of correctly rejecting H_0 when it is false is measured by $1 - \beta$ and is called *power* of the test.

BIBLIOGRAPHY

[1] E. Avventi, A. Lindquist, and B. Wahlberg. "ARMA Identification of Graphical Models." In: *IEEE Trans. Aut. Control* 58.5 (2013), pp. 1167 –1178.

[2] P. Baracca, N. Laurenti, and S. Tomasin. "Physical layer authentication over MIMO fading wiretap channels." In: *IEEE Trans. Wireless Commun.* 11.7 (2012), pp. 2564 –2573.

[3] M. Barni and B. Tondi. "The Source Identification Game: An Information-Theoretic Perspective." In: *IEEE Trans. on Inform. Forens. Security* 8.3 (2013), pp. 450–463.

[4] M. Basseville. "Distance Measures for Signal Processing and Pattern Recognition." In: *Signal Processing* 18 (1989), pp. 349–369.

[5] A. Blomqvist, A. Lindquist, and R. Nagamune. "Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach." In: *IEEE Trans. Aut. Control* 48 (2003), pp. 2172–2190.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[7] D.R. Brillinger. "Remarks concerning graphical models for time series and point processes." In: *Rivista de econometria* 16 (1996), pp. 1–23.

[8] J. Burg, D. Luenberger, and D. Wenger. "Estimation of structured covariance matrices." In: *Proceedings of the IEEE* 70 (1982), pp. 963–974.

[9] C. I. Byrnes, P. Enqvist, and A. Linquist. "Cepstral coefficients, covariance lags and pole-zero models for finite data strings." In: *IEEE Trans. Sig. Proc.* 50 (2001), pp. 677–693.

[10] C. I. Byrnes, P. Enqvist, and A. Linquist. "Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach." In: *SIAM J. Control and Optimization* 41 (2002), pp. 23–59.

[11] C. I. Byrnes, T. Georgiou, and A. Lindquist. "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint: A convex optimization approach to certain problems in systems and control." In: *IEEE Trans. Aut. Control* 46 (2001), pp. 822–839.

[12] C. I. Byrnes, T. Georgiou, and A. Lindquist. "A new approach to spectral estimation: A tunable high-resolution spectral estimator." In: *IEEE Trans. Sig. Proc.* 49 (2000), pp. 3189–3205.

[13] C. I. Byrnes, S. Gusev, and A. Lindquist. "A convex optimization approach to the rational covariance extension problem." In: *SIAM J. Control and Opimization* 37 (1999), pp. 211–229.

[14] C. I. Byrnes, S. Gusev, and A. Lindquist. "From finite covariance windows to modeling filters: A convex optimization approach." In: *SIAM Review* 43 (2001), pp. 645–675.

[15] C. I. Byrnes and A. Lindquist. "The generalized moment problem with complexity constraint." In: *Integral Equations and Operator Theory* 56(2) (2006), pp. 163–180.

[16] Modern Analysis and Application: Mark Krein Centenary Conference, eds. *The moment problem for rational measures: convexity in the spirit of Krein*. Vol. I: Operator Theory and Related Topics. Operator Theory Advances and Applications. Birkhäuser, 2009, pp. 157 –169.

[17] C. I. Byrnes and A. Linquist. "Important moments in systems and control." In: *SIAM J. Control and Optimization* 47.5 (2008), pp. 2458–2469.

[18] C. I. Byrnes and A. Linquist. "Interior point solutions of variational problems and global inverse function theorems." In: *International Journal of Robust and Nonlinear Control* 17 (2007), pp. 463–481.

[19] C. I. Byrnes et al. "Generalized interpolation in H-infinity with a complexity constraint." In: *Trans. American Math. Society* 358(3) (2006), pp. 965–987.

[20] C.I. Byrnes and A. Lindquist. "On the partial stochastic realization problem." In: *IEEE Trans. Aut. Contr.* 42 (1997), pp. 1049–1070.

[21] C.I. Byrnes et al. "A complete parameterization of all positive rational extensions of a covariance sequence." In: *IEEE Trans. Aut. Contr.* 40 (1995), pp. 1841–1857.

[22] C. Cachin. "An Information-Theoretic Model for Steganography." In: *International Workshop on Information Hiding*. Vol. LNCS-1525. Portland, OR, 1998.

[23] F. Carli and T.T. Georgiou. "On the covariance completion problem under a circulant structure." In: *IEEE Trans. Aut. Cont.* 56(4) (2011), pp. 918–922.

[24] F. Carli et al. "A Maximum Entropy solution of the Covariance Extension Problem for Reciprocal Processes." In: *IEEE Trans. Aut. Control* 56 (2011), pp. 1999 –2012.

[25] F.P. Carli et al. "An efficient algorithm for maximum entropy extension of block-circulant covariance matrices." In: *Linear Algebra and its Applications* 439.8 (2013), pp. 2309 –2329. ISSN: 0024-3795.

[26] A. Chiuso, A. Ferrante, and G. Picci. "Reciprocal realization and modeling of textured images." In: *Proc. 44rd IEEE CDC*. 2005.

[27] T. M. Cover and J. A. Thomas. *Information Theory*. New York: Wiley, 1991.

[28] R. Dahlhaus. "Graphical interaction models for multivariate time series." In: *Metrika* 51.2 (2000), pp. 157–172.

[29] P. Dai Pra. *Private communication*. June 2011.

[30] T. Daniels, M. Mina, and S.F. Russell. "A Signal Fingerprinting Paradigm for General Physical Layer and Sensor Network Security and Assurance." In: *IEEE SECURECOMM*. Athens, Greece, Sep. 2005, pp. 1–3.

[31] A. Dembo and O. Stroock. *Large Deviation Techniques and Applications*. Jones and Bartlett Publishers, 1993.

[32] A. P. Dempster. "Covariance selection." In: *Biometrics* 28 (1972), pp. 157–175.

[33] J. D. Deuschel and D. W. Stroock. *Large deviations*. New York: Academic Press, 1989.

[34] P. Enqvist. "A homotopy approach to rational covariance extension with degree constraint." In: *Int. J. Appl. Math. and Comp. Sci.* 11 (2001), pp. 1173–1201.

[35] P. Enqvist. "On the simultaneous realization problem: Markov parameter and covariance interpolation." In: *Signal Processing* 86 (10) (2006), pp. 3043–3054.

[36] P. Enqvist. "Spectral estimation by Geometric, Topological and Optimization Methods." PhD thesis. 2001.

[37] P. Enqvist and J. Karlsson. "Minimal Itakura-Saito distance and covariance interpolation." In: *47th IEEE Conference on Decision and Control, CDC.* 2008, pp. 137 –142.

[38] D.B. Faria and D.R. Cheriton. "Detecting identity-based attacks in wireless networks using signalprints." In: *ACM WiSe*. Los Angeles, 2006, pp. 43–52.

[39] A. Ferrante and M. Pavon. "Matrix Completion *à la* Dempster by the Principle of Parsimony." In: *IEEE Trans. Information Theory* 57.6 (2011), pp. 3925–3931.

[40] A. Ferrante, M. Pavon, and F. Ramponi. "Hellinger vs. Kullback-Leibler multivariable spectrum approximation." In: *IEEE Trans. Aut. Control* 53 (2008), pp. 954–967.

[41] A. Ferrante, M. Pavon, and M. Zorzi. "A maximum entropy enhancement for a family of high-resolution spectral estimators." In: *IEEE Trans. Aut. Control* 57 (2012), pp. 318 –329.

[42] A. Ferrante, F. Ramponi, and F. Ticozzi. "On the convergence of an efficient algorithm for Kullback-Leibler approximation of spectral densities." In: *IEEE Trans. Aut. Control* 56 (2011), pp. 506–515.

[43] D.R. Fuhrmann. *Complex Random Variables and Stochastic Processes*. CRC Press LLC, 1999.

[44] T. Georgiou. "Distances and Riemannian Metrics for Spectral Density Functions." In: *IEEE Trans. on Signal Processing* 55(8) (August 2007), pp. 3995–4003.

[45] T. Georgiou. "Distances between power spectral densities." In: *IEEE Trans. Aut. Control* 47 (2006), pp. 1056–1066.

[46] T. Georgiou. "Partial Realization of Covariance Sequences." PhD Thesis. CMST, University of Florida, 1983.

[47] T. Georgiou. "Realization of power spectra from partial covariance sequences." In: *IEEE Trans. on Acoustics, Speech, and Signal Processing* 35 (1987), pp. 438–449.

[48] T. Georgiou. "Relative entropy and the multivariable multidimensional moment problem." In: *IEEE Trans. Inform. Theory* 52 (2006), pp. 1052–1066.

[49] T. Georgiou. "Solution of the general moment problem via a one-parameter imbedding." In: *IEEE Trans. Aut. Control* 50 (2005), pp. 811–826.

[50] T. Georgiou. "Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization." In: *IEEE Trans. Aut. Control* 47 (2002), pp. 1811–1823.

[51] T. Georgiou. "Spectral estimation by selective harmonic amplification." In: *IEEE Trans. Aut. Control* 46 (2001), pp. 29–42.

[52] T. Georgiou. "The interpolation problem with a degree constraint." In: *IEEE Trans. Aut. Control* 44 (1999), pp. 631–635.

[53] T. Georgiou. "The Meaning of Distance in Spectral Analysis." In: *46th IEEE Conference on Decision and Control, New Orleans, U.S.A.* 2007, http://www.ieeecss–oll.org/video/meaning–distances–spectral–analysis.

[54] T. Georgiou. "The structure of state covariances and its relation to the power spectrum of the input." In: *IEEE Trans. Aut. Control* 47 (2002), pp. 1056–1066.

[55] T. Georgiou and A. Lindquist. "A convex optimization approach to ARMA modeling." In: *IEEE Trans. Aut. Control* AC-53 (2008), pp. 1108–1119.

[56] T. Georgiou and A. Lindquist. "Kullback-Leibler approximation of spectral density functions." In: *IEEE Trans. Inform. Theory* 49 (2003), pp. 2910–2917.

[57] T. Georgiou and A. Lindquist. "Remarks on control design with degree constraint." In: *IEEE Trans. Aut. Control* AC-51 (2006), pp. 1150–1156.

[58] R. Gray et al. "Distortion measures for speech processing." In: *IEEE Trans. Acoustics, Speech and Signal Proc.* 28 (1980), pp. 367–376.

[59] S. Ihara. *Information Theory for Continuous Systems*. Singapore: World Scientific, 1993.

[60] X. Jiang, L. Ning, and T. Georgiou. "Distances and Riemannian metrics for multivariate spectral densities." In: *IEEE Trans. Aut. Control* 57 (2012), pp. 1723 –1735.

[61] R. E. Kalman. "Realization of Covariance Sequences." In: *Proc. Toeplitz Memorial Conference, Tel Aviv, Israel*. 1981.

[62] J. Karlsson and T. Georgiou. "Uncertainty Bounds for Spectral Estimation." In: *IEEE Trans. Aut. Control* 58.7 (2013).

[63] J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman & Hall/CRC, 2008.

[64] A.N. Kolmogorov. "On the Shannon theory of information in the case of continuous signals." In: *IRE Trans. Inform. Theory* 2 (1956), pp. 102–108.

[65] H. Kramer and M. R. Leadbetter. *Stationary and Related Stochastic Processes*. New York: Wiley, 1966.

[66] A. J. Krener. "Reciprocal Processes and the stochastic realization problem for acausal systems." In: *Modeling Identification and Robust Control*. Ed. by C.I. Byrnes and A. Lindquist. North-Holland, 1986, pp. 197–211.

[67] A. J. Krener, R. Frezza, and B.C. Levy. "Gaussian reciprocal processes andself-adjoint differential equations of second order." In: *Stochastics and Stochastics Reports* 34 (1991), pp. 29–56.

[68] S. Kullback. *Information Theory and Statistics 2nd ed.* Mineola NY: Dover, 1968.

[69] L. Lai, H. El Gamal, and H. V. Poor. "Authentication Over Noisy Channels." In: *IEEE Trans. Inf. Theory* 55.2 (2009), pp. 906–916.

[70] B.C. Levy and A. Ferrante. "Characterization of stationary discrete-time Gaussian Reciprocal Processes over a finite interval." In: *SIAM J. Matrix Anal. Appl.* 24 (2002), pp. 334–355.

[71] B.C. Levy, R. Frezza, and A.J. Krener. "Modeling and Estimation of discrete-time Gaussian Reciprocal Processes." In: *IEEE Trans. Aut. Control* 35 (1990), pp. 1013–1023.

[72] A. Lindquist. "Prediction-error approximation by convex optimization." In: *Modeling, Estimation and Control: Festschrift in honor of Giorgio Picci on the occasion of his sixty-fifth birthday*. Ed. by A. Chiuso, A. Ferrante, and S. Pinzoni. Springer-Verlag, 2007, pp. 265–275.

[73] A. Lindquist and G. Picci. *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*. In preparation: preprint available in http://www.math.kth.se/~alq/LPbook.

[74] A. Lindquist and G. Picci. "The circulant rational covariance extension problem: the complete solution." In: *IEEE Trans. Aut. Control* 58 (2013).

[75] D. Materassi and G. Innocenti. "Topological identification in networks of dynamical systems." In: *Proc. of the 47th IEEE Conference on Decision and Control*. 2008.

[76] U.M. Maurer. "Authentication theory and hypothesis testing." In: *IEEE Trans. Inf. Theory* 46 (2000), pp. 1350–1356.

[77] J. H. McClellan. "Multidimensional spectral estimation." In: *Proc. IEEE* 70 (1982), pp. 1029–1039.

[78] A. J. Menezes, P. C. V. Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press, 1996.

[79] B. R. Musicus and A. M. Kabel. *Maximum entropy pole-zero estimation*. Technical Report 510. MIT Research Lab. Electronics, 1985.

[80] A. Nasiri Amini, E. Ebbini, and T. Georgiou. "Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques." In: *IEEE Trans. on Biomedical Engineering* 52 (2005), pp. 221–228.

[81] F. D. Neeser and J. L. Massey. "Proper Complex Random Processes with Applications to Information Theory." In: *IEEE Trans. Inf. Theory* 39.4 (1993), pp. 1293–1302.

[82] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Univ. Press, 2000.

[83] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 2002.

[84] M. Pavon and A. Ferrante. "On the Geometry of Maximum Entropy Problems." In: *SIAM REVIEW* Vol. 55.No. 3 (2013), pp. 415–439.

[85] M. Pavon and A. Ferrante. "On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities." In: *IEEE Trans. Aut. Control* 51 (2006), pp. 639–644.

[86] G. Picci and F. Carli. "Modelling and simulation of images by reciprocal processes." In: *Proc. Tenth International Conference on Computer Modeling and Simulation UKSIM*. 2008, pp. 513–518.

[87] B. Picinbono and P. Bondon. "Second-Order Statistics of Complex Signals." In: *IEEE Trans. Sig. Proc.* 45.2 (1997).

[88] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated by A. Feinstein. San Francisco: Holden-Day, 1964.

[89] F. Ramponi, A. Ferrante, and M. Pavon. "A Globally Convergent Matricial Algorithm for Multivariate Spectral Estimation." In: *IEEE Trans. Aut. Control* 54.10 (2009), pp. 2376–2388.

[90] R. T. Rockafellar. *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.

[91] O. Rosen and D. Stoffer. "Automatic Estimation of Multivariate Spectra via Smoothing Splines." In: *Biometrika* 94 (2007), pp. 335–345.

[92] Y. A. Rozanov. *Stationary Random Processes*. San Francisco: Holden-Day, 1967.

[93] J. Sand. "Reciprocal realizations on the circle." In: *SIAM J. Control and Optimization* 34.2 (1996), pp. 505–520.

[94] R. Sandhu, T. T. Georgiou, and A. R. Tannenbaum. "A new distribution metric for image segmentation." In: vol. 6914. 2008.

[95] J. Songsiri and L. Vandenberghe. "Topology selection in graphical models of autoregressive processes." In: *Journal of Machine Learning Research* 11 (2010), pp. 2671–2705.

[96] T. P. Speed and H. T. Kiiveri. "Gaussian Markov distributions over finite graphs." In: *Annals of Statistics* 14.1 (1986), pp. 138–150.

[97] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. New York: Prentice Hall, 1997.

[98] A. A. Stoorvogel and J. H. Van Schuppen. "System identification with information theoretic criteria." In: *Identification, Adaptation, Learning: The Science of Learning Models from Data*. Ed. by S. Bittanti and G. Picci. Berlin-Heidelberg: Springer, 1996.

[99] D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge Univ. Press, 2005.

[100] V. Vedral. "The role of relative entropy in quantum information theory." In: *Rev. Mod. Phys* 74 (2002), pp. 197–213.

[101] J. Von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press, 1955.

[102] P. Wahlberg and P.J. Schreier. "Spectral Relations for Multidimensional Complex Improper Stationary and (Almost) Ciclostationary Processes." In: *IEEE Trans. Inform. Theory* 54.4 (2008).

[103] L. Xiao et al. "Channel-based spoofing detection in frequency-selective Rayleigh channels." In: *IEEE Trans. Wireless Commun.* 8 (2009), pp. 5948–5956.

[104] H. Zimmermann. "OSI Reference Model–The ISO Model of Architecture for Open Systems Interconnection." In: *IEEE Trans. Comm.* 28(4) (1986), pp. 425–432.

[105] M. Zorzi. "A New Family of High-Resolution Multivariate Spectral Estimators." In: *IEEE Trans. Aut. Control* (In press). DOI: 10.1109/TAC.2013.2293218.

[106] M. Zorzi. "Rational approximations of spectral densities based on the Alpha divergence." In: *Mathematics of Control, Signals, and Systems* (2013), pp. 1–20.

ACKNOWLEDGMENTS

First of all I would like to thank Augusto Ferrante for being a kind and very competent advisor. I deeply appreciate his dedication and solid methodological attitude, and I am grateful for everything he taught me. I would also like to express my sincere gratitude to Michele Pavon. His commitment to both scientific research and teaching is admirable and was really motivating for me. During my Ph.D. I had the pleasure of working with Giorgio Picci and Anders Lindquist. Our passionate conversations were always inspiring. I would also like to thank Anders for the warm hospitality in Shanghai.

I am deeply grateful to my officemates at the Department of Information Engineering in Padova. They are brilliant, passionate and above all very good friends. I learned a lot from them. This journey would not have been so stimulating without their presence. I am sure I am going to miss the wonderful working environment, our coffee-break conversations, and all the times I doubled up with laughter!

I would also like to thank my fellows at Slow Food Bassano del Grappa for all the convivial moments we shared and the exciting challenges we faced in order to raise awareness about quality food and sustainability. I am very proud of everything we achieved together!

I could not have asked for better traveling companions than my roommates Federica, Anna and Francesca. I want to thank them for the sharing, our friendly chats and the precious support they gave me.

I am grateful to all my friends from *Boars* for the time we spent together in the last three years. I also want to thank them for everything they did in order to support me closely, even when I was 9000 kilometers away. Special thanks to Elena for the magnificent Advent calendar she created for me!

I am deeply grateful to my family. I am sure I will never forget the Christmas skype call in 2012. In particular, I want to thank my parents Anna and Paolo for their support, that has been invaluable. They gave me wise advice, even during hard times, and really motivated me throughout this journey. To them I dedicate this thesis.

Finally, I want to thank Manuel for being at my side all the time (even on the Great Wall in Jinshanling, despite the freezing cold!). The last three years were very demanding for the both of us, but I knew I could count on him and this really made me feel stronger.